

Volume 1

Encyclopedia of
Public
Administration
and
Public Policy
Second Edition

Volume 2

Encyclopedia of
Public
Administration
and
Public Policy
Second Edition

Volume 3

Encyclopedia of
Public
Administration
and
Public Policy
Second Edition

Encyclopedias from Taylor & Francis Group

Agriculture Titles

Dekker Agropedia Collection (Eight Volume Set)

ISBN: 978-0-8247-2194-7 Cat. No.: DK803X

Encyclopedia of Agricultural, Food, and Biological Engineering

Edited by Dennis R. Heldman

ISBN: 978-0-8247-0938-9 Cat. No.: DK9381

Encyclopedia of Animal Science

Edited by Wilson G. Pond and Alan Bell

ISBN: 978-0-8247-5496-9 Cat. No.: DK2206

Encyclopedia of Pest Management

Edited by David Pimentel

ISBN: 978-0-8247-0632-6 Cat. No.: DK6323

Encyclopedia of Pest Management, Volume II

Edited by David Pimentel

ISBN: 978-1-4200-5361-6 Cat. No.: 53612

Encyclopedia of Plant and Crop Science

Edited by Robert M. Goodman

ISBN: 978-0-8247-0944-0 Cat. No.: DK1190

Encyclopedia of Soil Science, Second Edition (Two Volume Set)

Edited by Rattan Lal

ISBN: 978-0-8493-3830-4 Cat. No.: DK830X

Encyclopedia of Water Science

Edited by Stanley W. Trimble

ISBN: 978-0-8493-9627-4 Cat. No.: DK9627

Chemistry Titles

Encyclopedia of Chromatography, Second Edition (Two Volume Set)

Edited by Jack Cazes

ISBN: 978-0-8247-2785-7 Cat. No.: DK7851

Encyclopedia of Supramolecular Chemistry (Two Volume Set)

Edited by Jerry L. Atwood and Jonathan W. Steed

ISBN: 978-0-8247-5056-5 Cat. No.: DK056X

Encyclopedia of Surface and Colloid Science, Second Edition (Eight Volume Set)

Edited by P. Somasundaran

ISBN: 978-0-8493-9615-1 Cat. No.: DK9615

Engineering Titles

Encyclopedia of Chemical Processing (Five Volume Set)

Edited by Sunggyu Lee

ISBN: 978-0-8247-5563-8 Cat. No.: DK2243

Encyclopedia of Corrosion Technology, Second Edition

Edited by Philip A. Schweitzer, P.E.

ISBN: 978-0-8247-4878-4 Cat. No.: DK1295

Encyclopedia of Energy Engineering and Technology (Three Volume Set)

Edited by Barney L. Capehart

ISBN: 978-0-8493-3653-9 Cat. No.: DK653X

Dekker Encyclopedia of Nanoscience and Nanotechnology (Five Volume Set)

Edited by James A. Schwarz, Cristian I. Contescu, and Karol Putyera

ISBN: 978-0-8247-5055-8 Cat. No.: DK0551

Encyclopedia of Optical Engineering (Three Volume Set)

Edited by Ronald G. Driggers

ISBN: 978-0-8247-0940-2 Cat. No.: DK9403

Business Titles

Encyclopedia of Library and Information Science, Second Edition (Four Volume Set)

Edited by Miriam Drake

ISBN: 978-0-8247-2075-9 Cat. No.: DK075X

Encyclopedia of Library and Information Science, First Update Supplement

Edited by Miriam Drake

ISBN: 978-0-8493-3894-6 Cat. No.: DK8948

Encyclopedia of Corrosion Technology, Second Edition

Edited by Philip A. Schweitzer, P.E.

ISBN: 978-0-8247-4878-4 Cat. No.: DK1295

Encyclopedia of Public Administration and Public Policy, Second Edition (Three Volume Set)

Edited by Evan M. Berman

ISBN: 978-0-4200-5275-6 Cat. No.: AU5275

Encyclopedia of Wireless and Mobile Communications (Three Volume Set)

Edited by Borko Furht

ISBN: 978-0-4200-4326-6 Cat. No.: AU4326

These titles are available both in print and online. To order, visit:

www.crcpress.com

Telephone: 1-800-272-7737

Fax: 1-800-374-3401

E-Mail: orders@crcpress.com

Volume 1
1993 - Environment

Encyclopedia of
Public
Administration
and
Public Policy

Second Edition

Edited by Evan M. Berman
Jack Rabin, Founding Editor



Taylor & Francis
Taylor & Francis Group
New York London

Taylor & Francis is an imprint of the
Taylor & Francis Group, an **informa** business

Volume 2
Equal - Organizational

Encyclopedia of
Public
Administration
and
Public Policy
Second Edition

Edited by Evan M. Berman
Jack Rabin, Founding Editor



Taylor & Francis
Taylor & Francis Group
New York London

Taylor & Francis is an imprint of the
Taylor & Francis Group, an **informa** business

Volume 3
Papers - Zero

Encyclopedia of
Public
Administration
and
Public Policy

Second Edition

Edited by Evan M. Berman
Jack Rabin, Founding Editor



Taylor & Francis
Taylor & Francis Group
New York London

Taylor & Francis is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2008 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-1-4200-5275-6 (Hardcover)

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC) 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>

Volume 1

1993 Budget Reconciliation Bill and the Federal Budget Deficit / 1
360-Degree Performance Appraisal Systems / 6
Accountability and Ethics: Reconsidering the Relationships / 10
Accounting and Reporting for Private Non-Profit Organizations—Balancing Economic Efficiency with Social Mission / 25
Accounting, Financial / 32
Achieving Productivity through Budgeting / 38
Acquiring Resources through Price Negotiation: A Public Sector Approach / 43
Activity-Based Costing in the Public Sector / 47
Administration of State Lotteries / 53
Administrative Discretion / 58
Administrative Law Judges and Agency Adjudication / 62
Administrative Reform in Southeast Asia / 66
Agenda Setting / 71
Agricultural Policy / 76
Alternative Dispute Resolution Processes / 80
Annexation / 85
Appearance of Impropriety / 94
Appreciative Inquiry / 99
Arbitrage / 103
Asset Pricing / 110
Association of Southeast Asian Nations (ASEAN) / 114
Attractors, Strange Attractors and Fractals / 118
Auditing / 123
Balanced Scorecard for Public Administration and Policy / 130
Balanced Scorecard Tools for Organizational Assessment in Public Administration and Policy / 136
Benchmarking in Public Administration and Public Policy / 141
Biomedical Ethics and Public Policy Discourse / 147
Bond Ratings / 151
Bond Valuation / 154
Budget Execution and Management Control / 158
Budget Reform / 166
Budget Stabilization Fund / 172
Budgetary Analysis and Economics / 177
Budgetary Balance and Deficit Controls / 181
Budgetary Reforms in Southeast Asia / 186
Budgeting and Public Choice / 191
Budgeting, Line-Item and Object-of-Expenditure Controls / 195
Budgeting, Performance-Based / 197
Budgeting, Target-Based / 201
Budgeting: Allocation Formula / 205
Budgets and Accountability / 209
Bureaucracy, Economic Theories of / 217
Bureaucracy: Perspectives / 221
Bureaucratic Agencies: Delegation of Authority / 225
Bureaucratic Discretion: Political Control and Influence / 230
Bureaucratic Politics / 236
Bureaucrats and Politicians in Southeast Asia / 242
Burnout in Public Agencies / 247
Cambodia / 253
Capital Budgeting / 257
Capital Markets / 264
Capital Programming and Budgeting: Comparative Local Government Perspectives / 270
Capital Purchases / 279
Career Development / 283
Cash Flow Forecasting: Principles / 285
Cash Management for State and Local Governments / 291
Chinese Public Administration: History and Current Trends / 297
Citizen Participation in Local Government / 301
Citizen Participation in the Budget Process: Exit, Voice, and Loyalty / 305
City Government / 309
Civic Duty / 313
Civil Disobedience / 321
Civil Service / 327
Civil Service Reform Act of 1978 / 329
Civil Service Reform and Reinvention / 335
Collaborative Governance / 343
Collective Bargaining / 348
Collective Bargaining, The Politics of / 351
Community-Based Planning for HIV/AIDS / 355
Compensation: Pay and Benefits / 364
Comprehensive Annual Financial Report / 369
Conflict of Interest / 374
Consolidated Governments / 377
Constitutional Constraints on Administrative Behavior in the United States / 382
Constitutional Constraints on the Administrations of Public Policy / 391
Constructive Change Proposals, Assessing the Validity of / 395

Volume 1 (Continued)

Contracting for Social Services / 399
Control Systems / 404
Cooperative Purchasing / 410
Correctional Administration, The Challenges of / 413
Council/Commission Representation / 422
County Government / 426
Court Administrators as Managers and Executives / 432
Court Budgeting: Judicial Branch Independence and Accountability / 435
Court System Strategic Planning / 442
Court Technology and Information Systems / 448
Court Unification / 454
Courts and Policy in the United States / 459
Criminal Justice Public Policy / 463
Crisis Policy-Making and Management in Southeast Asia / 470
Customer Service / 475
Data Integrity / 480
Data Mining as a Decision-Making Aid / 485
Debt Affordability / 489
Debt Issuance / 494
Decentralization in Developing Countries / 498
Decentralization in Southeast Asia / 504
Decision Making in Administrative Organizations: Theories / 508
Decision Making, Incrementalism and Transformational Change / 514
Decision Making, Open System and Non-Equilibrium / 519
Decision Making, Rational / 524
Decision Making: Administrative Dispute Resolution / 530
Default / 535
Democracy and Public Policy / 538
Deontology / 545
Development Administration in Southeast Asia / 547
Development Policy / 552
Discipline and Dismissal / 555
Diversity / 558
Economic Development and Organization / 565
Economic Development, Citizen Participation and / 569
Economic Development, Community Organizations and / 577
Economic Development, Financing / 582
Economic Development, Immigration, Self-Employment and / 585
Economic Development, Infrastructure and / 593
Economic Development, State Government Administration of / 598
Economic Development: Governance Concepts and Issues / 602
Economic Development: International / 610
Economic Efficiency / 614
Economic Theories of Public Leadership / 618
Economics, Transaction Cost / 623
Education Policy / 629
Education, State Administration of Primary and Secondary / 633
Efficiency / 637
E-Government in East Asia / 639
Electric Industry Restructuring / 643
Electronic Government / 648
Employee Benefits / 654
Enterprise: Public versus Private / 661
Entitlements and the Congressional Budget Process / 665
Environmental Ethics / 669
Environmental Policy / 672
Environmental Pollution Control, State Government Administration of / 676

Volume 2

Equal Employment Opportunity and Affirmative Action / 681
Ethics and Administrative Reform / 686
Ethics and Information and Communication Technology / 691
Ethics and Public Policy / 696
Ethics in Organizations, Implementation of / 701
Ethics in Public Administration, Teaching of / 705
Ethics, Feminist Perspective on / 709
Ethics, Governance Structures / 712
Evolutionary Theory and Public Policy / 716
Executive Budgeting / 720
Executive Leadership in Local Government / 725
Federal Subsidies for New Entrepreneurial High-Growth Firms / 731
Federalism in "Homeland Security" and Crisis Management / 740
Feminism and Chaos Theory / 744
Feminist Comparative Policy / 747
Financial Condition / 753
Financial Emergencies / 759

Volume 2 (Continued)

Financial Resource Management in Local Government / 766
Financial Statement Analysis / 771
Fiscal Distress of Local Governments in Israel / 778
Fiscal Federalism / 783
Fiscal Transparency / 788
Foreign Policy Analysis / 796
Forms of Government / 800
Freedom of Information Act—Federal / 805
Futures and Options / 809
Game Theory / 812
Gaza Disengagement: Israel Government Forced Resettlement (2005) / 816
Gender and Public Administration / 821
Gender Bias, Ethical Analysis of and in Public Policy / 824
Generally Accepted Accounting Principles (GAAP) / 828
Geographic Information Systems and Public Policy / 833
Global Ethics and Corruption / 838
Governance Networks / 844
Government Failure: Theory / 851
Government Performance and Results Act of 1993: A Retrospective / 854
Government–Business Relations in East Asia / 860
Government-Sponsored Enterprises (GSEs) / 866
Governors as Chief Administrators, The Changing Role of / 871
Gramm–Rudman–Hollings: Impact on the Federal Budgetary Process / 877
Grievance Procedures and Administration / 883
Groups and Their Properties / 888
Health Care Decision Analysis, Alternatives in / 893
Health Care Decision Making Behavior, Emerging Paradigms of / 897
Health Care Finance in the Twenty-First Century / 902
Health Care Organizations, Managing / 906
Health Care Policy / 912
Health Care Settings, Change in / 916
Health Care, Assessment and Evaluation of / 923
Health Policy in Singapore / 929
Home Rule / 933
Human Relations Management Theory / 938
Human Resource Management / 942
Human Resource Management for Productivity / 945
Human Resource Planning / 949
Human Resources Administrators in State Government, Changing Roles of / 953
Human Resources Management in Local Government / 962
Human Rights, Ethical Analysis of and in Public Policy / 965
Human Subjects Research: United States Public Policy / 970
Humanitarian Intervention / 975
Impacts of Bureaucratic Reform on State Government Administration / 981
Impasse Resolution Procedures / 987
Incrementalism, Budgetary / 992
Indonesia / 997
Information Sources and State Policy Making / 1001
Information Systems / 1005
Innovation and Public Policy / 1009
Integrated Healthcare Systems: New Trends, Emerging Models, and Future Shocks / 1020
Internal Controls / 1024
International Public Sector Accounting Standards / 1029
Inventory Management / 1034
Israel Civil Society: Characteristics / 1037
Israel: Peace and Conflict Resolution Organizations / 1042
Israel: Public Law / 1046
Israel: Spatial Policy / 1049
Job Classification / 1054
John Rawls / 1063
Judicial Ethics / 1067
Judicial Independence / 1071
Judicial Selection and Merit Selection / 1076
Just War / 1079
Justice through Emancipation: Abraham Lincoln as a Model of Justice / 1084
Juvenile Stalking / 1089
Juvenile Waiver / 1093
Knowledge Utilization: Issues and Recommendations / 1097
Labor Policy in Thailand / 1113
Labor–Management Partnerships and Cooperation / 1117
Labor–Management Rights and Responsibilities / 1121
Lao People's Democratic Republic (PDR) / 1128
Law and Ethics / 1132
Leadership / 1136
Leadership in Organizations / 1139
Leadership: High-Integrity / 1145
Learning Organizations / 1149
Lebanon: Corruption / 1153

Volume 2 (Continued)

Legal Ethics / 1157
Life Cycle Costing / 1164
Linear Programming / 1168
Local Governments: State Supervision of Budgets and Finance / 1171
Local Services, Provision and Production / 1175
Logical Positivism and Public Administration / 1180
Logistics and Transportation / 1184
Malaysia / 1189
Management for Quality in Government / 1194
Managing Mutual Commitments and Expectations: Contract Administration in Public Labor Relations / 1198
Market Failure, Theory of / 1203
Mathematical Modeling / 1207
Media and Bureaucracy in the United States / 1210
Medicaid Reforms in the States, Administration of / 1214
Merit System / 1219
Metropolitan Governance / 1221
Metropolitan Governmental Fragmentation / 1225
Microenterprise Finance and Economic Development / 1230
Middle East and Regulation / 1234
Middle East: Corruption / 1239
Milgram Experiments / 1243
Mitigating HIV/AIDS' Impact on Teachers and Administrators in Sub-Saharan Africa / 1247
Model Contracts to Reduce the Risks in Complex Information Technology Procurements / 1252
Models of Spatial Finance and "Voting with Your Feet" / 1256
Models of the Policy Process / 1260
Models, Types and Attributes of / 1264
Moral Development Theory / 1266
Motivation / 1270
Motives and Incentives: Rethinking Budget Theory in a New Institutional Context / 1275
Municipal Securities / 1280
Narrative Inquiry / 1285
Narrative Policy Analysis / 1289
National Security Policy / 1293
Natural Resource Policy: Institutional Aspects / 1298
Naturalistic Decision Making / 1303
Net Assets and Fund Balance / 1307
Networks, Organizational / 1312
New Institutional Economics / 1316
New Public Management / 1321
New Public Service / 1326
New Sciences Approach and Public Administration / 1331
Newtonian Paradigm / 1336
No Child Left Behind Act of 2001 / 1339
Non-Monetary Incentives and Productivity / 1346
Non-Profit Organizations, External Relations in / 1351
Nuremberg Trials and Charter / 1356
Offset Requirements in International Public Procurement / 1361
Ombuds and Ombuds Programs / 1366
Open Systems Theory / 1369
Operations Research / 1374
Orange County Bankruptcy / 1377
Organization Development / 1383
Organization Theories, Types of / 1387
Organizational Citizenship Behavior / 1392
Organizational Commitment / 1396
Organizational Culture / 1402

Volume 3

Papers on the Science of Administration / 1407
Parental Rights and Obligations / 1409
Participatory Policy Analysis / 1413
Patronage and Spoils / 1418
Pay for Performance and Merit Pay / 1423
Pay-As-You-Go Financing / 1428
Pension Under-Funding: Generational Accounting Perspective / 1432
Performance Appraisal / 1437
Performance Audits / 1442
Performance Indicators in State Administration, The Use of / 1447
Performance Measurement / 1450
Performance Measurement and Benchmarking in Local Government / 1455
Performance Measurement: Citizen-Driven / 1459
Performance Reporting / 1465
Performance-Based Budgeting / 1469

Volume 3 (Continued)

Policy Change / 1473
Policy Design / 1477
Policy Diffusion / 1482
Policy Implementation / 1487
Policy Networks / 1491
Policy Sciences Approach / 1495
Policy Termination / 1499
Policymaking in Local Government / 1503
Policy-Oriented Social Science Research: Measuring the Benefits / 1508
Popular Reporting / 1516
Portfolio Theory / 1519
Postmodernism and Public Policy / 1524
Postpositivist Perspectives in Policy Analysis / 1530
Practical Organization Theory / 1534
Pragmatism and Public Administration / 1538
Principal-Agent Perspective / 1542
Privatization / 1546
Privatization and Economic Development / 1549
Privatization and Public-Private Partnerships for Local Services / 1556
Privatization and Reinvention Paradigms for State Governmental Administration, Consequences of / 1561
Probability Theory / 1566
Procurement / 1571
Productivity and Effectiveness / 1575
Program Budgeting (PPBS) / 1579
Program Evaluation and Review Technique (PERT) and Critical Path Method (CPM) / 1583
Public Administration as Management / 1588
Public Administration, A Twenty-First Century Lens of Understanding / 1595
Public Authorities / 1601
Public Budgeting, Politics in / 1605
Public Choice / 1614
Public Choice Theory / 1618
Public Funds Investment Strategies / 1623
Public Goods / 1628
Public Policy / 1632
Public Policy Formation Under Conditions of Ambiguity / 1634
Public Procurement / 1642
Public Procurement Ethics / 1647
Public Reporting / 1652
Public Sector Productivity: Barriers / 1657
Public Service / 1661
Public-Private Partnerships for Economic Development / 1664
Public-Private Partnerships in Developing Countries / 1669
Punctuated Equilibrium / 1674
Q Methodology / 1678
Quantitative Decision Making: Benefit-Cost Analysis / 1682
Reciprocal Relations Among Peace, Prosperity, and Democracy / 1688
Reduction in Force / 1692
Regulatory Policy: Role and Importance / 1694
Representative Bureaucracy / 1698
Restorative Justice / 1702
Revenue Forecasting and Revenue Estimation / 1709
Risk Management / 1717
Risk Pools / 1721
Rulemaking / 1727
Rural Economies, Developing / 1732
Rural Health Policy / 1738
Science of Administration / 1742
Seasonality / 1746
Senior Executive Service / 1757
Seniority / 1762
Sexual Harassment / 1765
Singapore / 1773
Singapore: Economic Policy / 1778
Social Networks / 1784
Special District Governments / 1788
Stage Heuristics in Policy Sciences / 1795

Volume 3 (Continued)

Standard of Award / 1799
Standards of Conduct / 1802
State Administration of Mental Retardation and Developmental Disabilities Policies / 1806
State and Local Public Pension Fund Management / 1811
State Enterprise Zones / 1816
State Government Administration in a Federal System / 1823
State Government Administration of Rural Development Policy / 1827
State Government Administration, Impacts of Reform Movements / 1832
State Government Adoption of "Managing for Results" Reforms / 1838
State Government Policy Initiatives: Improving Access for the Medically Underserved / 1843
State Government, Aspects of Managing Strategic Change in / 1847
State Government: Complex Policy Network Management / 1853
State Government: Promotion of Innovations / 1858
State Legislative Fiscal Offices and State Government Administration / 1864
Strategic Management and Productivity / 1867
Subnational Counter-Cyclical Fiscal Policy in the United States / 1875
Supplemental Appropriation / 1880
Sustainable Development / 1884
Tax Expenditures / 1887
Tax Incidence / 1893
Team Building / 1899
Telehealth and State Government Policy / 1904
Teleology and Utilitarianism / 1908
Thailand / 1911
Theory, Types and Attributes of / 1916
Third-Party Interventions / 1920
Time Value of Money / 1923
Times Series Analysis / 1931
Trade Policy / 1936
Training and Development / 1942
Training for Productivity / 1949
Transparency and Corruption in Southeast Asia / 1955
Transportation Policy / 1960
Trust in Organizations / 1965
Truth and Reconciliation Commission / 1969
Truthfulness and Lying / 1972
Tuskegee Study / 1977
U.S. Office of Personnel Management / 1981
Uncertainty in Decision Making / 1984
Uncertainty, Concept of / 1988
Understanding the Basics of Refunding in the Municipal Bond Market / 1991
Uniform Guidelines on Employee Selection Procedures / 1995
Unit Determination / 1998
United States Treasury Securities / 2001
Urban Planning and Ethics / 2006
Use and Management of Information Technology in State Government / 2012
Using Econometric Models for Public-Sector Decision Making / 2017
Validation / 2024
Values and Policy Analysis / 2026
Virtue Ethics / 2028
Watergate / 2033
Welfare and Social Security Policy / 2041
Welfare Economics / 2047
Welfare Reform, Management of / 2054
Welfare State / 2060
Whistleblowing, Corporate, and Public Policy / 2066
Win-Win Approaches to Domestic Peace, Prosperity, and Democracy / 2070
Win-Win Approaches to International Peace, Prosperity, and Democracy / 2076
Win-Win Policy / 2083
Work Motivation / 2089
Working Capital / 2093
Yield Curve for Public-Sector Debt Instruments / 2097
Zero-Base Budgeting / 2102

In Memory of Jack M. Rabin
1945–2006

Encyclopedia of Public Administration and Public Policy, Second Edition

Evan M. Berman, Editor
Jack Rabin, Founding Editor

Editorial Advisory Board

Julia Beckett

Department of Public Administration and Urban Studies, University of Akron, Akron, Ohio, U.S.A.

Terry F. Buss

Center for International Affairs, National Academy of Public Administration, Washington, District of Columbia, U.S.A.

Hon S. Chan

Department of Public & Social Administration, City University of Hong Kong, Kowloon, Hong Kong

John L. Daly

Department of Government & International Affairs, University of South Florida, Tampa, Florida, U.S.A.

Ali Farazmand

School of Public Administration, Florida Atlantic University, Fort Lauderdale, Florida, U.S.A.

Gerald T. Gabris

Division of Public Administration, Northern Illinois University, DeKalb, Illinois, U.S.A.

John J. Gargan

Political Science Department, Kent State University, Kent, Ohio, U.S.A.

James A. Gazell

School of Public Administration and Urban Studies, San Diego State University, San Diego, California, U.S.A.

Anna C. Goldoff

Department of Public Management, John Jay College of Criminal Justice, City University of New York, New York, New York, U.S.A.

W. Bartley Hildreth

Hugo Wall School of Urban and Public Affairs and the W. Frank Barton School of Business, Wichita State University, Wichita, Kansas, U.S.A.

Marc Holzer

Department of Public Administration, Rutgers University, Newark, New Jersey, U.S.A.

Michael Howell-Moroney

Department of Government, University of Alabama at Birmingham, Birmingham, Alabama, U.S.A.

Ahmed Shafiqul Huque

Department of Political Science, McMaster University, Hamilton, Ontario, Canada

Douglas Ihrke

Department of Political Science and Urban Studies Program, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, U.S.A.

Richard C. Kearney

School of Public and International Affairs, North Carolina University, Raleigh, North Carolina, U.S.A.

Jack Kem

Department of Joint and Multinational Operations, U.S. Army Command and General Staff College, Fort Leavenworth, Kansas, U.S.A.

Anne Osborne Kilpatrick

Department of Health Administration and Policy, Medical University of South Carolina, Charleston, South Carolina, U.S.A.

Carol W. Lewis

Department of Political Science, University of Connecticut, Storrs, Connecticut, U.S.A.

Gerald J. Miller

Department of Public Administration, Rutgers University, Newark, New Jersey, U.S.A.

Göktuğ Morçöl

School of Public Affairs, Pennsylvania State University at Harrisburg, Middletown, Pennsylvania, U.S.A.

Ricardo S. Morse

School of Government, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, U.S.A.

Bruce J. Perlman

School of Public Administration, University of New Mexico, Albuquerque, New Mexico, U.S.A.

Donijo Robbins

School of Public & Nonprofit Administration Grand Valley State University, Grand Rapids, Michigan, U.S.A.

James Ruiz

School of Public Affairs, Pennsylvania State University at Harrisburg, Middletown, Pennsylvania, U.S.A.

Barbara Sims

School of Public Affairs, Pennsylvania State University at Harrisburg, Middletown, Pennsylvania, U.S.A.

Kenneth A. Smith

Atkinson Graduate School of Management, Willamette University, Salem, Oregon, U.S.A.

Joseph L. Staats

Department of Political Science, Texas Tech University, Lubbock, Texas, U.S.A.

Khi V. Thai

School of Public Administration, Florida Atlantic University, Fort Lauderdale, Florida, U.S.A.

Eran Vigoda-Gadot

School of Political Sciences, University of Haifa, Haifa, Israel

Jeffrey A. Weber

Department of Political University, East Stroudsburg University, East Stroudsburg, Pennsylvania, U.S.A.

Contributors

- Brian E. Adams** / *Department of Political Science, San Diego State University, San Diego, California, U.S.A.*
- Stephen C. Adams** / *Office of Strategic Projects and Planning, Florida Department of Environmental Protection, Tallahassee, Florida, U.S.A.*
- John Adler** / *Department of Transportation, Arizona Department of Administration, Phoenix, Arizona, U.S.A.*
- Troels Pind Adrian** / *School of Public Affairs, Baruch College, City University of New York, New York, New York, U.S.A.*
- Robert Agranoff** / *School of Public and Environmental Affairs, Indiana University, Bloomington, Indiana, U.S.A.*
- William G. Albrecht** / *Department of Political Science and Public Administration, University of North Carolina at Pembroke, Pembroke, North Carolina, U.S.A.*
- Patricia M. Alt** / *Department of Health Science, Towson University, Towson, Maryland, U.S.A.*
- David N. Ammons** / *Institute of Government, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, U.S.A.*
- Ori Arbel-Ganz** / *Department of Political Science, Bar-Ilan University, Ramat-Gan, Israel*
- Michael Asner** / *Michael Asner Consulting, Surrey, British Columbia, Canada*
- Michael W. Austin** / *Department of Philosophy and Religion, Eastern Kentucky University, Richmond, Kentucky, U.S.A.*
- A. Hunter Bacot** / *Department of Political Science, University of North Carolina at Charlotte, Charlotte, North Carolina, U.S.A.*
- David L. Baker** / *School of Public Affairs, Arizona State University, Tempe, Arizona, U.S.A.*
- Larry Bakken** / *School of Law, Hamline University, St. Paul, Minnesota, U.S.A.*
- Daryl Balia** / *Professional Ethics and Risk Management, Public Service Commission, Pretoria, South Africa*
- Irene J. Barnett** / *Department of Political Science, Kent State University, Kent, Ohio, U.S.A.*
- John R. Bartle** / *School of Public Administration, University of Nebraska at Omaha, Omaha, Nebraska, U.S.A. and Center for Public Administration, Sun Yat-Sen University, Guangzhou, Guangdong, P.R. China*
- C. Fred Baughman** / *Department of Public Administration, University of Central Florida–Daytona Beach, Daytona Beach, Florida, U.S.A.*
- Roger J. Beck** / *Department of Agribusiness Economics, Southern Illinois University Carbondale, Carbondale, Illinois, U.S.A.*
- Julia Beckett** / *Department of Public Administration and Urban Studies, University of Akron, Akron, Ohio, U.S.A.*
- Jane Beckett-Camarata** / *Department of Political Science, Kent State University, Kent, Ohio, U.S.A.*
- Christina M. Bellon** / *Department of Philosophy, California State University, Sacramento, Sacramento, California, U.S.A.*
- J. Edwin Benton** / *Department of Government and International Affairs, University of South Florida, Tampa, Florida, U.S.A.*
- Evan M. Berman** / *Public Administration Institute, Louisiana State University, Baton Rouge, Louisiana, U.S.A.*
- Curtis R. Berry** / *Department of Political Science, Shippensburg University, Shippensburg, Pennsylvania, U.S.A.*

- Bradley J. Best** / *Department of Political Science, Buena Vista University, Storm Lake, Iowa, U.S.A.*
- Lisa B. Bingham** / *School of Public and Environmental Affairs, Indiana University, Bloomington, Indiana, U.S.A.*
- Thomas A. Birkland** / *Rockefeller College of Public Affairs and Policy, University at Albany, State University of New York, Albany, New York, U.S.A.*
- Frayda S. Bluestein** / *School of Government, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, U.S.A.*
- Keith Boeckelman** / *Department of Political Science, Western Illinois University, Macomb, Illinois, U.S.A.*
- Amiram Bogot** / *Former Registrar, Non-Profit Organizations, Jerusalem, Israel*
- David E. Booher** / *Center for Collaborative Policy, California State University Sacramento, Davis, California, U.S.A.*
- R. Wayne Boss** / *Leeds School of Business, University of Colorado at Boulder, Boulder, Colorado, U.S.A.*
- Deborah A. Botch** / *New York State Unified Court System, Albany, New York, U.S.A.*
- Cynthia J. Bowling** / *Department of Political Science, Auburn University, Auburn, Alabama, U.S.A.*
- Ann O'M. Bowman** / *Department of Government and International Studies, University of South Carolina, Columbia, South Carolina, U.S.A.*
- James S. Bowman** / *Askew School of Public Administration, Florida State University, Tallahassee, Florida, U.S.A.*
- Neil M. Boyd** / *School of Education and Behavioral Sciences, Pennsylvania State University, Capital College, Middletown, Pennsylvania, U.S.A.*
- Dana Burr Bradley** / *Department of Political Science and Gerontology, University of North Carolina at Charlotte, Charlotte, North Carolina, U.S.A.*
- Paul J. Brennan** / *Department of Accounting, Austin Peay State University, Clarksville, Tennessee, U.S.A.*
- Brian Brewer** / *Department of Public and Social Administration, City University of Hong Kong, Kowloon, Hong Kong*
- Gene A. Brewer** / *School of Public and International Affairs, University of Georgia, Athens, Georgia, U.S.A.*
- Stacey M. Brewer** / *Hugo Wall School of Urban and Public Affairs, Wichita State University, Wichita, Kansas, U.S.A.*
- Andrew Brown** / *School of Social Science, University of New England, Armidale, New South Wales, Australia*
- Judith R. Brown** / *Planning, Research and Evaluation Division, Governor's Office of Planning and Budget, Atlanta, Georgia, U.S.A.*
- Mary Maureen Brown** / *Department of Political Science, University of North Carolina at Charlotte, Charlotte, North Carolina, U.S.A.*
- Raymond Brown** / *University of Oregon, Eugene, Oregon, U.S.A.*
- William P. Browne** / *Department of Political Science, Central Michigan University, Mount Pleasant, Michigan, U.S.A.*
- James A. Buford Jr.** / *Ellis-Harper Management, Auburn, Alabama, U.S.A.*
- Beverly S. Bunch** / *Director of Public Administration Program, PAC 324, University of Illinois at Springfield, Springfield, Illinois, U.S.A.*
- Brendan F. Burke** / *Department of Political Science, Bridgewater State College, Bridgewater, Massachusetts, U.S.A.*
- DeLysa Burnier** / *Department of Political Science, Ohio University, Athens, Ohio, U.S.A.*
- Terry F. Buss** / *Center for International Affairs, National Academy of Public Administration, Washington, District of Columbia, U.S.A.*
- Thomas Buttross** / *Department of Accounting, Pennsylvania State University at Harrisburg, Middletown, Pennsylvania, U.S.A.*

- Kathe Callahan** / *School of Public Affairs and Administration, Rutgers University, Newark, New Jersey, U.S.A.*
- Ledivina V. Cariño** / *National College of Public Administration and Governance, University of the Philippines, Quezon City, Philippines*
- Abraham Carmeli** / *Graduate School of Business Administration, Bar-Ilan University, Ramat-Gan, Israel*
- David G. Carnevale** / *Department of Political Science, University of Oklahoma, Norman, Oklahoma, U.S.A.*
- Jered B. Carr** / *Department of Political Science, Wayne State University, Detroit, Michigan, U.S.A.*
- Keith Carrington** / *Criminal Justice and Legal Studies Division, St. John's University, Jamaica, New York, U.S.A.*
- Paul J. Castellani** / *Department of Public Administration and Policy, University at Albany, State University of New York, Albany, New York, U.S.A.*
- Thomas J. Catlaw** / *School of Public Affairs, Arizona State University, Phoenix, Arizona, U.S.A.*
- Anita Chadha** / *Department of Social Sciences, University of Houston - Downtown, Houston, Texas, U.S.A.*
- Paul M. Chalekian** / *Department of Political Science, University of Nevada, Reno, Carson City, Nevada, U.S.A.*
- James L. Chan** / *Department of Accounting, University of Illinois at Chicago, Chicago, Illinois, U.S.A.*
- Piper S. Charles** / *Department of Political Science, University of North Carolina at Charlotte, Charlotte, North Carolina, U.S.A.*
- D. S. Chauhan** / *Department of Political Science, Bowling Green State University, Bowling Green, Ohio, U.S.A.*
- Greg Chen** / *School of Public Affairs, Baruch College, City University of New York, New York, New York, U.S.A.*
- Keon S. Chi** / *Center for State Governance Transformation, Council of State Governments, Lexington, Kentucky, U.S.A.*
- Rupert F. Chisholm** / *Department of Management, Pennsylvania State University at Harrisburg, Middletown, Pennsylvania, U.S.A.*
- Robert K. Christensen** / *Political Science Department, University of North Carolina at Charlotte, Charlotte, North Carolina, U.S.A.*
- Cristina E. Ciocirlan** / *Department of Public Administration, Pennsylvania State University at Harrisburg, Middletown, Pennsylvania, U.S.A.*
- Cal Clark** / *Department of Political Science, Auburn University, Auburn, Alabama, U.S.A.*
- Jill Clark** / *Department of Political Science, University of Texas at Arlington, Arlington, Texas, U.S.A.*
- Joy A. Clay** / *School of Urban Affairs and Public Policy, The University of Memphis, Memphis, Tennessee, U.S.A.*
- Charles K. Coe** / *Department of Political Science and Public Administration, North Carolina State University, Raleigh, North Carolina, U.S.A.*
- Jerrell D. Cogburn** / *Department of Public Administration, University of Texas at San Antonio, San Antonio, Texas, U.S.A.*
- Richard K. Common** / *The Business School, University of Hull, Hull, U.K.*
- Stephen E. Condrey** / *Carl Vinson Institute of Government, University of Georgia, Athens, Georgia, U.S.A.*
- Charles Conteh** / *Department of Political Science, McMaster University, Hamilton, Ontario, Canada*
- Alethia H. Cook** / *Department of Political Science, University of Akron, Akron, Ohio, U.S.A.*
- Peter L. Cruise** / *Department of Health and Community Services, California State University, Chico, California, U.S.A.*
- Dennis M. Daley** / *Political Science and Public Administration, North Carolina State University, Raleigh, North Carolina, U.S.A.*
- Glenn A. Daley** / *RAND Graduate School, Santa Monica, California, U.S.A.*

- Mark R. Daniels** / *Department of Government and Public Affairs, Slippery Rock University of Pennsylvania, Slippery Rock, Pennsylvania, U.S.A.*
- Patria de Lancer Julnes** / *Department of Political Science, Utah State University, Logan, Utah, U.S.A.*
- Ilka M. Decker** / *Carl Vinson Institute of Government, University of Georgia, Athens, Georgia, U.S.A.*
- Peter deLeon** / *Graduate School of Public Affairs, University of Colorado at Denver and Health Sciences Center, Denver, Colorado, U.S.A.*
- Dwight V. Denison** / *Robert F. Wagner Graduate School of Public Service, New York University, New York, New York, U.S.A.*
- Lynda M. Dennis** / *Public Administration Department, University of Central Florida, Orlando, Florida, U.S.A.*
- Jeff Leon Dense** / *Department of Political Science, Eastern Oregon University, La Grande, Oregon, U.S.A.*
- Victor S. DeSantis** / *Department of Political Science, Bridgewater State College, Bridgewater, Massachusetts, U.S.A.*
- Krishna S. Dhir** / *Campbell School of Business, Berry College, Mount Berry, Georgia, U.S.A.*
- Laurent Dobuzinskis** / *Department of Political Science, Simon Fraser University, Burnaby, British Columbia, Canada*
- Jennifer Dodge** / *Robert F. Wagner Graduate School of Public Service, New York University, New York, New York, U.S.A.*
- Brian E. Dollery** / *School of Economics, University of New England, Armidale, New South Wales, Australia*
- Amy K. Donahue** / *Institute of Public and Urban Affairs and Department of Political Science, University of Connecticut, Storrs, Connecticut, U.S.A.*
- Melvin J. Dubnick** / *Departments of Political Science and Public Administration, Rutgers University, Newark, New Jersey, U.S.A.*
- Larkin S. Dudley** / *Center for Public Administration and Policy, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, U.S.A.*
- William Duncombe** / *Department of Public Administration, Syracuse University, Syracuse, New York, U.S.A.*
- Robert F. Durant** / *School of Public Affairs, Department of Public Administration and Policy Analysis, American University, Washington, District of Columbia, U.S.A.*
- Dan Durning** / *Carl Vinson Institute of Government, University of Georgia, Athens, Georgia, U.S.A.*
- R. Gregg Dwyer** / *Department of Neuropsychiatry and Behavioral Science, University of South Carolina School of Medicine, Columbia, South Carolina, U.S.A.*
- Carol Ebdon** / *Department of Public Administration, University of Nebraska at Omaha, Omaha, Nebraska, U.S.A.*
- Robert J. Eger** / *Public Administration and Urban Studies Department, Georgia State University, Atlanta, Georgia, U.S.A.*
- Virginia Eubanks** / *Department of Women's Studies, University at Albany, State University of New York, Albany, New York, U.S.A.*
- Richard C. Feiock** / *Askew School of Public Administration and Policy, Florida State University, Tallahassee, Florida, U.S.A.*
- Patrick Fisher** / *Department of Political Science/Philosophy, Monmouth University, West Long Branch, New Jersey, U.S.A.*
- H. Edward Flentje** / *Hugo Wall School of Urban and Public Affairs, Wichita State University, Wichita, Kansas, U.S.A.*
- Esther M. Forti** / *Department of Health Professions, Medical University of South Carolina, Charleston, South Carolina, U.S.A.*
- Scott Fritzen** / *Public Policy Programme, Lee Kuan Yew School of Public Policy, National University of Singapore, Singapore*

- Scott R. Furlong** / *Public and Environmental Affairs, University of Wisconsin-Green Bay, Green Bay, Wisconsin, U.S.A.*
- Gerald T. Gabris** / *Division of Public Administration, Northern Illinois University, DeKalb, Illinois, U.S.A.*
- James A. Gazell** / *School of Public Administration and Urban Studies, San Diego State University, San Diego, California, U.S.A.*
- Stuart C. Gilman** / *Ethics Resource Center, Washington, District of Columbia, U.S.A.*
- Anna C. Goldoff** / *Department of Public Management, John Jay College of Criminal Justice, New York, New York, U.S.A.*
- Robert T. Golembiewski** / *Department of Political Science, University of Georgia, Athens, Georgia, U.S.A.*
- Ray Gonzales** / *Graduate Department of Public Administration, Rutgers University, Newark, New Jersey, U.S.A.*
- Cole Blease Graham Jr.** / *Department of Government and International Studies, University of South Carolina, Columbia, South Carolina, U.S.A.*
- Lawrence S. Graham** / *Department of Government, University of Texas, Austin, Texas, U.S.A.*
- Roger Green** / *Division of Public Affairs, Florida Gulf Coast University, Fort Myers, Florida, U.S.A.*
- Ian Greene** / *Department of Political Science, York University, Toronto, Ontario, Canada*
- Daphne Greenwood** / *Department of Economics, University of Colorado at Colorado Springs, Colorado Springs, Colorado, U.S.A.*
- Thomas Greitens** / *American Farmland Trust, Center for Agriculture in the Environment, Northern Illinois University, DeKalb, Illinois, U.S.A.*
- George M. Guess** / *Open Society Institute, Local Government and Public Service Reform Initiative, Budapest, Hungary*
- A. Gunasekaran** / *Department of Management, University of Massachusetts Dartmouth, North Dartmouth, Massachusetts, U.S.A.*
- Krishan M. Gupta** / *Department of Accounting and Finance, University of Massachusetts Dartmouth, North Dartmouth, Massachusetts, U.S.A.*
- James S. Guseh** / *Department of Public Administration, North Carolina Central University, Durham, North Carolina, U.S.A.*
- Andrew Haas** / *Public Administration and Urban Studies Department, Georgia State University, Atlanta, Georgia, U.S.A.*
- Peter J. Haas** / *Department of Political Science, San Jose State University, San Jose, California, U.S.A.*
- Darrene L. Hackler** / *Department of Public and International Affairs, George Mason University, Fairfax, Virginia, U.S.A.*
- Arie Halachmi** / *Institute of Government, Tennessee State University, Nashville, Tennessee, U.S.A. and Zhongshan University, Guangzhou, P.R. China*
- Kathleen Hale** / *Department of Political Science, Auburn University, Auburn, Alabama, U.S.A.*
- Jeremy L. Hall** / *Department of Government, University of Alabama at Birmingham, Birmingham, Alabama, U.S.A.*
- Jay P. Hamilton** / *Department of Public Management, John Jay College of Criminal Justice, New York, New York, U.S.A.*
- Dimitri R. Hanna** / *Department of International Affairs and Diplomacy, Notre Dame University, Zouk Mosbeh, Lebanon*
- Akhlaque Haque** / *Department of Government, University of Alabama at Birmingham, Birmingham, Alabama, U.S.A.*
- M. Shamsul Haque** / *Department of Political Science, National University of Singapore, Singapore*
- Michael A. Harper** / *University of Arkansas, Fayetteville, Arkansas, U.S.A.*
- G.L.A. Harris** / *Mark O. Hatfield School of Government, Portland State University, Portland, Oregon, U.S.A.*
- Jean Harris** / *Department of Accounting, Pennsylvania State University at Harrisburg, Middletown, Pennsylvania, U.S.A.*

- John Harris** / *Department of Philosophy, University of Colorado at Boulder, Boulder, Colorado, U.S.A.*
- Wendy Haynes** / *Bridgewater State College, Bridgewater, Massachusetts, U.S.A.*
- Steven W. Hays** / *Department of Government and International Studies, University of South Carolina, Columbia, South Carolina, U.S.A.*
- Jiang He** / *Public Policy Program, University of North Carolina at Charlotte, Charlotte, North Carolina, U.S.A.*
- Robert Hebdon** / *Faculty of Management, McGill University, Montreal, Québec, Canada*
- James R. Heichelbech** / *Graduate School of Public Affairs, University of Colorado at Denver, Denver, Colorado, U.S.A.*
- April Hejka-Ekins** / *Department of Politics and Public Administration, California State University, Stanislaus, Turlock, California, U.S.A.*
- Craig Hemmens** / *Department of Criminal Justice Administration, Boise State University, Boise, Idaho, U.S.A.*
- J. J. Hendricks** / *Department of Politics and Public Administration, California State University, Stanislaus, Turlock, California, U.S.A.*
- Arie Hershcovich** / *Department of Political Science, Yezreel Valley College, Emek Yezreel, Israel*
- Frederick M. Hess** / *Virginia Center for Educational Policy Studies, University of Virginia, Charlottesville, Virginia, U.S.A.*
- Lyn Holley** / *Department of Gerontology, University of Nebraska at Omaha, Omaha, Nebraska, U.S.A.*
- Marc Holzer** / *Croton-on-Hudson, New York, U.S.A.*
- Annie Hondeghem** / *Katholieke Universiteit Leuven, Leuven, Belgium*
- Sung Gul Hong** / *Department of Public Administration, Kookmin University, Seoul, Korea*
- Yilin Hou** / *Department of Public Administration and Policy, University of Georgia, Athens, Georgia, U.S.A.*
- Leo W.J.C. Huberts** / *Department of Public Administration and Communication Science, Vrije Universiteit, Amsterdam, The Netherlands*
- John K. Hudzik** / *Global Engagement and Strategic Projects, Michigan State University, East Lansing, Michigan, U.S.A.*
- Ronald John Hy** / *College of Arts and Sciences, Texas A&M University-Kingsville, Kingsville, Texas, U.S.A.*
- Douglas Ihrke** / *Department of Political Science, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin, U.S.A.*
- Iryna Illiash** / *Graduate Department of Public Administration, Rutgers University, Newark, New Jersey, U.S.A.*
- Helen Ingram** / *School of Social Ecology, University of California, Irvine, California, U.S.A.*
- Hedy L. Isaacs** / *Department of Government, University of the West Indies, Kingston, Jamaica*
- Martin Ives** / *R.F. Wagner Graduate School of Public Service, New York University, New York, New York, U.S.A.*
- Edward T. Jennings Jr.** / *Department of Political Science, University of Kentucky, Lexington, Kentucky, U.S.A.*
- Courtney E. Jensen** / *Center for Public Administration and Policy, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, U.S.A.*
- Yijia Jing** / *Department of Public Administration, Fudan University, Shanghai, P.R. China*
- Gary A. Johnson** / *University of North Carolina, Charlotte, Charlotte, North Carolina, U.S.A.*
- David Seth Jones** / *Department of Public Policy, University of Brunei Darussalam, Gadong, Brunei*
- L. R. Jones** / *Graduate School of Business and Public Policy, Naval Postgraduate School, Monterey, California, U.S.A.*
- Walter J. Jones** / *Medical University of South Carolina, Charleston, South Carolina, U.S.A.*
- Carole L. Jurkiewicz** / *Public Administration Institute, Louisiana State University, Baton Rouge, Louisiana, U.S.A.*
- Jonathan B. Justice** / *School of Urban Affairs and Public Policy, University of Delaware, Newark, Delaware, U.S.A.*

- Hwang-Sun Kang** / *Department of Urban Management, Seoul Development Institute, Seoul, Republic of Korea*
- Lawrence F. Keller** / *Public Administration Program, Cleveland State University, Cleveland, Ohio, U.S.A.*
- Michael A. Kelley** / *Department of Social Sciences, Pittsburg State University, Pittsburg, Kansas, U.S.A.*
- J. Edward Kellough** / *School of Public and International Affairs, University of Georgia, Athens, Georgia, U.S.A.*
- Thomas W. Kent** / *Department of Management and Entrepreneurship, College of Charleston, Charleston, South Carolina, U.S.A.*
- Cornelius M. Kerwin** / *American University, Washington, District of Columbia, U.S.A.*
- Mozaffar Khan** / *University of Toronto, Toronto, Ontario, Canada*
- Peter J. Kiedrowski** / *Institutional Investments, Wells Fargo, Minneapolis, Minnesota, U.S.A.*
- Bruce W. Kieler** / *Wharton County Junior College, Wharton, Texas, U.S.A.*
- Henry E. Kilpatrick** / *Econpolicy, and School of Public Policy, George Mason University, Arlington, Fairfax, Virginia, U.S.A.*
- Jeongwoo Kim** / *School of Public Affairs, Arizona State University, Phoenix, Arizona, U.S.A.*
- Seok-Eun Kim** / *Department of Political Science, Kansas State University, Manhattan, Kansas, U.S.A.*
- Paul B. Kirby** / *University of Massachusetts Medical School, Center for Health Policy and Research, University of Massachusetts, Worcester, Massachusetts, U.S.A.*
- Judith J. Kirchhoff** / *MPA Program–SBPAIS, Long Island University, University Plaza, Brooklyn, New York, U.S.A.*
- Kenneth A. Klase** / *Department of Political Science, University of North Carolina at Greensboro, Greensboro, North Carolina, U.S.A.*
- William Earle Klay** / *Askew School of Public Administration and Policy, Florida State University, Tallahassee, Florida, U.S.A.*
- Donald E. Klingner** / *Graduate School of Public Affairs, University of Colorado at Colorado Springs, Colorado Springs, Colorado, U.S.A.*
- Paul J.M. Klumpes** / *Tanaka Business School, Imperial College London, London, U.K.*
- Yvonne J. Kochanowski** / *SteelEdge Business Consulting, Placerville, California, U.S.A.*
- Steven G. Koven** / *School of Urban and Public Affairs, University of Louisville, Louisville, Kentucky, U.S.A.*
- Dale Krane** / *Department of Public Administration, University of Nebraska at Omaha, Omaha, Nebraska, U.S.A.*
- Robert S. Kravchuk** / *Indiana University, Bloomington, Indiana, U.S.A.*
- Kenneth A. Kriz** / *School of Public Administration, University of Nebraska at Omaha, Omaha, Nebraska, U.S.A.*
- Robert Krol** / *Department of Economics, California State University, Northridge, California, U.S.A.*
- Claus Langfred** / *John M. Olin School of Business, Washington University, St. Louis, Missouri, U.S.A.*
- Deborah L. Laufersweiler-Dwyer** / *Department of Criminal Justice, University of Arkansas at Little Rock, Little Rock, Arkansas, U.S.A.*
- Wendell C. Lawther** / *Department of Public Administration, University of Central Florida, Orlando, Florida, U.S.A.*
- Richard L.B. LeCompte** / *Department of Finance, Real Estate and Decision Sciences, Wichita State University, Wichita, Kansas, U.S.A.*
- Mordecai Lee** / *Department of Governmental Affairs, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, U.S.A.*
- Seok-Hwan Lee** / *Department of Public Administration, Catholic University of Korea, Puchon City, Republic of Korea*
- Suzanne Leland** / *Department of Political Science, University of North Carolina at Charlotte, Charlotte, North Carolina, U.S.A.*
- Eric K. Leonard** / *Shenandoah University, Winchester, Virginia, U.S.A.*

- Carol W. Lewis** / *Department of Political Science, University of Connecticut, Storrs, Connecticut, U.S.A.*
- Jeremy R.T. Lewis** / *Department of Political Science, Huntingdon College, Montgomery, Alabama, U.S.A.*
- Kuotsai Tom Liou** / *Department of Public Administration, University of Central Florida, Orlando, Florida, U.S.A.*
- Robert E. Lloyd** / *U.S. Department of State, Ft. Lauderdale, Florida, U.S.A.*
- David J. Louscher** / *Department of Political Science, University of Akron, Akron, Ohio, U.S.A.*
- Richard A. Loverd** / *Acton, Massachusetts, U.S.A.*
- Bill Luker** / *Department of Economics, University of North Texas, Denton, Texas, U.S.A.*
- Craig Lundberg** / *School of Hotel Administration, Cornell University, Ithaca, New York, U.S.A.*
- Cynthia E. Lynch** / *Nelson Mandela School of Public Policy and Urban Affairs, Southern University, Baton Rouge, Louisiana, U.S.A.*
- Thomas D. Lynch** / *Public Administrations Institute, Louisiana State University, Baton Rouge, Louisiana, U.S.A.*
- Susan MacDonald** / *California State University, Stanislaus, Turlock, California, U.S.A.*
- Jeroen Maesschalck** / *Instituut voor de overheid, Katholieke Universiteit Leuven, Leuven, Belgium*
- Michelle Maiese** / *University of Colorado at Boulder, Boulder, Colorado, U.S.A.*
- Theo Edwin Maloy** / *Department of History and Political Science, West Texas A&M University, Canyon, Texas, U.S.A.*
- Nancy Mantell** / *Rutgers University, New Brunswick, New Jersey, U.S.A.*
- Justin Marlowe** / *Department of Public Administration, University of Kansas, Lawrence, Kansas, U.S.A.*
- Gary S. Marshall** / *Department of Social Sciences, Roskilde University, Roskilde, Denmark*
- Christine R. Martell** / *Graduate School of Public Affairs, University of Colorado at Denver and Health Sciences Center, Denver, Colorado, U.S.A.*
- Betsy A. Matthews** / *Department of Correctional and Juvenile Justice Studies, Eastern Kentucky University, Richmond, Kentucky, U.S.A.*
- Darin Matthews** / *Multnomah County Schools, Portland, Oregon, U.S.A.*
- Amy G. Mazur** / *Department of Political Science, Washington State University, Pullman, Washington, U.S.A.*
- James McBain** / *Department of Social Sciences, Pittsburg State University, Pittsburg, Kansas, U.S.A.*
- Jerry McCaffery** / *Department of Systems Management, Naval Postgraduate School, Monterey, California, U.S.A.*
- Sam M. McCall** / *City Auditor, City of Tallahassee, Tallahassee, Florida, U.S.A.*
- Nancy McCarthy Snyder** / *Hugo Wall School of Urban and Public Affairs, Wichita State University, Wichita, Kansas, U.S.A.*
- Mary Ann McClure** / *Department of Art, Music and Philosophy, John Jay College of Criminal Justice, New York, New York, U.S.A.*
- Mark L. McConkie** / *Graduate School of Public Affairs, University of Colorado at Colorado Springs, Colorado Springs, Colorado, U.S.A.*
- Carl L. McCoy** / *Carl Vinson Institute of Government, University of Georgia, Athens, Georgia, U.S.A.*
- Michael McGuire** / *Department of Public Administration, University of North Texas, Denton, Texas, U.S.A.*
- Dean Michael Mead** / *Governmental Accounting Standards Board, Norwalk, Connecticut, U.S.A.*
- Timothy D. Mead** / *Department of Political Science, University of North Carolina at Charlotte, Charlotte, North Carolina, U.S.A.*
- William J. Mead** / *Consultant, White Plains, New York, U.S.A.*
- Jack W. Meek** / *College of Business and Public Management, University of La Verne, La Verne, California, U.S.A.*

- Julia E. Melkers** / *College of Urban Planning and Public Affairs, University of Illinois at Chicago, Chicago, Illinois, U.S.A.*
- Donald Menzel** / *Division of Public Administration, Northern Illinois University, DeKalb, Illinois, U.S.A.*
- Assaf Meydani** / *School of Government and Society, Academic College of Tel-Aviv Yaffo, Tel-Aviv, Israel*
- Michael E. Milakovich** / *Department of Political Science, University of Miami, Coral Gables, Florida, U.S.A.*
- Gerald J. Miller** / *Department of Public Administration, Rutgers University, Newark, New Jersey, U.S.A.*
- Patricia Moore** / *Public Administration Department, Kean University, Union, New Jersey, U.S.A.*
- Göktuğ Morçöl** / *School of Public Affairs, Pennsylvania State at Harrisburg, Middletown, Pennsylvania, U.S.A.*
- Karen Mossberger** / *Department of Political Science, Kent State University, Kent, Ohio, U.S.A.*
- Donald P. Moynihan** / *Robert M. La Follette School of Public Affairs, University of Wisconsin-Madison, Madison, Wisconsin, U.S.A.*
- Phillip A. Mundo** / *Political Science Department, Drew University, Madison, New Jersey, U.S.A.*
- Tina Nabatchi** / *School of Public and Environmental Affairs, Indiana University, Bloomington, Indiana, U.S.A.*
- Nitza Nachmias** / *School of Political Science, University of Haifa, Haifa, Israel*
- Katherine C. Naff** / *Public Administration Program, San Francisco State University, San Francisco, California, U.S.A.*
- Stuart S. Nagel** / *Political Science Department, University of Illinois, Urbana, Illinois, U.S.A.*
- Ramzi N. Nasser** / *Notre Dame University, Zouk Mosbeh, Lebanon*
- Doron Navot** / *Political Science Department, Hebrew University, Jerusalem, Israel*
- Brian Negin** / *Israel Central Bureau of Statistics, Jerusalem, Israel*
- Becky A. Nesbit** / *School of Public and Environmental Affairs, Indiana University, Bloomington, Indiana, U.S.A.*
- Thomas S. Nesslein** / *Urban and Regional Studies, Department of Economics, University of Wisconsin-Green Bay, Green Bay, Wisconsin, U.S.A.*
- David C. Nice** / *Department of Political Science, Washington State University, Pullman, Washington, U.S.A.*
- Anthony J. Nownes** / *Department of Political Science, University of Tennessee, Knoxville, Tennessee, U.S.A.*
- Paul C. Nutt** / *Fisher College of Business, Ohio State University, Columbus, Ohio, U.S.A.*
- Holona L. Ochs** / *Department of Political Science, University of Kansas, Lawrence, Kansas, U.S.A.*
- David Ogula** / *Valley Stream, New York, U.S.A.*
- Cheol H. Oh** / *Department of Public Administration, Soongsil University, Seoul, South Korea*
- Joleen R. Okun** / *University of Virginia, Charlottesville, Virginia, U.S.A.*
- Patrick W. Ong** / *Peace and Equity Foundation, Quezon City, Philippines*
- Sonia M. Ospina** / *Robert F. Wagner Graduate School of Public Service, New York University, New York, New York, U.S.A.*
- James F. O'Sullivan** / *Robert F. Wagner Graduate School of Public Service, New York University, New York, New York, U.S.A.*
- Mumukshu Patel** / *National Academy of Public Administration, Washington, District of Columbia, U.S.A.*
- W. David Patton** / *Center for Public Policy and Administration, Boise State University, Boise, Idaho, U.S.A.*
- Gregory Paxton** / *School of Public Policy and Management, Ohio State University, Columbus, Ohio, U.S.A.*
- Jun Peng** / *School of Public Administration and Public Policy, University of Arizona, Tucson, Arizona, U.S.A.*

- J. L. Peters** / *Department of Public Management, John Jay College of Criminal Justice, City University of New York, New York, New York, U.S.A.*
- Steven A. Peterson** / *School of Public Affairs, Pennsylvania State University at Harrisburg, Middletown, Pennsylvania, U.S.A.*
- Suzanne J. Piotrowski** / *Graduate Department of Public Administration, Rutgers University, Newark, New Jersey, U.S.A.*
- Jeremy F. Plant** / *School of Public Affairs, Pennsylvania State University at Harrisburg, Middletown, Pennsylvania, U.S.A.*
- Amy R. Poteete** / *Department of Political Science, Concordia University, Montreal, Quebec, Canada*
- A. Premchand** / *Consultant, Irvine, California, U.S.A.*
- Ann Pushkin** / *College of Business and Economics, West Virginia University, Morgantown, West Virginia, U.S.A.*
- Joan E. Pynes** / *Department of Government and International Affairs, University of South Florida, Tampa, Florida, U.S.A.*
- Jon S.T. Quah** / *Department of Political Science, National University of Singapore, Singapore*
- Hal G. Rainey** / *Department of Political Science, University of Georgia, Athens, Georgia, U.S.A.*
- M. Ramesh** / *Lee Kuan Yew School of Public Policy, National University of Singapore, Singapore*
- F. Stevens Redburn** / *U.S. Office of Management and Budget, Executive Office of the President, Washington, District of Columbia, U.S.A.*
- B. J. Reed** / *College of Public Affairs and Community Service, University of Nebraska, Omaha, Nebraska, U.S.A.*
- Christine M. Reed** / *School of Public Administration, University of Nebraska at Omaha, Omaha, Nebraska, U.S.A.*
- Peter J. Reed** / *Bannock Consulting, London, U.K.*
- Helena K. Rene** / *School of Public Affairs, American University, Washington, District of Columbia, U.S.A.*
- Tari Renner** / *Department of Political Science, Illinois Wesleyan University, Bloomington, Illinois, U.S.A.*
- Rupert G. Rhodd** / *Department of Economics, Florida Atlantic University, Davie, Florida, U.S.A.*
- Norma M. Riccucci** / *School of Public Affairs and Administration, Rutgers University, Newark Campus, Newark, New Jersey, U.S.A.*
- Donijo Robbins** / *School of Public and Nonprofit Administration, Grand Valley State University, Grand Rapids, Michigan, U.S.A.*
- Mark D. Robbins** / *University of Connecticut, Storrs, Connecticut, U.S.A.*
- Gary E. Roberts** / *School of Global Leadership and Entrepreneurship, Regent University, Virginia Beach, Virginia, U.S.A.*
- Julia E. Robinson** / *Robinson Enterprises, Inc., Boise, Idaho, U.S.A.*
- Julie W. Robinson** / *Veteran Affairs Medical Center (151C), Tuscaloosa Research and Education Advancement Corporation, Tuscaloosa, Alabama, U.S.A.*
- Scott E. Robinson** / *Department of Political Science and Public Affairs, University of Texas-Dallas, Richardson, Texas, U.S.A.*
- David A. Rochefort** / *Department of Political Science, Northeastern University, Boston, Massachusetts, U.S.A.*
- Emery Roe** / *Public Policy, Mills College, Oakland, California, U.S.A.*
- Bruce D. Rogers** / *Institute of Government, Tennessee State University, Nashville, Tennessee, U.S.A.*
- Rainer Rohdewohld** / *German Agency for Technical Cooperation (Deutsche Gesellschaft fuer Technische Zusammenarbeit—GTZ) (Support for Decentralization Measures Project/SfDM) (GTZ-SfDM), Jakarta, Indonesia*
- John A. Rohr** / *Center for Public Administration and Policy, Virginia Polytechnic Institute, Blacksburg, Virginia, U.S.A.*

- Mark Carl Rom** / *Georgetown Public Policy Institute (GPPI), Georgetown University, Washington, District of Columbia, U.S.A.*
- Jeffrey A. Rosen** / *School of Public Policy and Management, Ohio State University, Columbus, Ohio, U.S.A.*
- David H. Rosenbloom** / *Department of Public Administration and Policy, American University, Washington, District of Columbia, U.S.A.*
- Herbert J. Rubin** / *Department of Sociology, Northern Illinois University, DeKalb, Illinois, U.S.A.*
- Irene Rubin** / *Division of Public Administration, Northern Illinois University, DeKalb, Illinois, U.S.A.*
- Marilyn Marks Rubin** / *Department of Public Management, John Jay College of Criminal Justice, The City University of New York, New York, New York, U.S.A.*
- Mark R. Rutgers** / *Department of Public Administration, Leiden University, Leiden, The Netherlands*
- John F. Sacco** / *Department of Public and International Affairs, George Mason University, Fairfax, Virginia, U.S.A.*
- Alan L. Saltzstein** / *Division of Political Science and Criminal Justice, California State University, Fullerton, California, U.S.A.*
- Ishak Saporta** / *Faculty of Management, Tel Aviv University, Tel Aviv, Israel*
- Grant T. Savage** / *Department of Management and Marketing, University of Alabama, Tuscaloosa, Alabama, U.S.A.*
- E. S. Savas** / *School of Public Affairs, Baruch College, City University of New York, New York, New York, U.S.A.*
- Teva J. Scheer** / *University of Colorado at Denver, Highlands Ranch, Colorado, U.S.A.*
- Mary Schmeida** / *Department of Nursing Research and Innovation, Cleveland Clinic, Cleveland, Ohio, U.S.A.*
- George Schmelzle** / *Department of Accounting and Finance, School of Business and Management Science, Indiana University Purdue University, Fort Wayne, Indiana, U.S.A.*
- Anne L. Schneider** / *College of Public Programs, Arizona State University, Tempe, Arizona, U.S.A.*
- William D. Schreckhise** / *Department of Political Science, University of Arkansas, Fayetteville, Arkansas, U.S.A.*
- Raviv Schwartz** / *Ben-Gurion, University of the Negev, Beer-Sheva, Israel*
- Alex Sekwat** / *Institute of Government, Tennessee State University, Nashville, Tennessee, U.S.A.*
- Sally Coleman Selden** / *School of Business and Economics, Lynchburg College, Lynchburg, Virginia, U.S.A.*
- Leiyu Shi** / *Department of Health Policy and Management, Johns Hopkins University, Baltimore, Maryland, U.S.A.*
- Noore Alam Siddiquee** / *Department of Political Science, International Islamic University Malaysia, Kuala Lumpur, Malaysia*
- Carlos Nunes Silva** / *Department of Geography, Faculdade de Letras, University of Lisbon, Lisbon, Portugal*
- James R. Simmons** / *Department of Political Science, College of Letters and Sciences, University of Wisconsin-Oshkosh, Oshkosh, Wisconsin, U.S.A.*
- Bill Simonsen** / *Center for Policy Analysis and Management, University of Connecticut, Storrs, Connecticut, U.S.A.*
- Barbara Sims** / *School of Public Affairs, Pennsylvania State University at Harrisburg, Middletown, Pennsylvania, U.S.A.*
- Richard G. Sims** / *Director of Tax Policy, Institute of Taxation and Economic Policy, Washington, District of Columbia, U.S.A.*
- Robert W. Smith** / *Department of Political Science, Clemson University, Clemson, South Carolina, U.S.A.*
- Keith F. Snider** / *Naval Postgraduate School, Monterey, California, U.S.A.*
- Douglas R. Snow** / *School of Public Management, Suffolk University, Boston, Massachusetts, U.S.A.*
- Hee Joon Song** / *Department of Public Administration, Ewha Womans University, Seoul, Korea*
- Linda L. Stanley** / *School of Business, Our Lady of the Lake University, San Antonio, Texas, U.S.A.*

- Catherine L. Staples** / *Department of Economics and Business, Randolph-Macon College, Ashland, Virginia, U.S.A.*
- Robert M. Stein** / *Department of Political Science, Rice University, Houston, Texas, U.S.A.*
- Benjamin Steiner** / *Division of Criminal Justice, University of Cincinnati, Cincinnati, Ohio, U.S.A.*
- G. Ross Stephens** / *Department of Political Science, University of Missouri-Kansas City, Kansas City, Missouri, U.S.A.*
- Margaret Stout** / *School of Public Affairs, Arizona State University, Phoenix, Arizona, U.S.A.*
- Ronald M. Stout Jr.** / *4th Judicial District, New York State Judiciary, Latham, New York, U.S.A.*
- Stuart C. Strother** / *School of Urban and Public Affairs, University of Louisville, Louisville, Kentucky, U.S.A.*
- Thomas C. Sutton** / *Department of Political Science, Baldwin-Wallace College, Berea, Ohio, U.S.A.*
- James H. Svava** / *Department of Political Science and Public Administration, North Carolina State University, Raleigh, North Carolina, U.S.A.*
- Liyan Tang** / *Tanaka Business School, Imperial College London, London, U.K.*
- Travis K. Taylor** / *Department of Accounting, Economics and Finance, Christopher Newport University, Newport News, Virginia, U.S.A.*
- Jon C. Teaford** / *Department of History, Purdue University, West Lafayette, Indiana, U.S.A.*
- Matthew Tedesco** / *Department of Philosophy, University of Colorado at Boulder, Boulder, Colorado, U.S.A.*
- Morton J. Tenzer** / *Department of Political Science, University of Connecticut, Storrs, Connecticut, U.S.A.*
- Fred Thompson** / *Atkinson Graduate School of Management, Willamette University, Salem, Oregon, U.S.A.*
- Costel D. Todor** / *Ministry of Justice, Bucharest, Romania*
- Marcela Tribble** / *School of Policy and Management, Florida International University, Miami, Florida, U.S.A.*
- James L. True** / *Department of Political Science, Lamar University, Beaumont, Texas, U.S.A.*
- Mark Turner** / *School of Management and Policy, University of Canberra, Canberra, Australian Capital Territory, Australia*
- Montgomery Van Wart** / *Department of Public Administration, California State University-San Bernardino, San Bernardino, California, U.S.A.*
- Walter Vance** / *General Accounting Office, Washington, District of Columbia, U.S.A.*
- Wouter Vandenabeele** / *Public Management Institute, Catholic University Leuven, Leuven, Belgium*
- Eran Vigoda** / *School of Political Sciences, University of Haifa, Haifa, Israel*
- Vera Vogelsang-Coombs** / *Maxine Goodman Levin College of Urban Affairs, Cleveland State University, Cleveland, Ohio, U.S.A.*
- William R. Voorhees** / *Department of Political Science and Public Administration, University of North Florida, Jacksonville, Florida and School of Public Affairs, Arizona State University, Tempe, Arizona, U.S.A.*
- Joe L. Wallis** / *Department of Economics, University of Otago, Dunedin, New Zealand*
- Kenneth F. Warren** / *Department of Political Science, Saint Louis University, St. Louis, Missouri, U.S.A.*
- Ryan J. Watson** / *Middletown, Maryland, U.S.A.*
- William L. Waugh Jr.** / *Department of Public Administration and Urban Studies, Georgia State University, Atlanta, Georgia, U.S.A.*
- Jeffrey A. Weber** / *Department of Political Science, East Stroudsburg University, East Stroudsburg, Pennsylvania, U.S.A.*
- Marcie Weinstein** / *College of Health Professions, Towson University, Towson, Maryland, U.S.A.*
- Clay Wescott** / *Government and Regional Cooperation Division, Asian Development Bank (ADB), Manila, Philippines*
- Jonathan P. West** / *Department of Political Science, University of Miami, Coral Gables, Florida, U.S.A.*

- Richard D. White Jr.** / *Public Administration Institute, Louisiana State University, Baton Rouge, Louisiana, U.S.A.*
- Andrew B. Whitford** / *Department of Political Science, University of Kansas, Lawrence, Kansas and Public Administration and Policy, University of Georgia, Athens, Georgia, U.S.A.*
- Daniel Williams** / *School of Public Affairs, Baruch College, City University of New York, New York, New York, U.S.A.*
- Frank J. Williams** / *Rhode Island Supreme Court, Providence, Rhode Island, U.S.A.*
- Russell L. Williams** / *Askew School of Public Administration and Policy, Florida State University, Tallahassee, Florida, U.S.A.*
- Katherine G. Willoughby** / *Department of Public Administration and Urban Studies, Georgia State University, Atlanta, Georgia, U.S.A.*
- Matthew Woessner** / *Department of Public Affairs, Pennsylvania State University at Harrisburg, Middletown, Pennsylvania, U.S.A.*
- John D. Wong** / *Hugo Wall School of Urban and Public Affairs, Wichita State University, Wichita, Kansas, U.S.A.*
- E. J. Woodhouse** / *Science and Technology Studies, Rensselaer Polytechnic Institute, Troy, New York, U.S.A.*
- John Abbott Worthley** / *Department of Public Administration, Seton Hall University, South Orange, New Jersey, U.S.A.*
- Elisabeth Wright** / *School of Business and Public Management, George Washington University, Washington, District of Columbia, U.S.A.*
- David A. Yalof** / *Department of Political Science, University of Connecticut, Storrs, Connecticut, U.S.A.*
- Thomas E. Yatsco** / *U.S. Department of Health and Human Services, Washington, District of Columbia, U.S.A.*
- Jong-In Yoon** / *Department of Political Science, University of Georgia, Athens, Georgia, U.S.A.*
- Peter C. Young** / *Graduate School of Business, University of St. Thomas, Minneapolis, Minnesota, U.S.A.*
- Habib Zafarullah** / *School of Social Science, University of New England, Armidale, New South Wales, Australia*
- Nikolaos Zahariadis** / *Department of Government, University of Alabama at Birmingham, Birmingham, Alabama, U.S.A.*
- Stephen J. Ziegler** / *Public and Environmental Affairs, Indiana University/Purdue University, Fort Wayne, Indiana, U.S.A.*
- Ulf Zimmermann** / *Department of Political Science and International Affairs, Kennesaw State University, Kennesaw, Georgia, U.S.A.*
- Cyrus Ernesto Zirakzadeh** / *Department of Political Science, University of Connecticut, Storrs, Connecticut, U.S.A.*

Contents

<i>Contributors</i>	ix
<i>Topical Table of Contents</i>	xxxiii
<i>Preface</i>	xlix
1993 Budget Reconciliation Bill and the Federal Budget Deficit / Patrick Fisher	1
360-Degree Performance Appraisal Systems / Neil M. Boyd	6
Accountability and Ethics: Reconsidering the Relationships / Melvin J. Dubnick	10
Accounting and Reporting for Private Non-Profit Organizations—Balancing Economic Efficiency with Social Mission / John F. Sacco and Walter Vance	25
Accounting, Financial / Mozaffar Khan	32
Achieving Productivity through Budgeting / Gerald J. Miller and Donijo Robbins	38
Acquiring Resources through Price Negotiation: A Public Sector Approach / Rupert G. Rhodd	43
Activity-Based Costing in the Public Sector / Thomas Buttross and George Schmelzle	47
Administration of State Lotteries / Jeff Leon Dense	53
Administrative Discretion / Kenneth F. Warren	58
Administrative Law Judges and Agency Adjudication / William D. Schreckhise	62
Administrative Reform in Southeast Asia / M. Shamsul Haque	66
Agenda Setting / David A. Rochefort	71
Agricultural Policy / William P. Browne	76
Alternative Dispute Resolution Processes / Tina Nabatchi and Lisa B. Bingham	80
Annexation / Jered B. Carr and Richard C. Feiock	85
Appearance of Impropriety / Stuart C. Gilman	94
Appreciative Inquiry / Thomas W. Kent	99
Arbitrage / Susan MacDonald	103
Asset Pricing / Greg Chen	110
Association of Southeast Asian Nations (ASEAN) / Richard K. Common	114
Attractors, Strange Attractors and Fractals / David Ogula	118
Auditing / Jean Harris and Ann Pushkin	123
Balanced Scorecard for Public Administration and Policy / Peter J. Reed	130
Balanced Scorecard Tools for Organizational Assessment in Public Administration and Policy / Peter J. Reed	136
Benchmarking in Public Administration and Public Policy / A. Gunasekaran and Krishan M. Gupta	141
Biomedical Ethics and Public Policy Discourse / John Abbott Worthley	147
Bond Ratings / Robert J. Eger and Andrew Haas	151
Bond Valuation / Richard L.B. LeCompte	154
Budget Execution and Management Control / L. R. Jones, Jerry McCaffery and Fred Thompson	158
Budget Reform / Jerry McCaffery	166
Budget Stabilization Fund / Yilin Hou	172

Budgetary Analysis and Economics / John R. Bartle	177
Budgetary Balance and Deficit Controls / William R. Voorhees	181
Budgetary Reforms in Southeast Asia / David Seth Jones	186
Budgeting and Public Choice / John R. Bartle	191
Budgeting, Line-Item and Object-of-Expenditure Controls / Suzanne Leland	195
Budgeting, Performance-Based / Julia E. Melkers	197
Budgeting, Target-Based / B. J. Reed	201
Budgeting: Allocation Formula / Patricia Moore	205
Budgets and Accountability / Jonathan Justice	209
Bureaucracy, Economic Theories of / James R. Simmons	217
Bureaucracy: Perspectives / Larkin S. Dudley	221
Bureaucratic Agencies: Delegation of Authority / Scott R. Furlong	225
Bureaucratic Discretion: Political Control and Influence / Scott R. Furlong	230
Bureaucratic Politics / Gene A. Brewer	236
Bureaucrats and Politicians in Southeast Asia / Scott Fritzen	242
Burnout in Public Agencies / Robert T. Golembiewski	247
Cambodia / Clay Wescott	253
Capital Budgeting / Jane Beckett-Camarata	257
Capital Markets / Kenneth A. Kriz and Peter J. Kiedrowski	264
Capital Programming and Budgeting: Comparative Local Government Perspectives / George M. Guess and Costel D. Todor	270
Capital Purchases / Wendell C. Lawther and John Adler	279
Career Development / Teva J. Scheer	283
Cash Flow Forecasting: Principles / William R. Voorhees and Jeongwoo Kim	285
Cash Management for State and Local Governments / Dwight V. Denison	291
Chinese Public Administration: History and Current Trends / Yijia Jing	297
Citizen Participation in Local Government / Anthony J. Nownes	301
Citizen Participation in the Budget Process: Exit, Voice, and Loyalty / Carol Ebdon	305
City Government / Jon C. Teafor	309
Civic Duty / Vera Vogelsang-Coombs and Larry Bakken	313
Civil Disobedience / Matthew Tedesco and John Harris	321
Civil Service / J. Edward Kellough	327
Civil Service Reform Act of 1978 / Jerrell D. Cogburn	329
Civil Service Reform and Reinvention / Jerrell D. Cogburn	335
Collaborative Governance / David E. Booher	343
Collective Bargaining / Robert Hebdon	348
Collective Bargaining, The Politics of / Robert Hebdon	351
Community-Based Planning for HIV/AIDS / Bruce W. Kieler and Ishak Saporta	355
Compensation: Pay and Benefits / James A. Buford Jr.	364
Comprehensive Annual Financial Report / Robert S. Kravchuk	369
Conflict of Interest / Russell L. Williams	374
Consolidated Governments / Dan Durning	377
Constitutional Constraints on Administrative Behavior in the United States / David H. Rosenbloom and Helena K. Rene	382
Constitutional Constraints on the Administrations of Public Policy / Matthew Woessner	391
Constructive Change Proposals, Assessing the Validity of / Elisabeth Wright	395
Contracting for Social Services / Nancy McCarthy Snyder	399
Control Systems / Lynda M. Dennis	404

Cooperative Purchasing / Alex Sekwat	410
Correctional Administration, The Challenges of / Betsy A. Matthews	413
Council/Commission Representation / Tari Renner	422
County Government / J. Edwin Benton	426
Court Administrators as Managers and Executives / Steven W. Hays	432
Court Budgeting: Judicial Branch Independence and Accountability / John K. Hudzik	435
Court System Strategic Planning / Deborah A. Botch	442
Court Technology and Information Systems / Ronald M. Stout Jr.	448
Court Unification / James A. Gazell	454
Courts and Policy in the United States / Robert K. Christensen and Jiang He	459
Criminal Justice Public Policy / Barbara Sims	463
Crisis Policy-Making and Management in Southeast Asia / Scott Fritzen	470
Customer Service / Evan M. Berman	475
Data Integrity / Theo Edwin Maloy	480
Data Mining as a Decision-Making Aid / William J. Mead	485
Debt Affordability / Bill Simonsen, Mark D. Robbins and Raymond Brown	489
Debt Issuance / Beverly S. Bunch	494
Decentralization in Developing Countries / Scott Fritzen and Patrick W. Ong	498
Decentralization in Southeast Asia / Ledivina V. Cariño	504
Decision Making in Administrative Organizations: Theories / James S. Guseh	508
Decision Making, Incrementalism and Transformational Change / Keith Carrington	514
Decision Making, Open System and Non-Equilibrium / Keith Carrington	519
Decision Making, Rational / E. J. Woodhouse	524
Decision Making: Administrative Dispute Resolution / Becky A. Nesbit and Lisa B. Bingham	530
Default / Robert J. Eger and Andrew Haas	535
Democracy and Public Policy / Dale Krane and Gary S. Marshall	538
Deontology / James R. Heichelbech	545
Development Administration in Southeast Asia / Mark Turner	547
Development Policy / Lawrence S. Graham	552
Discipline and Dismissal / Steven W. Hays	555
Diversity / Sonia M. Ospina and James F. O'Sullivan	558
Economic Development and Organization / C. Fred Baughman	565
Economic Development, Citizen Participation and / Terry F. Buss, F. Stevens Redburn and Marcela Tribble	569
Economic Development, Community Organizations and / Herbert J. Rubin	577
Economic Development, Financing / William L. Waugh Jr.	582
Economic Development, Immigration, Self-Employment and / Marcela Tribble and Terry F. Buss	585
Economic Development, Infrastructure and / Robert Krol	593
Economic Development, State Government Administration of / Keith Boeckelman	598
Economic Development: Governance Concepts and Issues / Kuotsai Tom Liou	602
Economic Development: International / Cal Clark	610
Economic Efficiency / Brian E. Dollery and Joe L. Wallis	614
Economic Theories of Public Leadership / Brian E. Dollery and Joe L. Wallis	618
Economics, Transaction Cost / Glenn A. Daley	623
Education Policy / Frederick M. Hess and Joleen R. Okun	629
Education, State Administration of Primary and Secondary / Thomas C. Sutton	633

Efficiency / <i>Hwang-Sun Kang</i>	637
E-Government in East Asia / <i>Hee Joon Song</i>	639
Electric Industry Restructuring / <i>John D. Wong and Stacey M. Brewer</i>	643
Electronic Government / <i>Mary Maureen Brown</i>	648
Employee Benefits / <i>Jerrell D. Cogburn and Dennis M. Daley</i>	654
Enterprise: Public versus Private / <i>Thomas S. Nesselin</i>	661
Entitlements and the Congressional Budget Process / <i>Patrick Fisher</i>	665
Environmental Ethics / <i>Christine M. Reed</i>	669
Environmental Policy / <i>Thomas Greitens</i>	672
Environmental Pollution Control, State Government Administration of / <i>A. Hunter Bacot</i>	676
Equal Employment Opportunity and Affirmative Action / <i>Norma M. Riccucci</i>	681
Ethics and Administrative Reform / <i>Jeroen Maesschalck</i>	686
Ethics and Information and Communication Technology / <i>Brian Negin</i>	691
Ethics and Public Policy / <i>Amy K. Donahue</i>	696
Ethics in Organizations, Implementation of / <i>Evan M. Berman</i>	701
Ethics in Public Administration, Teaching of / <i>Donald Menzel</i>	705
Ethics, Feminist Perspective on / <i>Dana Burr Bradley</i>	709
Ethics, Governance Structures / <i>Carole L. Jurkiewicz</i>	712
Evolutionary Theory and Public Policy / <i>Steven A. Peterson</i>	716
Executive Budgeting / <i>Brendan F. Burke</i>	720
Executive Leadership in Local Government / <i>James H. Svava</i>	725
Federal Subsidies for New Entrepreneurial High-Growth Firms / <i>Terry F. Buss</i>	731
Federalism in “Homeland Security” and Crisis Management / <i>Alethia H. Cook</i> and <i>David J. Louscher</i>	740
Feminism and Chaos Theory / <i>Mary Ann McClure</i>	744
Feminist Comparative Policy / <i>Amy G. Mazur</i>	747
Financial Condition / <i>Jane Beckett-Camarata</i>	753
Financial Emergencies / <i>Jane Beckett-Camarata</i>	759
Financial Resource Management in Local Government / <i>Kenneth A. Klase</i>	766
Financial Statement Analysis / <i>Martin Ives</i>	771
Fiscal Distress of Local Governments in Israel / <i>Abraham Carmeli</i>	778
Fiscal Federalism / <i>William Duncombe</i>	783
Fiscal Transparency / <i>A. Premchand</i>	788
Foreign Policy Analysis / <i>Eric K. Leonard</i>	796
Forms of Government / <i>Victor S. DeSantis</i>	800
Freedom of Information Act—Federal / <i>Suzanne J. Piotrowski</i>	805
Futures and Options / <i>Greg Chen</i>	809
Game Theory / <i>Michael A. Kelley</i>	812
Gaza Disengagement: Israel Government Forced Resettlement (2005) / <i>Nitza Nachmias</i>	816
Gender and Public Administration / <i>DeLysa Burnier</i>	821
Gender Bias, Ethical Analysis of and in Public Policy / <i>J. J. Hendricks</i> and <i>April Hejka-Ekins</i>	824
Generally Accepted Accounting Principles (GAAP) / <i>Catherine L. Staples</i>	828
Geographic Information Systems and Public Policy / <i>Akhlaque Haque</i>	833
Global Ethics and Corruption / <i>Leo W.J.C. Huberts</i>	838
Governance Networks / <i>Robert Agranoff and Michael McGuire</i>	844
Government Failure: Theory / <i>Thomas S. Nesselin</i>	851
Government Performance and Results Act of 1993: A Retrospective / <i>G.L.A. Harris</i>	854

Government–Business Relations in East Asia / <i>Sung Gul Hong</i>	860
Government-Sponsored Enterprises (GSEs) / <i>Stuart C. Strother and Steven G. Koven</i>	866
Governors as Chief Administrators, The Changing Role of / <i>H. Edward Flentje</i>	871
Gramm–Rudman–Hollings: Impact on the Federal Budgetary Process / <i>G.L.A. Harris</i>	877
Grievance Procedures and Administration / <i>Lisa B. Bingham and Tina Nabatchi</i>	883
Groups and Their Properties / <i>Claus Langfred</i>	888
Health Care Decision Analysis, Alternatives in / <i>Krishna S. Dhir</i>	893
Health Care Decision-Making Behavior, Emerging Paradigms of / <i>Krishna S. Dhir</i>	897
Health Care Finance in the Twenty-First Century / <i>Judith J. Kirchhoff</i>	902
Health Care Organizations, Managing / <i>Grant T. Savage and Julie W. Robinson</i>	906
Health Care Policy / <i>Walter J. Jones</i>	912
Health Care Settings, Change in / <i>R. Wayne Boss and Mark L. McConkie</i>	916
Health Care, Assessment and Evaluation of / <i>Leiyu Shi</i>	923
Health Policy in Singapore / <i>M. Ramesh</i>	929
Home Rule / <i>Dale Krane</i>	933
Human Relations Management Theory / <i>Gerald T. Gabris</i>	938
Human Resource Management / <i>Stephen E. Condrey, Ilka M. Decker and Carl L. McCoy</i>	942
Human Resource Management for Productivity / <i>Seok-Hwan Lee</i>	945
Human Resource Planning / <i>Joan E. Pynes</i>	949
Human Resources Administrators in State Government, Changing Roles of / <i>Curtis R. Berry</i>	953
Human Resources Management in Local Government / <i>J. Edward Kellough</i>	962
Human Rights, Ethical Analysis of and in Public Policy / <i>Cyrus Ernesto Zirakzadeh</i>	965
Human Subjects Research: United States Public Policy / <i>Patricia M. Alt</i> and <i>Marcie Weinstein</i>	970
Humanitarian Intervention / <i>Michelle Maiese</i>	975
Impacts of Bureaucratic Reform on State Government Administration / <i>Brendan F. Burke</i>	981
Impasse Resolution Procedures / <i>Gary E. Roberts</i>	987
Incrementalism, Budgetary / <i>Carol W. Lewis</i>	992
Indonesia / <i>Rainer Rohdewohld</i>	997
Information Sources and State Policy Making / <i>Jill Clark</i>	1001
Information Systems / <i>Mary Maureen Brown and Piper S. Charles</i>	1005
Innovation and Public Policy / <i>Stuart S. Nagel</i>	1009
Integrated Healthcare Systems: New Trends, Emerging Models, and Future Shocks / <i>Yvonne J. Kochanowski</i>	1020
Internal Controls / <i>Robert W. Smith</i>	1024
International Public Sector Accounting Standards / <i>James L. Chan</i>	1029
Inventory Management / <i>Ronald John Hy</i>	1034
Israel Civil Society: Characteristics / <i>Nitza Nachmias and Amiram Bogot</i>	1037
Israel: Peace and Conflict Resolution Organizations / <i>Raviv Schwartz</i>	1042
Israel: Public Law / <i>Assaf Meydani</i>	1046
Israel: Spatial Policy / <i>Arie Hershcovich</i>	1049
Job Classification / <i>Lyn Holley</i>	1054
John Rawls / <i>James R. Heichelbech</i>	1063
Judicial Ethics / <i>David A. Yalof</i>	1067
Judicial Independence / <i>Ian Greene</i>	1071
Judicial Selection and Merit Selection / <i>Cole Blease Graham Jr.</i>	1076
Just War / <i>Cyrus Ernesto Zirakzadeh</i>	1079
Justice through Emancipation: Abraham Lincoln as a Model of Justice / <i>Frank J. Williams</i>	1084

Juvenile Stalking / <i>Deborah L. Laufersweiler-Dwyer and R. Gregg Dwyer</i>	1089
Juvenile Waiver / <i>Craig Hemmens and Benjamin Steiner</i>	1093
Knowledge Utilization: Issues and Recommendations / <i>Cheol H. Oh</i>	1097
Labor Policy in Thailand / <i>Andrew Brown</i>	1113
Labor–Management Partnerships and Cooperation / <i>W. David Patton</i>	1117
Labor–Management Rights and Responsibilities / <i>D. S. Chauhan</i>	1121
Lao People’s Democratic Republic (PDR) / <i>Clay Wescott</i>	1128
Law and Ethics / <i>John A. Rohr</i>	1132
Leadership / <i>Thomas W. Kent</i>	1136
Leadership in Organizations / <i>Douglas Ihrke</i>	1139
Leadership: High-Integrity / <i>Mark L. McConkie and R. Wayne Boss</i>	1145
Learning Organizations / <i>Montgomery Van Wart</i>	1149
Lebanon: Corruption / <i>Dimitri R. Hanna and Ramzi N. Nasser</i>	1153
Legal Ethics / <i>Christina M. Bellon</i>	1157
Life Cycle Costing / <i>Greg Chen, Daniel Williams and Troels Pind Adrian</i>	1164
Linear Programming / <i>Ronald John Hy</i>	1168
Local Governments: State Supervision of Budgets and Finance / <i>Julia Beckett</i>	1171
Local Services, Provision and Production / <i>Robert M. Stein</i>	1175
Logical Positivism and Public Administration / <i>Laurent Dobuzinkis</i>	1180
Logistics and Transportation / <i>Linda L. Stanley and Darin Matthews</i>	1184
Malaysia / <i>Noore Alam Siddiquee and Habib Zafarullah</i>	1189
Management for Quality in Government / <i>Michael E. Milakovich</i>	1194
Managing Mutual Commitments and Expectations: Contract Administration in Public Labor Relations / <i>D. S. Chauhan</i>	1198
Market Failure, Theory of / <i>Brian E. Dollery and Joe L. Wallis</i>	1203
Mathematical Modeling / <i>Roger J. Beck</i>	1207
Media and Bureaucracy in the United States / <i>Mordecai Lee</i>	1210
Medicaid Reforms in the States, Administration of / <i>Thomas E. Yatsco and Irene J. Barnett</i>	1214
Merit System / <i>J. Edward Kellough</i>	1219
Metropolitan Governance / <i>Ulf Zimmermann</i>	1221
Metropolitan Governmental Fragmentation / <i>Stephen J. Ziegler and David C. Nice</i>	1225
Microenterprise Finance and Economic Development / <i>Douglas R. Snow</i>	1230
Middle East and Regulation / <i>Ori Arbel-Ganz</i>	1234
Middle East: Corruption / <i>Doron Navot</i>	1239
Milgram Experiments / <i>Holona L. Ochs and Andrew B. Whitford</i>	1243
Mitigating HIV/AIDS’ Impact on Teachers and Administrators in Sub-Saharan Africa / <i>Terry F. Buss and Mumukshu Patel</i>	1247
Model Contracts to Reduce the Risks in Complex Information Technology Procurements / <i>Michael Asner</i>	1252
Models of Spatial Finance and “Voting with Your Feet” / <i>Ray Gonzales</i>	1256
Models of the Policy Process / <i>Thomas A. Birkland</i>	1260
Models, Types and Attributes of / <i>Jeffrey A. Weber</i>	1264
Moral Development Theory / <i>Richard D. White Jr.</i>	1266
Motivation / <i>Evan M. Berman</i>	1270
Motives and Incentives: Rethinking Budget Theory in a New Institutional Context / <i>William Earle Klay</i>	1275
Municipal Securities / <i>Mark D. Robbins and Bill Simonsen</i>	1280
Narrative Inquiry / <i>Sonia M. Ospina and Jennifer Dodge</i>	1285

Narrative Policy Analysis / <i>Emery Roe</i>	1289
National Security Policy / <i>Eric K. Leonard</i>	1293
Natural Resource Policy: Institutional Aspects / <i>Amy R. Poteete</i>	1298
Naturalistic Decision Making / <i>Anna C. Goldoff</i>	1303
Net Assets and Fund Balance / <i>Dean Michael Mead</i>	1307
Networks, Organizational / <i>Seok-Eun Kim</i>	1312
New Institutional Economics / <i>Brian E. Dollery and Joe L. Wallis</i>	1316
New Public Management / <i>Eran Vigoda</i>	1321
New Public Service / <i>Wouter Vandenabeele and Annie Hondeghem</i>	1326
New Sciences Approach and Public Administration / <i>Göktuğ Morçöl</i>	1331
Newtonian Paradigm / <i>Mary Ann McClure</i>	1336
No Child Left Behind Act of 2001 / <i>G.L.A. Harris</i>	1339
Non-Monetary Incentives and Productivity / <i>Hedy L. Isaacs</i>	1346
Non-Profit Organizations, External Relations in / <i>Mordecai Lee</i>	1351
Nuremberg Trials and Charter / <i>Morton J. Tenzer</i>	1356
Offset Requirements in International Public Procurement / <i>Travis K. Taylor</i>	1361
Ombuds and Ombuds Programs / <i>Tina Nabatchi and Lisa B. Bingham</i>	1366
Open Systems Theory / <i>Rupert F. Chisholm</i>	1369
Operations Research / <i>William L. Waugh Jr.</i>	1374
Orange County Bankruptcy / <i>Susan MacDonald</i>	1377
Organization Development / <i>Gerald T. Gabris</i>	1383
Organization Theories, Types of / <i>Robert T. Golembiewski</i>	1387
Organizational Citizenship Behavior / <i>Courtney E. Jensen and Larkin S. Dudley</i>	1392
Organizational Commitment / <i>Jong-In Yoon</i>	1396
Organizational Culture / <i>Craig Lundberg</i>	1402
Papers on the Science of Administration / <i>Mark L. McConkie and R. Wayne Boss</i>	1407
Parental Rights and Obligations / <i>Michael W. Austin</i>	1409
Participatory Policy Analysis / <i>Virginia Eubanks</i>	1413
Patronage and Spoils / <i>Donald E. Klingner</i>	1418
Pay for Performance and Merit Pay / <i>Joan E. Pynes</i>	1423
Pay-As-You-Go Financing / <i>David L. Baker</i>	1428
Pension Under-Funding: Generational Accounting Perspective / <i>Paul J.M. Klumpes</i> and <i>Liyan Tang</i>	1432
Performance Appraisal / <i>Dennis M. Daley</i>	1437
Performance Audits / <i>Brendan F. Burke and Wendy Haynes</i>	1442
Performance Indicators in State Administration, The Use of / <i>Peter J. Haas</i>	1447
Performance Measurement / <i>Patria de Lancer Julnes</i>	1450
Performance Measurement and Benchmarking in Local Government / <i>David N. Ammons</i>	1455
Performance Measurement: Citizen-Driven / <i>Kathe Callahan</i>	1459
Performance Reporting / <i>William Earle Klay and Sam M. McCall</i>	1465
Performance-Based Budgeting / <i>Judith R. Brown</i>	1469
Policy Change / <i>James L. True</i>	1473
Policy Design / <i>Anne L. Schneider and Helen Ingram</i>	1477
Policy Diffusion / <i>Karen Mossberger</i>	1482
Policy Implementation / <i>Ann O'M. Bowman</i>	1487
Policy Networks / <i>Jack W. Meek</i>	1491
Policy Sciences Approach / <i>Peter deLeon and Christine R. Martell</i>	1495
Policy Termination / <i>Mark R. Daniels</i>	1499

Policymaking in Local Government / <i>Jeremy L. Hall</i>	1503
Policy-Oriented Social Science Research: Measuring the Benefits / <i>Henry E. Kilpatrick</i>	1508
Popular Reporting / <i>Joy A. Clay</i>	1516
Portfolio Theory / <i>Richard L.B. LeCompte</i>	1519
Postmodernism and Public Policy / <i>Thomas J. Catlaw and Margaret Stout</i>	1524
Postpositivist Perspectives in Policy Analysis / <i>Göktuğ Morçöl</i>	1530
Practical Organization Theory / <i>Gerald T. Gabris</i>	1534
Pragmatism and Public Administration / <i>Keith F. Snider</i>	1538
Principal-Agent Perspective / <i>Arie Halachmi</i>	1542
Privatization / <i>Marc Holzer and Hwang-Sun Kang</i>	1546
Privatization and Economic Development / <i>Bill Luker</i>	1549
Privatization and Public-Private Partnerships for Local Services / <i>E. S. Savas</i>	1556
Privatization and Reinvention Paradigms for State Governmental Administration, Consequences of / <i>Julia E. Robinson</i>	1561
Probability Theory / <i>Bradley J. Best and Paul J. Brennan</i>	1566
Procurement / <i>Charles K. Coe</i>	1571
Productivity and Effectiveness / <i>Patria de Lancer Julnes</i>	1575
Program Budgeting (PPBS) / <i>Katherine G. Willoughby</i>	1579
Program Evaluation and Review Technique (PERT) and Critical Path Method (CPM) / <i>Bradley J. Best and Jeremy R.T. Lewis</i>	1583
Public Administration as Management / <i>Lawrence F. Keller</i>	1588
Public Administration, A Twenty-First Century Lens of Understanding / <i>Thomas D. Lynch, Cynthia E. Lynch and Peter L. Cruise</i>	1595
Public Authorities / <i>Suzanne Leland and Gary A. Johnson</i>	1601
Public Budgeting, Politics in / <i>Irene Rubin</i>	1605
Public Choice / <i>Darrene L. Hackler</i>	1614
Public Choice Theory / <i>Thomas S. Nesslein</i>	1618
Public Funds Investment Strategies / <i>William G. Albrecht</i>	1623
Public Goods / <i>Cristina E. Ciocirlan</i>	1628
Public Policy / <i>Steven A. Peterson</i>	1632
Public Policy Formation Under Conditions of Ambiguity / <i>Nikolaos Zahariadis</i>	1634
Public Procurement / <i>Robert E. Lloyd</i>	1642
Public Procurement Ethics / <i>David Seth Jones</i>	1647
Public Reporting / <i>Mordecai Lee</i>	1652
Public Sector Productivity: Barriers / <i>David N. Ammons</i>	1657
Public Service / <i>Richard A. Loverd</i>	1661
Public-Private Partnerships for Economic Development / <i>Steven G. Koven and Stuart C. Strother</i>	1664
Public-Private Partnerships in Developing Countries / <i>Steven G. Koven and Stuart C. Strother</i>	1669
Punctuated Equilibrium / <i>Scott E. Robinson</i>	1674
Q Methodology / <i>Dan Durning</i>	1678
Quantitative Decision Making: Benefit-Cost Analysis / <i>Jay P. Hamilton</i>	1682
Reciprocal Relations Among Peace, Prosperity, and Democracy / <i>Stuart S. Nagel</i>	1688
Reduction in Force / <i>Teva J. Scheer</i>	1692
Regulatory Policy: Role and Importance / <i>Scott R. Furlong</i>	1694
Representative Bureaucracy / <i>Katherine C. Naff</i>	1698
Restorative Justice / <i>Michelle Maiese</i>	1702
Revenue Forecasting and Revenue Estimation / <i>Marilyn Marks Rubin and Nancy Mantell</i>	1709

Risk Management / <i>David L. Baker</i>	1717
Risk Pools / <i>Peter C. Young</i>	1721
Rulemaking / <i>Cornelius M. Kerwin</i>	1727
Rural Economies, Developing / <i>Terry F. Buss and Marcela Tribble</i>	1732
Rural Health Policy / <i>Esther M. Forti</i>	1738
Science of Administration / <i>Mark R. Rutgers</i>	1742
Seasonality / <i>Daniel Williams</i>	1746
Senior Executive Service / <i>Robert F. Durant and Jean C. Accius II</i>	1757
Seniority / <i>Alan L. Saltzstein</i>	1762
Sexual Harassment / <i>Sally Coleman Selden</i>	1765
Singapore / <i>Jon S.T. Quah</i>	1773
Singapore: Economic Policy / <i>Charles Conteh</i>	1778
Social Networks / <i>Brian E. Adams</i>	1784
Special District Governments / <i>G. Ross Stephens</i>	1788
Stage Heuristics in Policy Sciences / <i>Christine R. Martell and Peter deLeon</i>	1795
Standard of Award / <i>Frayda S. Bluestein</i>	1799
Standards of Conduct / <i>Jeremy F. Plant</i>	1802
State Administration of Mental Retardation and Developmental Disabilities	
Policies / <i>Paul J. Castellani</i>	1806
State and Local Public Pension Fund Management / <i>Jun Peng</i>	1811
State Enterprise Zones / <i>Ryan J. Watson and Terry F. Buss</i>	1816
State Government Administration in a Federal System / <i>Timothy D. Mead</i>	1823
State Government Administration of Rural Development Policy / <i>Michael McGuire</i>	1827
State Government Administration, Impacts of Reform Movements / <i>Cynthia J. Bowling</i>	1832
State Government Adoption of “Managing for Results” Reforms / <i>Donald P. Moynihan</i>	1838
State Government Policy Initiatives: Improving Access for the	
Medically Underserved / <i>Mary Schmeida</i>	1843
State Government, Aspects of Managing Strategic Change in / <i>Paul C. Nutt,</i>	
<i>Jeffrey A. Rosen and Gregory Paxton</i>	1847
State Government: Complex Policy Network Management / <i>Kathleen Hale</i>	1853
State Government: Promotion of Innovations / <i>Keon S. Chi</i>	1858
State Legislative Fiscal Offices and State Government Administration / <i>Anita Chadha</i>	1864
Strategic Management and Productivity / <i>Arie Halachmi</i>	1867
Subnational Counter-Cyclical Fiscal Policy in the United States / <i>Yilin Hou</i>	1875
Supplemental Appropriation / <i>Jerry McCaffery</i>	1880
Sustainable Development / <i>Richard C. Feiock and Stephen C. Adams</i>	1884
Tax Expenditures / <i>Iryna Illiash</i>	1887
Tax Incidence / <i>Daphne Greenwood</i>	1893
Team Building / <i>R. Wayne Boss and Mark L. McConkie</i>	1899
Telehealth and State Government Policy / <i>Mary Schmeida</i>	1904
Teleology and Utilitarianism / <i>James R. Heichelbech</i>	1908
Thailand / <i>Brian Brewer</i>	1911
Theory, Types and Attributes of / <i>Jeffrey A. Weber</i>	1916
Third-Party Interventions / <i>R. Wayne Boss and Mark L. McConkie</i>	1920
Time Value of Money / <i>John D. Wong</i>	1923
Times Series Analysis / <i>Bruce D. Rogers</i>	1931
Trade Policy / <i>Philip A. Mundo</i>	1936
Training and Development / <i>Joan E. Pynes</i>	1942

Training for Productivity / <i>Montgomery Van Wart</i>	1949
Transparency and Corruption in Southeast Asia / <i>Habib Zafarullah and Noore Alam Siddiquee</i> .	1955
Transportation Policy / <i>Jeremy F. Plant</i>	1960
Trust in Organizations / <i>David G. Carnevale</i>	1965
Truth and Reconciliation Commission / <i>Daryl Balia</i>	1969
Truthfulness and Lying / <i>Roger Green</i>	1972
Tuskegee Study / <i>Holona L. Ochs and Andrew B. Whitford</i>	1977
U.S. Office of Personnel Management / <i>Teva J. Scheer</i>	1981
Uncertainty in Decision Making / <i>William J. Mead</i>	1984
Uncertainty, Concept of / <i>James McBain</i>	1988
Understanding the Basics of Refunding in the Municipal Bond Market / <i>Jun Peng</i>	1991
Uniform Guidelines on Employee Selection Procedures / <i>Teva J. Scheer</i>	1995
Unit Determination / <i>Jonathan P. West</i>	1998
United States Treasury Securities / <i>Theo Edwin Maloy</i>	2001
Urban Planning and Ethics / <i>Carlos Nunes Silva</i>	2006
Use and Management of Information Technology in State Government / <i>Paul M. Chalekian</i>	2012
Using Econometric Models for Public-Sector Decision Making / <i>Richard G. Sims</i>	2017
Validation / <i>Teva J. Scheer</i>	2024
Values and Policy Analysis / <i>Steven A. Peterson</i>	2026
Virtue Ethics / <i>James S. Bowman</i>	2028
Watergate / <i>Carol W. Lewis</i>	2033
Welfare and Social Security Policy / <i>Mark Carl Rom</i>	2041
Welfare Economics / <i>Glenn A. Daley</i>	2047
Welfare Reform, Management of / <i>Edward T. Jennings Jr.</i>	2054
Welfare State / <i>Paul B. Kirby</i>	2060
Whistleblowing, Corporate, and Public Policy / <i>Michael W. Austin and Michael A. Harper</i>	2066
Win-Win Approaches to Domestic Peace, Prosperity, and Democracy / <i>Stuart S. Nagel</i>	2070
Win-Win Approaches to International Peace, Prosperity, and Democracy / <i>Stuart S. Nagel</i>	2076
Win-Win Policy / <i>Stuart S. Nagel</i>	2083
Work Motivation / <i>Hal G. Rainey</i>	2089
Working Capital / <i>Justin Marlowe</i>	2093
Yield Curve for Public-Sector Debt Instruments / <i>Theo Edwin Maloy</i>	2097
Zero-Base Budgeting / <i>J. L. Peters</i>	2102

Topical Table of Contents

Budgeting

Auditing and Reporting

Accounting and Reporting for Private Non-Profit Organizations—Balancing Economic Efficiency with Social Mission / John F. Sacco and Walter Vance	25
Auditing / Jean Harris and Ann Pushkin	123
Budgets and Accountability / Jonathan Justice	209
Comprehensive Annual Financial Report / Robert S. Kravchuk	369
Popular Reporting / Joy A. Clay	1516

Budget Process

Budget Execution and Management Control / L. R. Jones, Jerry McCaffery and Fred Thompson	158
Budget Stabilization Fund / Yilin Hou	172
Budgetary Balance and Deficit Controls / William R. Voorhees	181
Budgeting: Allocation Formula / Patricia Moore	205
Entitlements and the Congressional Budget Process / Patrick Fisher	665
Executive Budgeting / Brendan F. Burke	720
Incrementalism, Budgetary / Carol W. Lewis	992
Internal Controls / Robert W. Smith	1024
Public Budgeting, Politics in / Irene Rubin	1605
Supplemental Appropriation / Jerry McCaffery	1880

Budget Reform

Budget Reform / Jerry McCaffery	166
Budgetary Reforms in Southeast Asia / David Seth Jones	186

Cash Management

Cash Flow Forecasting: Principles / William R. Voorhees and Jeongwoo Kim	285
Cash Management for State and Local Governments / Dwight V. Denison	291

Costing and Pricing

Activity-Based Costing in the Public Sector / Thomas Buttross and George Schmelzle	47
Asset Pricing / Greg Chen	110
Life Cycle Costing / Greg Chen, Daniel Williams and Troels Pind Adrian	1164

Budgeting (*cont'd.*)

Laws and Regulation

1993 Budget Reconciliation Bill and the Federal Budget Deficit / Patrick Fisher	1
Generally Accepted Accounting Principles (GAAP) / Catherine L. Staples	828
Gramm–Rudman–Hollings: Impact on the Federal Budgetary Process / G.L.A. Harris	877
International Public Sector Accounting Standards / James L. Chan	1029

Theory

Achieving Productivity through Budgeting / Gerald J. Miller and Donijo Robbins	38
Budgeting and Public Choice / John R. Bartle	191
Budgets and Accountability / Jonathan Justice	209
Motives and Incentives: Rethinking Budget Theory in a New Institutional Context / William Earle Klay	1275

Types of Budgets

Budgeting, Line-Item and Object-of-Expenditure Controls / Suzanne Leland	195
Budgeting, Performance-Based / Julia E. Melkers	197
Budgeting, Target-Based / B. J. Reed	201
Capital Budgeting / Jane Beckett-Camarata	257
Performance-Based Budgeting / Judith R. Brown	1469
Program Budgeting (PPBS) / Katherine G. Willoughby	1579
Zero-Base Budgeting / J. L. Peters	2102

Bureaucracy

Bureaucracy, Economic Theories of / James R. Simmons	217
Bureaucracy: Perspectives / Larkin S. Dudley	221
Bureaucratic Agencies: Delegation of Authority / Scott R. Furlong	225
Bureaucratic Discretion: Political Control and Influence / Scott R. Furlong	230
Bureaucratic Politics / Gene A. Brewer	236
Impacts of Bureaucratic Reform on State Government Administration / Brendan F. Burke	981
Media and Bureaucracy in the United States / Mordecai Lee	1210
Representative Bureaucracy / Katherine C. Naff	1698

Conflict Resolution

Alternative Dispute Resolution Processes / Tina Nabatchi and Lisa B. Bingham	80
Impasse Resolution Procedures / Gary E. Roberts	987

Countries and Regions

Administrative Reform in Southeast Asia / M. Shamsul Haque	66
Association of Southeast Asian Nations (ASEAN) / Richard K. Common	114
Cambodia / Clay Wescott	253

Chinese Public Administration: History and Current Trends	/ Yijia Jing	297
Decentralization in Developing Countries	/ Scott Fritzen and Patrick W. Ong	498
Development Administration in Southeast Asia	/ Mark Turner	547
Indonesia	/ Rainer Rohdewohld	997
Lao People's Democratic Republic (PDR)	/ Clay Wescott	1128
Malaysia	/ Noore Alam Siddiquee and Habib Zafarullah	1189
Singapore	/ Jon S.T. Quah	1773
Thailand	/ Brian Brewer	1911

Specific Topics

Budgetary Reforms in Southeast Asia	/ David Seth Jones	186
Bureaucrats and Politicians in Southeast Asia	/ Scott Fritzen	242
Crisis Policy-Making and Management in Southeast Asia	/ Scott Fritzen	470
Decentralization in Southeast Asia	/ Ledivina V. Cariño	504
E-Government in East Asia	/ Hee Joon Song	639
Fiscal Distress of Local Governments in Israel	/ Abraham Carmeli	778
Gaza Disengagement: Israel Government Forced Resettlement (2005)	/ Nitza Nachmias	816
Government-Business Relations in East Asia	/ Sung Gul Hong	860
Israel Civil Society: Characteristics	/ Nitza Nachmias and Amiram Bogot	1037
Israel: Peace and Conflict Resolution Organizations	/ Raviv Schwartz	1042
Israel: Spatial Policy	/ Arie Hershcovich	1049
Labor Policy in Thailand	/ Andrew Brown	1113
Lebanon: Corruption	/ Dimitri R. Hanna and Ramzi N. Nasser	1153
Middle East and Regulation	/ Ori Arbel-Ganz	1234
Middle East: Corruption	/ Doron Navot	1239
Mitigating HIV/AIDS' Impact on Teachers and Administrators in Sub-Saharan Africa	/ Terry F. Buss and Mumukshu Patel	1247
Singapore: Economic Policy	/ Charles Conteh	1778
Transparency and Corruption in Southeast Asia	/ Habib Zafarullah and Noore Alam Siddiquee	1955

Court Administration

Court Administrators as Managers and Executives	/ Steven W. Hays	432
Court Budgeting: Judicial Branch Independence and Accountability	/ John K. Hudzik	435
Court System Strategic Planning	/ Deborah A. Botch	442
Court Technology and Information Systems	/ Ronald M. Stout Jr.	448
Court Unification	/ James A. Gazell	454
Courts and Policy in the United States	/ Robert K. Christensen and Jiang He	459
Judicial Independence	/ Ian Greene	1071
Judicial Selection and Merit Selection	/ Cole Blease Graham Jr.	1076

Decision-Making

Decision Making in Administrative Organizations: Theories	/ James S. Guseh	508
Decision Making, Incrementalism and Transformational Change	/ Keith Carrington	514

Decision-Making (*cont'd.*)

Decision Making, Open System and Non-Equilibrium / Keith Carrington	519
Decision Making, Rational / E. J. Woodhouse	524
Decision Making: Administrative Dispute Resolution / Becky A. Nesbit and Lisa B. Bingham	530
Naturalistic Decision Making / Anna C. Goldoff	1303
Quantitative Decision Making: Benefit–Cost Analysis / Jay P. Hamilton	1682

Economic Development

Economic Development and Organization / C. Fred Baughman	565
Economic Development, Citizen Participation and / Terry F. Buss, F. Stevens Redburn and Marcela Tribble	569
Economic Development, Community Organizations and / Herbert J. Rubin	577
Economic Development, Financing / William L. Waugh Jr.	582
Economic Development, Immigration, Self-Employment and / Marcela Tribble and Terry F. Buss	585
Economic Development, Infrastructure and / Robert Krol	593
Economic Development, State Government Administration of / Keith Boeckelman	598
Economic Development: Governance Concepts and Issues / Kuotsai Tom Liou	602
Economic Development: International / Cal Clark	610
Microenterprise Finance and Economic Development / Douglas R. Snow	1230
Privatization and Economic Development / Bill Luker	1549
Public–Private Partnerships for Economic Development / Steven G. Koven and Stuart C. Strother	1664
Public–Private Partnerships in Developing Countries / Steven G. Koven and Stuart C. Strother	1669
Rural Economies, Developing / Terry F. Buss and Marcela Tribble	1732
Singapore: Economic Policy / Charles Conteh	1778
Sustainable Development / Richard C. Feiock and Stephen C. Adams	1884

Ethics

Specific Topics

Environmental Ethics / Christine M. Reed	669
Ethics and Information and Communication Technology / Brian Negin	691
Ethics and Public Policy / Amy K. Donahue	696
Gender Bias, Ethical Analysis of and in Public Policy / J. J. Hendricks and April Hejka-Ekins	824
Global Ethics and Corruption / Leo W.J.C. Huberts	838
Humanitarian Intervention / Michelle Maiese	975

Cases

Biomedical Ethics and Public Policy Discourse / John Abbott Worthley	147
Lebanon: Corruption / Dimitri R. Hanna and Ramzi N. Nasser	1153
Middle East: Corruption / Doron Navot	1239
Milgram Experiments / Holona L. Ochs and Andrew B. Whitford	1243

Nuremberg Trials and Charter / Morton J. Tenzer	1356
Transparency and Corruption in Southeast Asia / Habib Zafarullah and Noore Alam Siddiquee	1955
Tuskegee Study / Holona L. Ochs and Andrew B. Whitford	1977
Urban Planning and Ethics / Carlos Nunes Silva	2006
Watergate / Carol W. Lewis	2033

Ethics Management

Conflict of Interest / Russell L. Williams	374
Ethics and Administrative Reform / Jeroen Maesschalck	686
Ethics in Organizations, Implementation of / Evan M. Berman	701
Ethics, Governance Structures / Carole L. Jurkiewicz	712
Human Rights, Ethical Analysis of and in Public Policy / Cyrus Ernesto Zirakzadeh	965
Human Subjects Research: United States Public Policy / Patricia M. Alt and Marcie Weinstein	970
Ombuds and Ombuds Programs / Tina Nabatchi and Lisa B. Bingham	1366
Standards of Conduct / Jeremy F. Plant	1802
Whistleblowing, Corporate, and Public Policy / Michael W. Austin and Michael A. Harper	2066

Legal Aspects

Appearance of Impropriety / Stuart C. Gilman	94
Judicial Ethics / David A. Yalof	1067
Law and Ethics / John A. Rohr	1132
Legal Ethics / Christina M. Bellon	1157

Teaching

Ethics in Public Administration, Teaching of / Donald Menzel	705
---	-----

Theory

Accountability and Ethics: Reconsidering the Relationships / Melvin J. Dubnick	10
Civic Duty / Vera Vogelsang-Coombs and Larry Bakken	313
Ethics, Feminist Perspective on / Dana Burr Bradley	709
John Rawls / James R. Heichelbech	1063
Moral Development Theory / Richard D. White Jr.	1266
Teleology and Utilitarianism / James R. Heichelbech	1908
Virtue Ethics / James S. Bowman	2028

Trust

Trust in Organizations / David G. Carnevale	1965
Truth and Reconciliation Commission / Daryl Balia	1969
Truthfulness and Lying / Roger Green	1972

Finance

Specific Topics

Financial Condition / Jane Beckett-Camarata	753
Financial Emergencies / Jane Beckett-Camarata	759

Finance (*cont'd.*)

Specific Topics (cont'd.)

Financial Resource Management in Local Government / Kenneth A. Klase	766
Fiscal Distress of Local Governments in Israel / Abraham Carmeli	778
Orange County Bankruptcy / Susan MacDonald	1377
Pay-As-You-Go Financing / David L. Baker	1428
Pension Under-Funding: Generational Accounting Perspective / Paul J.M. Klumpes and Liyan Tang	1432
Subnational Counter-Cyclical Fiscal Policy in the United States / Yilin Hou	1875
Tax Expenditures / Iryna Illiash	1887
Working Capital / Justin Marlowe	2093

Bonds

Bond Ratings / Robert J. Eger and Andrew Haas	151
Bond Valuation / Richard L.B. LeCompte	154
Municipal Securities / Mark D. Robbins and Bill Simonsen	1280
Understanding the Basics of Refunding in the Municipal Bond Market / Jun Peng	1991
United States Treasury Securities / Theo Edwin Maloy	2001

Debt

Debt Affordability / Bill Simonsen, Mark D. Robbins and Raymond Brown	489
Debt Issuance / Beverly S. Bunch	494

Markets

Arbitrage / Susan MacDonald	103
Capital Markets / Kenneth A. Kriz and Peter J. Kiedrowski	264
Futures and Options / Greg Chen	809
Yield Curve for Public-Sector Debt Instruments / Theo Edwin Maloy	2097

Reporting

Accounting, Financial / Mozaffar Khan	32
Financial Statement Analysis / Martin Ives	771
Fiscal Transparency / A. Premchand	788
Net Assets and Fund Balance / Dean Michael Mead	1307

Theory

Fiscal Federalism / William Duncombe	783
Portfolio Theory / Richard L.B. LeCompte	1519
Public Funds Investment Strategies / William G. Albrecht	1623
Revenue Forecasting and Revenue Estimation / Marilyn Marks Rubin and Nancy Mantell	1709
Risk Management / David L. Baker	1717
Risk Pools / Peter C. Young	1721
Time Value of Money / John D. Wong	1923

Gender Issues

Feminism and Chaos Theory / Mary Ann McClure	744
Feminist Comparative Policy / Amy G. Mazur	747
Gender and Public Administration / DeLysa Burnier	821
Gender Bias, Ethical Analysis of and in Public Policy / J. J. Hendricks and April Hejka-Ekins	824

Health Care

Health Care Decision Analysis, Alternatives in / Krishna S. Dhir	893
Health Care Decision-Making Behavior, Emerging Paradigms of / Krishna S. Dhir	897
Health Care Finance in the Twenty-First Century / Judith J. Kirchhoff	902
Health Care Organizations, Managing / Grant T. Savage and Julie W. Robinson	906
Health Care Policy / Walter J. Jones	912
Health Care Settings, Change in / R. Wayne Boss and Mark L. McConkie	916
Health Care, Assessment and Evaluation of / Leiyu Shi	923
Health Policy in Singapore / M. Ramesh	929
Integrated Healthcare Systems: New Trends, Emerging Models, and Future Shocks / Yvonne J. Kochanowski	1020
Rural Health Policy / Esther M. Forti	1738

Human Resource Management

Human Relations Management Theory / Gerald T. Gabris	938
Human Resource Management / Stephen E. Condrey, Ilka M. Decker and Carl L. McCoy	942
Human Resource Planning / Joan E. Pynes	949
Human Resources Administrators in State Government, Changing Roles of / Curtis R. Berry	953
Human Resources Management in Local Government / J. Edward Kellough	962
Labor Policy in Thailand / Andrew Brown	1113
U.S. Office of Personnel Management / Teva J. Scheer	1981

Appraisal

360-Degree Performance Appraisal Systems / Neil M. Boyd	6
Performance Appraisal / Dennis M. Daley	1437

Civil Service

Civil Service / J. Edward Kellough	327
Civil Service Reform Act of 1978 / Jerrell D. Cogburn	329
Civil Service Reform and Reinvention / Jerrell D. Cogburn	335
Job Classification / Lyn Holley	1054
Merit System / J. Edward Kellough	1219
New Public Service / Wouter Vandenabeele and Annie Hondegheem	1326
Patronage and Spoils / Donald E. Klingner	1418
Public Service / Richard A. Loverd	1661
Reduction in Force / Teva J. Scheer	1692

Human Resource Management (*cont'd.*)

Civil Service (cont'd.)

Senior Executive Service / Robert F. Durant and Jean C. Accius II.	1757
Seniority / Alan L. Saltzstein	1762
Uniform Guidelines on Employee Selection Procedures / Teva J. Scheer	1995

Compensation

Compensation: Pay and Benefits / James A. Buford Jr.	364
Employee Benefits / Jerrell D. Cogburn and Dennis M. Daley	654
Pay for Performance and Merit Pay / Joan E. Pynes	1423

Development

Burnout in Public Agencies / Robert T. Golembiewski	247
Career Development / Teva J. Scheer	283
Human Resource Management for Productivity / Seok-Hwan Lee	945
Non-Monetary Incentives and Productivity / Hedy L. Isaacs	1346
Team Building / R. Wayne Boss and Mark L. McConkie	1899
Training and Development / Joan E. Pynes	1942

Discipline

Discipline and Dismissal / Steven W. Hays	555
Grievance Procedures and Administration / Lisa B. Bingham and Tina Nabatchi	883
Sexual Harassment / Sally Coleman Selden	1765

Diversity

Diversity / Sonia M. Ospina and James F. O'Sullivan	558
Equal Employment Opportunity and Affirmative Action / Norma M. Riccucci	681

Labor–Management Relations

Collective Bargaining / Robert Hebdon	348
Collective Bargaining, The Politics of / Robert Hebdon	351
Labor–Management Partnerships and Cooperation / W. David Patton	1117
Labor–Management Rights and Responsibilities / D. S. Chauhan	1121
Managing Mutual Commitments and Expectations: Contract Administration in Public Labor Relations / D. S. Chauhan	1198
Unit Determination / Jonathan P. West	1998

Law

Administrative Discretion / Kenneth F. Warren	58
Administrative Law Judges and Agency Adjudication / William D. Schreckhise	62
Civil Disobedience / Matthew Tedesco and John Harris	321
Constitutional Constraints on Administrative Behavior in the United States / David H. Rosenbloom and Helena K. Rene	382
Constitutional Constraints on the Administrations of Public Policy / Matthew Woessner	391

Freedom of Information Act–Federal / <i>Suzanne J. Piotrowski</i>	805
Israel: Public Law / <i>Assaf Meydani</i>	1046
Just War / <i>Cyrus Ernesto Zirakzadeh</i>	1079
Justice through Emancipation: Abraham Lincoln as a Model of Justice / <i>Frank J. Williams</i>	1084
Juvenile Stalking / <i>Deborah L. Lauferseweiler-Dwyer and R. Gregg Dwyer</i>	1089
Juvenile Waiver / <i>Craig Hemmens and Benjamin Steiner</i>	1093
Restorative Justice / <i>Michelle Maiese</i>	1702
Rulemaking / <i>Cornelius M. Kerwin</i>	1727
Sexual Harassment / <i>Sally Coleman Selden</i>	1765

Local Government

Annexation / <i>Jered B. Carr and Richard C. Feiock</i>	85
Capital Programming and Budgeting: Comparative Local Government Perspectives / <i>George M. Guess and Costel D. Todor</i>	270
City Government / <i>Jon C. Teafor</i>	309
Consolidated Governments / <i>Dan Durning</i>	377
Council/Commission Representation / <i>Tari Renner</i>	422
County Government / <i>J. Edwin Benton</i>	426
Executive Leadership in Local Government / <i>James H. Svara</i>	725
Forms of Government / <i>Victor S. DeSantis</i>	800
Government-Sponsored Enterprises (GSEs) / <i>Stuart C. Strother and Steven G. Koven</i>	866
Home Rule / <i>Dale Krane</i>	933
Local Governments: State Supervision of Budgets and Finance / <i>Julia Beckett</i>	1171
Metropolitan Governance / <i>Ulf Zimmermann</i>	1221
Metropolitan Governmental Fragmentation / <i>Stephen J. Ziegler and David C. Nice</i>	1225
Public Authorities / <i>Suzanne Leland and Gary A. Johnson</i>	1601
Special District Governments / <i>G. Ross Stephens</i>	1788

Methods

Appreciative Inquiry / <i>Thomas W. Kent</i>	99
Data Integrity / <i>Theo Edwin Maloy</i>	480
Data Mining as a Decision-Making Aid / <i>William J. Mead</i>	485
Deontology / <i>James R. Heichelbech</i>	545
Linear Programming / <i>Ronald John Hy</i>	1168
Mathematical Modeling / <i>Roger J. Beck</i>	1207
Narrative Inquiry / <i>Sonia M. Ospina and Jennifer Dodge</i>	1285
Operations Research / <i>William L. Waugh Jr.</i>	1374
Q Methodology / <i>Dan Durning</i>	1678
Seasonality / <i>Daniel Williams</i>	1746
Times Series Analysis / <i>Bruce D. Rogers</i>	1931
Using Econometric Models for Public-Sector Decision Making / <i>Richard G. Sims</i>	2017
Validation / <i>Teva J. Scheer</i>	2024

Methods (*cont'd.*)

Policy Analysis

Foreign Policy Analysis / Eric K. Leonard	796
Narrative Policy Analysis / Emery Roe	1289
Participatory Policy Analysis / Virginia Eubanks	1413
Policy-Oriented Social Science Research: Measuring the Benefits / Henry E. Kilpatrick	1508
Stage Heuristics in Policy Sciences / Christine R. Martell and Peter deLeon	1795
Values and Policy Analysis / Steven A. Peterson	2026

Organization

Learning Organizations / Montgomery Van Wart	1149
Networks, Organizational / Seok-Eun Kim	1312
Non-Profit Organizations, External Relations in / Mordecai Lee	1351
Organization Development / Gerald T. Gabris	1383
Organization Theories, Types of / Robert T. Golembiewski	1387
Organizational Citizenship Behavior / Courtney E. Jensen and Larkin S. Dudley	1392
Organizational Commitment / Jong-In Yoon	1396
Organizational Culture / Craig Lundberg	1402
Practical Organization Theory / Gerald T. Gabris	1534

Leadership

Leadership / Thomas W. Kent	1136
Leadership in Organizations / Douglas Ihrke	1139
Leadership: High-Integrity / Mark L. McConkie and R. Wayne Boss	1145

Performance

Efficiency / Hwang-Sun Kang	637
New Public Management / Eran Vigoda	1321
Productivity and Effectiveness / Patria de Lancer Julnes	1575
Public Sector Productivity: Barriers / David N. Ammons	1657

Information Technology

E-Government in East Asia / Hee Joon Song	639
Electronic Government / Mary Maureen Brown	648
Geographic Information Systems and Public Policy / Akhlaque Haque	833
Information Systems / Mary Maureen Brown and Piper S. Charles	1005

Other Strategies

Achieving Productivity through Budgeting / Gerald J. Miller and Donijo Robbins	38
Constructive Change Proposals, Assessing the Validity of / Elisabeth Wright	395

Customer Service / <i>Evan M. Berman</i>	475
Inventory Management / <i>Ronald John Hy</i>	1034
Knowledge Utilization: Issues and Recommendations / <i>Cheol H. Oh</i>	1097
Logistics and Transportation / <i>Linda L. Stanley and Darin Matthews</i>	1184
Management for Quality in Government / <i>Michael E. Milakovich</i>	1194
Privatization / <i>Marc Holzer and Hwang-Sun Kang</i>	1546
Privatization and Reinvention Paradigms for State Governmental Administration, Consequences of / <i>Julia E. Robinson</i>	1561
Program Evaluation and Review Technique (PERT) and Critical Path Method (CPM) / <i>Bradley J. Best and Jeremy R.T. Lewis</i>	1583
Strategic Management and Productivity / <i>Arie Halachmi</i>	1867

People

Human Resource Management for Productivity / <i>Seok-Hwan Lee</i>	945
Motivation / <i>Evan M. Berman</i>	1270
Non-Monetary Incentives and Productivity / <i>Hedy L. Isaacs</i>	1346
Team Building / <i>R. Wayne Boss and Mark L. McConkie</i>	1899
Training and Development / <i>Joan E. Pynes</i>	1942
Training for Productivity / <i>Montgomery Van Wart</i>	1949
Work Motivation / <i>Hal G. Rainey</i>	2089

Performance Measurement

Balanced Scorecard for Public Administration and Policy / <i>Peter J. Reed</i>	130
Balanced Scorecard Tools for Organizational Assessment in Public Administration and Policy / <i>Peter J. Reed</i>	136
Benchmarking in Public Administration and Public Policy / <i>A. Gunasekaran and Krishan M. Gupta</i>	141
Government Performance and Results Act of 1993: A Retrospective / <i>G.L.A. Harris</i>	854
Performance Audits / <i>Brendan F. Burke and Wendy Haynes</i>	1442
Performance Indicators in State Administration, The Use of / <i>Peter J. Haas</i>	1447
Performance Measurement / <i>Patria de Lancer Julnes</i>	1450
Performance Measurement and Benchmarking in Local Government / <i>David N. Ammons</i>	1455
Performance Measurement: Citizen-Driven / <i>Kathe Callahan</i>	1459
Performance Reporting / <i>William Earle Klay and Sam M. McCall</i>	1465

Policy Analysis

Administration of State Lotteries / <i>Jeff Leon Dense</i>	53
Agricultural Policy / <i>William P. Browne</i>	76
Correctional Administration, The Challenges of / <i>Betsy A. Matthews</i>	413
Criminal Justice Public Policy / <i>Barbara Sims</i>	463
Development Policy / <i>Lawrence S. Graham</i>	552
Education Policy / <i>Frederick M. Hess and Joleen R. Okun</i>	629
Electric Industry Restructuring / <i>John D. Wong and Stacey M. Brewer</i>	643
Environmental Policy / <i>Thomas Greitens</i>	672

Policy Analysis (*cont'd.*)

Federal Subsidies for New Entrepreneurial High-Growth Firms / Terry F. Buss	731
Federalism in “Homeland Security” and Crisis Management / Alethia H. Cook and David J. Louscher	740
Gaza Disengagement: Israel Government Forced Resettlement (2005) / Nitza Nachmias	816
Mitigating HIV/AIDS’ Impact on Teachers and Administrators in Sub-Saharan Africa / Terry F. Buss and Mumukshu Patel	1247
National Security Policy / Eric K. Leonard	1293
Natural Resource Policy: Institutional Aspects / Amy R. Poteete	1298
No Child Left Behind Act of 2001 / G.L.A. Harris	1339
Parental Rights and Obligations / Michael W. Austin	1409
Regulatory Policy: Role and Importance / Scott R. Furlong	1694
Telehealth and State Government Policy / Mary Schmeida	1904
Trade Policy / Philip A. Mundo	1936
Transportation Policy / Jeremy F. Plant	1960
Welfare and Social Security Policy / Mark Carl Rom	2041
Welfare Reform, Management of / Edward T. Jennings Jr.	2054
Welfare State / Paul B. Kirby	2060

Policy-Making Process

Agenda Setting / David A. Rochefort	71
Biomedical Ethics and Public Policy Discourse / John Abbott Worthley	147
Citizen Participation in Local Government / Anthony J. Nownes	301
Citizen Participation in the Budget Process: Exit, Voice, and Loyalty / Carol Ebdon	305
Collaborative Governance / David E. Booher	343
Community-Based Planning for HIV/AIDS / Bruce W. Kieler and Ishak Saporta	355
Ethics and Public Policy / Amy K. Donahue	696
Governance Networks / Robert Agranoff and Michael McGuire	844
Innovation and Public Policy / Stuart S. Nagel	1009
Models of the Policy Process / Thomas A. Birkland	1260
Policy Change / James L. True	1473
Policy Design / Anne L. Schneider and Helen Ingram	1477
Policy Diffusion / Karen Mossberger	1482
Policy Implementation / Ann O’M. Bowman	1487
Policy Networks / Jack W. Meek	1491
Policy Sciences Approach / Peter deLeon and Christine R. Martell	1495
Policy Termination / Mark R. Daniels	1499
Polymaking in Local Government / Jeremy L. Hall	1503
Public Policy / Steven A. Peterson	1632
Public Policy Formation Under Conditions of Ambiguity / Nikolaos Zahariadis	1634
Public Reporting / Mordecai Lee	1652
Social Networks / Brian E. Adams	1784
Urban Planning and Ethics / Carlos Nunes Silva	2006

Procurement

Acquiring Resources through Price Negotiation: A Public Sector Approach / <i>Rupert G. Rhodd</i>	43
Capital Purchases / <i>Wendell C. Lawther and John Adler</i>	279
Constructive Change Proposals, Assessing the Validity of / <i>Elisabeth Wright</i>	395
Contracting for Social Services / <i>Nancy McCarthy Snyder</i>	399
Cooperative Purchasing / <i>Alex Sekwat</i>	410
Model Contracts to Reduce the Risks in Complex Information Technology Procurements / <i>Michael Asner</i>	1252
Offset Requirements in International Public Procurement / <i>Travis K. Taylor</i>	1361
Privatization and Public-Private Partnerships for Local Services / <i>E. S. Savas</i>	1556
Procurement / <i>Charles K. Coe</i>	1571
Public Procurement / <i>Robert E. Lloyd</i>	1642
Public Procurement Ethics / <i>David Seth Jones</i>	1647
Standard of Award / <i>Frayda S. Bluestein</i>	1799

State Government

Education, State Administration of Primary and Secondary / <i>Thomas C. Sutton</i>	633
Environmental Pollution Control, State Government Administration of / <i>A. Hunter Bacot</i>	676
Governors as Chief Administrators, The Changing Role of / <i>H. Edward Flentje</i>	871
Information Sources and State Policy Making / <i>Jill Clark</i>	1001
Medicaid Reforms in the States, Administration of / <i>Thomas E. Yatsco and Irene J. Barnett</i>	1214
State Administration of Mental Retardation and Developmental Disabilities Policies / <i>Paul J. Castellani</i>	1806
State and Local Public Pension Fund Management / <i>Jun Peng</i>	1811
State Enterprise Zones / <i>Ryan J. Watson and Terry F. Buss</i>	1816
State Government Administration in a Federal System / <i>Timothy D. Mead</i>	1823
State Government Administration of Rural Development Policy / <i>Michael McGuire</i>	1827
State Government Administration, Impacts of Reform Movements / <i>Cynthia J. Bowling</i>	1832
State Government Adoption of "Managing for Results" Reforms / <i>Donald P. Moynihan</i>	1838
State Government Policy Initiatives: Improving Access for the Medically Underserved / <i>Mary Schmeida</i>	1843
State Government, Aspects of Managing Strategic Change in / <i>Paul C. Nutt, Jeffrey A. Rosen and Gregory Paxton</i>	1847
State Government: Complex Policy Network Management / <i>Kathleen Hale</i>	1853
State Government: Promotion of Innovations / <i>Keon S. Chi</i>	1858
State Legislative Fiscal Offices and State Government Administration / <i>Anita Chadha</i>	1864
Use and Management of Information Technology in State Government / <i>Paul M. Chalekian</i>	2012

Theories

Attractors, Strange Attractors and Fractals / <i>David Ogula</i>	118
Control Systems / <i>Lynda M. Dennis</i>	404
Default / <i>Robert J. Eger and Andrew Haas</i>	535

Theories (*cont'd.*)

Evolutionary Theory and Public Policy / Steven A. Peterson	716
Feminism and Chaos Theory / Mary Ann McClure	744
Government Failure: Theory / Thomas S. Nesselin	851
Groups and Their Properties / Claus Langfred	888
Local Services, Provision and Production / Robert M. Stein	1175
Logical Positivism and Public Administration / Laurent Dobuzinskis	1180
Models, Types and Attributes of / Jeffrey A. Weber	1264
Newtonian Paradigm / Mary Ann McClure	1336
Open Systems Theory / Rupert F. Chisholm	1369
Papers on the Science of Administration / Mark L. McConkie and R. Wayne Boss	1407
Postmodernism and Public Policy / Thomas J. Catlaw and Margaret Stout	1524
Postpositivist Perspectives in Policy Analysis / Göktuğ Morçöl	1530
Punctuated Equilibrium / Scott E. Robinson	1674
Reciprocal Relations Among Peace, Prosperity, and Democracy / Stuart S. Nagel Deceased	1688
Theory, Types and Attributes of / Jeffrey A. Weber	1916

Administration

Democracy and Public Policy / Dale Krane and Gary S. Marshall	538
New Sciences Approach and Public Administration / Göktuğ Morçöl	1331
Pragmatism and Public Administration / Keith F. Snider	1538
Public Administration as Management / Lawrence F. Keller	1588
Public Administration, A Twenty-First Century Lens of Understanding / Thomas D. Lynch, Cynthia E. Lynch and Peter L. Cruise	1595
Science of Administration / Mark R. Rutgers	1742

Economic

Budgetary Analysis and Economics / John R. Bartle	177
Budgeting and Public Choice / John R. Bartle	191
Economic Efficiency / Brian E. Dollery and Joe L. Wallis	614
Economic Theories of Public Leadership / Brian E. Dollery and Joe L. Wallis	618
Economics, Transaction Cost / Glenn A. Daley	623
Enterprise: Public versus Private / Thomas S. Nesselin	661
Market Failure, Theory of / Brian E. Dollery and Joe L. Wallis	1203
New Institutional Economics / Brian E. Dollery and Joe L. Wallis	1316
Public Goods / Cristina E. Ciocirlan	1628
Tax Incidence / Daphne Greenwood	1893
Welfare Economics / Glenn A. Daley	2047

Rational Actor, Non-Economic

Game Theory / Michael A. Kelley	812
Models of Spatial Finance and “Voting with Your Feet” / Ray Gonzales	1256
Principal-Agent Perspective / Arie Halachmi	1542
Public Choice / Darrene L. Hackler	1614

Public Choice Theory / <i>Thomas S. Nessel</i>	1618
Third-Party Interventions / <i>R. Wayne Boss and Mark L. McConkie</i>	1920
Win-Win Approaches to Domestic Peace, Prosperity, and Democracy / <i>Stuart S. Nagel</i>	2070
Win-Win Approaches to International Peace, Prosperity, and Democracy / <i>Stuart S. Nagel</i>	2076
Win-Win Policy / <i>Stuart S. Nagel</i>	2083

Statistical and Uncertainty

Probability Theory / <i>Bradley J. Best and Paul J. Brennan</i>	1566
Uncertainty in Decision Making / <i>William J. Mead</i>	1984
Uncertainty, Concept of / <i>James McBain</i>	1988

Preface

This edition is dedicated to the memory of Jack M. Rabin, who founded the Book Series in Public Administration and Public Policy, of which this encyclopedia is part. Jack sadly and prematurely passed away on November 13, 2006 at age 61. He is very much missed.

Jack's legacy is one of collegiality and productivity – not least of which are the about 130 volumes that make up the Book Series in Public Administration and Public Policy. This is among the largest series in the discipline. Jack started this series and remained passionately committed to it until his death. He also created numerous journals such as the *International Journal of Public Administration*, which he served as the Editor-in-Chief. Many scholars in public administration, if not most, have published in one or more Jack's works. Directly or indirectly, it is hard to find anyone who has not benefited from Jack's dedication to expand scholarship and opportunity to all in public administration.

I am one of those who are indebted to Jack. I remember talking to him, about a week before his death, about a new project. He encouraged me to continue with it, and he made it possible for to do so. It is this spirit of support and collaboration, his spirit, that I hope to continue through this volume of *Encyclopedia of Public Administration and Public Policy*, and indeed in the Book Series itself.

With this volume, I introduce myself as the new Editor-in-Chief of the Book Series in Public Administration and Public Policy. This book series, which originally started at Marcel Dekker, is now part of the Taylor & Francis Group, at CRC Press. We will continue to solicit quality manuscripts and work with authors to contribute to the professional practice and intellectual development of public administration through books. The Book Series in Public Administration and Public Policy lives on.

The *Encyclopedia of Public Administration and Public Policy* serves multiple purposes for our field. First, leading scholars define and describe key topics of our field in what we hope are authoritative ways. The encyclopedia provides the reader with entries concerning the salient theories, issues, and concepts within the broad fields of public administration and public policy.

Second, each entry is a coherent article, rather than a short paragraph. By doing so, this work also serves as a gateway into our vast field, giving the reader both an introduction to the various subjects and references to additional bodies of knowledge, and making it relevant to experts and new scholars and practitioners alike.

Third, not only is this encyclopedia also available on the Internet, we are on-going soliciting new entries and taking the opportunity to update existing entries, too. As such, the *Encyclopedia of Public Administration and Public Policy* becomes a “living thing,” that is constantly changing and improving. All authors have had the opportunity to update their entries for this volume.

I commend Taylor & Francis for its continuing support of this series. Jack regarded this Encyclopedia as the crown of the book series, and I agree. It is thus a special pleasure to dedicate this edition to him. I express my gratitude to the contributing editors as well as to all the authors. We will continue to ask a broad range of authors to contribute. Many thanks to all!

Evan M. Berman

Encyclopedia of

Public Administration and Public Policy

Second Edition

Volume 1

*Pages 1 through 680
1993–Environment*

1993–Alternative

Annexation–
Budget Stab

Budgetary–
Burnout

Cambodia–Civil

Collaborative–
County

Court–Decision

Default–
Economic Dev

Economic Eff–
Environment

Encyclopedia of

Public Administration and Public Policy

Second Edition

Volume 2

Pages 681 through 1406
Equal–Organizational

Equal–Feminist

Financial–Global

Governance–Home

Human–
International

Inventory–Law

Leadership–Merit

Metropolitan–
Networks

New–
Organizational

Encyclopedia of

Public Administration and Public Policy

Second Edition

Volume 3

Pages 1407 through 2106

Papers–Zero

Papers–Policy

Popular–Program

Public–Punctuated

Q–Sexual

Singapore–
State Gov

State Legis–Trade

Training–Using

Validation–Zero

1993 Budget Reconciliation Bill and the Federal Budget Deficit

Patrick Fisher

*Department of Political Science/Philosophy, Monmouth University,
West Long Branch, New Jersey, U.S.A.*

INTRODUCTION

When Bill Clinton was elected president in 1992, the United States had the largest peacetime budget deficit in the nation's history. Although reducing the deficit was not a major theme of Clinton's 1992 campaign for president, Clinton decided to make deficit reduction a major priority for his administration. The 1993 Budget Reconciliation Bill, the congressionally modified version of Clinton's budget blueprint, was passed in August 1993, without a vote to spare in Congress. The key component of the bill was its tax increases on those making more than \$125,000 a year. This was a significant departure from the tax policies of the Reagan and Bush administrations, which stressed the economic benefits of reducing taxes on upper incomes.

Although the tax increases of the 1993 Budget Reconciliation Bill were not politically popular and may have played an important role in the Democrats' dismal showing in the 1994 congressional elections, evidence suggests that the tax increases have played an important role in reducing the federal budget deficit. Since the 1993 Budget Reconciliation Bill became law, annual budget deficits have declined significantly. By 1998, the federal government actually had a budget surplus, the first time the federal budget ran a surplus in nearly 30 years. In 1999, the federal government had a surplus of \$120 billion. This has dramatically changed the debate over the budget process; the deficit-reducing conventions of the 1990s are being replaced by a new set of ideas for fiscal policy. Although spending cuts have also helped in deficit reduction, a dramatic increase in federal budget revenues has been the most important reason for the reduction of the deficit.

PASSAGE OF THE 1993 BUDGET RECONCILIATION BILL

Bill Clinton, unlike most other recent presidents, demonstrated that he would immerse himself in the details of the budget. Despite the fact that it was not an issue that he emphasized while running for the presidency in 1992, Clinton also demonstrated an early effort to reduce the deficit. It was widely held that if

Clinton was to get a handle on the deficit as he promised, he must do so in the first year of his presidency, when his political capital was at its peak.

In a joint session of Congress on February 17, 1993, President Clinton unveiled his budget proposal, which included deep spending cuts but relied overwhelmingly on tax increases to bring the deficit downward. At the same time, Clinton proposed to quickly boost short-term job creation by pumping billions of dollars into new spending programs. Clinton's deficit-cutting plan was the largest in history, proposing to save nearly \$500 billion over 4 years. Of that amount, roughly two-thirds would go to reduce the deficit, while another third would be used to pay for increased job creation and long-term investment spending, making net deficit reduction at the end of the 4 years of the plan about \$325 billion.

The deficit-reduction package proposed a cut of \$493 billion over 4 years, \$247 of it coming from spending cuts and \$246 billion from tax increases, almost exactly a 1-to-1 ratio. The ratio of tax increases to spending cuts quickly emerged as the major conflict point in congressional reaction to the plan. Republicans and conservative Democrats were upset that the ratio of cuts to taxes was much less than the 2-to-1 ratio that Panetta had advocated during his confirmation hearings. Although the deficit-reduction plan made notable spending cuts, its heavy reliance on tax increases displays the difficulties the Clinton economic team had coming up with acceptable spending cuts.

Clinton's call for a tax increase was a direct repudiation of the economic philosophies of his two Republican predecessors. By aiming the taxes primarily at corporations and the well-off, Clinton is suggesting that the programs of Ronald Reagan and George Bush, which were designed to stimulate economic growth through tax cuts, came at the price of high deficits. Clinton believed that he could convince the American public—and a majority in Congress—that the economic expansion of the 1980s held negative consequences in the long run. Clinton proposed to raise most of the new revenue with an array of higher taxes on upper-income Americans and corporations, including \$126.3 billion over 6 years, mainly through a new top income tax bracket of 36% and a surtax

on income over \$250,000. Overall, more than one half of the new taxes were projected to fall on families making more than \$200,000 a year.

President Clinton's proposed budget faced its biggest obstacle in Congress with the vote on the Budget Reconciliation Bill. The budget resolution only locked in the broad deficit-reduction numbers, but left virtually all the specifics to the reconciliation process. The reconciliation bill was designed to reconcile tax and spending policy with deficit-reduction goals outlined in the budget resolution. The measure was the heart of Clinton's plan to reshape the nation's economic policy.

In the end, Clinton's economic plan emerged victorious, although just barely. The Omnibus Budget Reconciliation Act was approved in August 1993, without a single vote to spare in either chamber: it passed 218-217 in the House and 51-50 in the Senate (with Vice President Al Gore making the tie-breaking vote). The measure passed without any Republican votes, the first time in postwar congressional history and possibly the first time ever that the majority party has passed major legislation with absolutely no support from the opposition.

With the Republicans unwilling to compromise and unable to drive the process themselves, the struggle to get the antideficit package through Congress was exclusively one of rounding up enough Democrats. This was the case even though the proposal had something to offend almost every Democrat; conservatives were uncomfortable with the entire range of tax increases, and liberals were uncomfortable with some of the \$87 billion in cuts over 5 years in entitlement spending programs such as Medicare and Medicaid as well as the \$102 billion in cuts over 5 years in appropriated spending. Liberals and the Congressional Black Caucus, however, endorsed the bill early on, making moderates and conservatives the critical swing votes.

Overall, the Clinton economic plan was expected to shrink, but not eliminate the deficit. Annual deficits under the law were expected to be around \$200 billion and because nearly \$500 billion in deficits was to be eliminated over 5 years, the national debt was expected to rise by "only" \$1.1 trillion.^[1] Since enacted, the plan has reduced the deficit for the first time since the 1970s; 2000 was the eighth consecutive year with a declining deficit—a postwar record. More spending reductions and tax increases are needed to tackle the deficit problem in earnest, but the Clinton plan was a step in the right direction. The plan not only marked a major step away from the low-tax and high-deficit policies of his Republican predecessors, but also from the spending-oriented policies of Democratic Congresses. By embracing the plan, if only barely, congressional Democrats rolled the dice for their political futures,

betting that deficit reduction would improve the economy in the long run and improve their reelection prospects.

POLITICAL CONSEQUENCES OF THE 1993 BUDGET RECONCILIATION BILL

The difficulty of the passage of the 1993 Budget Reconciliation Bill demonstrates the degree that Members of Congress fear the electoral retribution for raising taxes. The anticipation of retrospective voting helps to focus legislative attention on the possible electoral consequences of voting to raise taxes even when they hear little from their constituents in advance of congressional action. Even though there was little evidence that members of Congress incurred significant electoral risks for approving new taxes prior to 1993,^[2] policy makers fear the potential consequences of voting for increased taxes.

As a result, Congress plans carefully for tax increases. They plan the timing, devise, and use processes that insulate them from interest groups, strategize to minimize the pain to voters, and build coalitions of support.^[3] Congress's aversion of risk extends to the type of taxes members of Congress are willing to raise when revenue must be increased. As experiences at the state level have shown, it is much more difficult to impose new taxes than to raise existing ones.^[4] This means that Congress tends to refuse to repeal established taxes because of the fear of losing revenue. At the same time, it does not want to introduce new taxes because a new tax can potentially have an unfavorable affect on the electorate. By making few changes, members of Congress keep past laws in effect as the best means of minimizing the political costs of taxation while maximizing tax revenue.^[5] By treating tax policy as a "non-decision" rather than as a matter of legislative choice, informal rules and behavioral patterns ensure that a policy already in place will continue in effect without new explicit decisions being made.^[6]

The fact that not a single Republican in either chamber voted in support of the 1993 budget indicates the degree of by which partisanship and ideology dominated the vote on the bill.^[7] Republicans saw the Clinton budget as a means of raising taxes and undoing the Reagan legacy that they revere. Democrats, although they were not enamored with the tax increases, believed that it was important to support the first Democratic president in 12 years. The result was a divide over the proposal that was along partisan and ideological lines. By embracing the plan, if only barely, congressional Democrats gambled on their political futures, betting that deficit reduction would improve the economy in the long run and improve their reelection prospects. Because they passed the

budget package without a single Republican vote, Democrats hoped that they would get sole credit for future success of the bill.

Democrats were buoyed by polls suggesting public support for deficit reduction, even at the cost of higher taxes. It was hoped by congressional Democrats that citizens may have come to realize that real deficit reduction would necessitate more taxes. After the Clinton budget was passed by Congress, citizens were asked about the budget deal. Despite the bitter controversy over rising taxes, Americans were surprisingly willing to pay for the new levies called for in the budget plan. According to a poll taken after the budget plan was implemented, 52% of the respondents said that they did not object to new taxes, and only 47% said that they objected. Of those who also claimed that the American public was partly to blame for the deficit, 60% said that they did not object to the tax increases (Table 1).

Democratic optimism that the 1993 budget would not be a political hindrance, however, evaporated as the 1994 congressional elections approached. As the 1994 elections proved, the Democrats' support of tax increases did not help them politically. Despite a relatively healthy economy in November 1994, Democrats who supported for the Clinton economic plan did

poorly in the midterm elections (although Democrats who voted against the 1993 Budget Reconciliation Bill fared just as poorly). More tellingly, not a single Republican incumbent lost a reelection bid for Congress in 1994. Even in an environment wherein voters were calling for deficit reduction, voting to raise taxes or cut programs proved to be a politically hazardous move for members of Congress.

DEFICIT REDUCTION SINCE THE PASSAGE OF THE 1993 BUDGET RECONCILIATION BILL

To reduce a budget deficit, policy makers must either raise taxes, cut spending, or hope that economic growth leads to increased revenues. The political battles over deficit reduction have tended to center on the degree by which spending should be reduced or taxes increased. Republicans tend to favor spending cuts, whereas Democrats, in an effort to prevent cuts of federal programs they favor, often support tax increases as a better option.

Overall, the 1993 Budget Reconciliation Bill was expected to shrink, but not eliminate, the deficit. Annual deficits have shrunk dramatically since the measure was enacted; indeed, they have shrunk much more than anyone, including the Clinton administration, predicted. The plan has reduced the deficit for the first time since the 1970s; 2000 was the eighth consecutive year with a declining deficit/increased surplus—a postwar record (Table 2). In a development that would have been believed impossible just a few years ago, there were budget surpluses in 1998–2000, the first time since 1969 that the federal government has not run in the red. Without question, since the 1993 Budget Reconciliation Bill was passed, federal budget deficits have been remarkably smaller than they had been the previous two decades.

The success of the 1993 budget at bringing in more revenue has been the most important reason for the elimination of the budget deficit. Since the 1993 budget was passed, budget revenues have significantly increased; in fact, they have increased much more than anyone, including the Clinton administration, expected. The reason for the significant increase in revenues is the combination of the progressive nature of the tax increases—they raised taxes on upper incomes—coupled with a strong economy. Those with higher incomes have done especially well since the 1993 budget was passed. Those making more than \$100,000 a year did particularly well in the economic environment of the mid-1990s; as a result, they paid more federal taxes at a higher tax rate. Since 1992, the last year the highest income tax rate was 31%, the federal government has gained substantially more revenue from those with higher incomes. The individual

Table 1 Willingness to pay new taxes

Question: Do you object to paying the new taxes, or not?		Yes	No	No opinion
Sex	Total	47	52	1
	Male	45	54	1
	Female	48	51	1
Education	College postgrad	33	67	0
	College graduate	41	59	0
	College inc.	77	20	3
	No college	46	53	1
Politics	Republicans	66	33	1
	Democrats	26	74	0
	Independents	49	50	1
Ideology	Liberal	32	68	0
	Moderate	39	61	0
	Conservative	59	40	1
Income	\$50,000 and over	46	54	0
	\$30,000–49,999	47	52	1
	\$20,000–29,999	41	59	0
	Under \$20,000	49	50	1

Source: Gallup poll (August 8–10, 1993; Survey GO 422004, Q.19). Based on national survey of 799 interviews. Poll number 335 of 1993.

Table 2 The federal deficit and the national debt (in billions of \$)

Year	Federal deficit	Total public debt	Debt as % of GNP
1970	12	283	28.1
1975	69	395	25.4
1980	61	710	26.1
1982	146	920	28.6
1984	170	1,300	34.0
1986	207	1,737	39.8
1988	146	2,051	41.4
1990	221	2,411	42.4
1992	290	2,999	48.8
1994	220	3,432	50.2
1996	107	3,733	49.9
1997	22	3,771	47.3
1998	+63	3,717	44.3
1999	+124	3,633	39.9
2000	+232	3,409	34.9

Source: Congressional Budget Office.^[9]

income taxes received by the federal government from those making over \$100,000 increased over 108% in the 5 years after the tax provisions of the 1993 budget became law; from those making between \$50,000 and \$100,000, income taxes received increased by 36%; and total income taxes received from those making less than \$50,000 actually decreased by almost 2% over the same time period. Millionaires, however, contributed more than twice as much to the federal treasury in 1997 than they did in 1992, the year prior to the passage of the 1993 Budget Reconciliation Bill. Thus, the progressive nature of the tax increases has played a significant role in the reduction of the deficit.

The reason for the significant increase in revenues since the passage of the 1993 Budget Reconciliation Bill is the combination of the progressive nature of the tax increases of the 1993 budget—coupled with a strong economy, in which those with upper incomes have tended to do particularly well. Deficit reduction has unquestionably been considerably aided by a healthy economy. The fact that the economy has performed so well after the 1993 Budget Reconciliation Bill was enacted, however, is a strong defense of the success of the bill. Congressional Republicans were adamant that the 1993 budget's tax increases would result in an economic downturn. As Representative Dick Armey (R-TX) stated, the 1993 budget is a recipe for disaster. "Taxes will go up. The economy will sputter along. Dreams will be put off, and all this for the hallow promise of deficit reduction and magical theories of lower interest rates." Newt Gingrich, who was House Minority Whip at the time, predicted

that the bill would lead to "a job-killing recession." The health of the economy after 1993, however, indicates that tax increases on upper incomes are not necessarily a hindrance to economic growth and are a legitimate mechanism by which to balance the nation's budget.

CONCLUSION

Congress has looked toward the president to lead the process of forming the national budget ever since the Budget and Accounting Act of 1921 made the president responsible for submitting a budget proposal to Congress. Yet, for the most part, presidents have tended to focus on other priorities. Many decisions a president could try to implement to improve the budget, such as reducing the deficit, would be wildly unpopular and therefore damage the president's standing with the public. Recent presidents have tended to show little willingness to expend their own popularity by making difficult decisions on the budget.

After he was elected president, Clinton proved to be an exception to this trend. Clinton proposed a politically risky budgetary blueprint that relied heavily on increased taxes. Without a vote to spare, Clinton was able to get Congressional approval of his budget blueprint in an extremely partisan environment. Republicans attacked the plan mercilessly, while most Democrats reluctantly supported the plan in hopes that it would help to reduce the national deficit.

The fact that taxes may be necessary, but at the same time unpopular, helps to explain the Clinton administration's lack of credit for the success of the 1993 Budget Reconciliation Bill. The elimination of a federal budget deficit may be Clinton's greatest accomplishment as president. In the years since the 1993 budget was enacted, the economy has flourished and the federal budget deficit eliminated. The Clinton administration, however, has failed to reap the benefits of the achievements of his budgetary blueprint. According to a Gallup poll conducted in January 1998, 32% said that Clinton deserved no credit at all for balancing the budget. When asked who deserved more credit for reducing the deficit, the number who said Clinton was virtually even with the number who claimed Republicans in Congress.

Increasing taxes, although concentrated mainly on upper incomes, are hard to sell as being positive, even if the goal is to reduce the deficit.^[8] Deficit reduction requires one to think in terms of the long run, which is difficult for voters and for policy makers to do. Tax increases, however, are evident immediately, and much more of a factor at the ballot box.

The passage of the 1993 Budget Reconciliation Bill leaves a mixed message. It has been remarkably

successful in doing what it was passed to do: reduce the deficit. But many of those who supported the bill paid a heavy political price for their support. As Speaker Thomas Foley (D-WA) argued shortly before the House vote on the 1993 budget in an attempt to rally Democrats behind the bill, “Tonight is the time to put away the old, easy ways. Tonight is the time for responsibility.” Yet, Foley became the first Speaker of the House to lose a reelection bid since the Civil War. A heavy price to ask a lawmaker to pay for deficit reduction.

REFERENCES

1. Lee, R.; Johnson, R. *Public Budgeting Systems*, 5th Ed.; Aspen Publishers: Gaithersburg, MD, 1994; 205 pp.
2. Douglas, A.R. *The Logic of Congressional Action*; Yale University Press: New Haven, 1990; 194 pp.
3. Rubin, I. *The Politics of Public Budgeting*; Chatham House: Chatham, NJ, 1990.
4. Hansen, S. Extraction: the politics of state taxation. In *Politics in the American States*; Gray, V., Jacob, H., Vines, K., Eds.; Scott Foresman and Company: Glenview, IL, 1983; 335–358.
5. Rose, R. Maximizing tax revenue while minimizing political costs. *J. Public Policy* **1986**, 5 (3), 289–320.
6. Schick, A. *The Capacity to Budget*; The Urban Institute: Washington, DC, 1990.
7. Fisher, P. The prominence of partisanship in the congressional budget process. *Party Polit.* **1999**, 5 (2), 225–236.
8. Fisher, P. Political explanations for the difficulties in congressional budgeting. *Soc. Sci. J.* **1999**, 36 (1), 149–161.
9. Congressional Budget Office. Historical Budget Data, www.cbo.gov, accessed January 31, 2002.

360-Degree Performance Appraisal Systems

Neil M. Boyd

School of Education and Behavioral Sciences, Pennsylvania State University, Capital College, Middletown, Pennsylvania, U.S.A.

INTRODUCTION

A 360-degree performance appraisal system is a multi-source assessment approach that taps the collective wisdom of those who work closely with an employee. The employee and their supervisors, colleagues, direct reports (subordinates), internal customers, external customers, and others may be part of the evaluation process.

Proponents of the 360-degree feedback approach offer it as a “progressive” means of conducting performance appraisal. Many contend that 360-degree feedback systems and other forms of multisource or multirater assessment methods in organizations have evolved from an innovative “nice-to-have” technique administered only to the most senior levels to a “must-have” tool for integration into overall performance and human resource management strategies. These systems appear well suited for the flexible, team-based, change-oriented organizational cultures of many organizations today.

360-Degree systems are gaining popularity because they tend to reduce the problems of previous generations of assessment methods. 360-Degree appraisal moves the manager (as appraiser) back into the “comfort zone” as she or he is now only one among a number of assessors. It greatly reduces the problems of central tendency, positive skewness, and “halo effects” in ratings that plague the boss-subordinate approach. 360-Degree appraisal reduces defensiveness on the part of the appraisee because there are a variety of assessors, feedback is presented as more balanced, and it is said to have greater face validity. It requires organizations to resolve the issue of what is meant by job effectiveness and what behaviors are causally related to it. It recognizes that subordinates are best placed to assess “leadership” or “people management” skills. The technique is helpful in defending legal challenges of the outcome of appraisals, it is presented as important in meeting the demands for employee empowerment and involvement, and finally, it is a useful tool in tapping employee opinions and attitudes.

HISTORY OF MULTISOURCE FEEDBACK AND 360-DEGREE APPRAISAL

One fact regarding 360-degree performance appraisal seems clear; the practice is not new. Contemporary 360-degree methods have roots as early as the 1940s, however, there is some disagreement regarding the exact genesis of the technique. For example, Fleenor and Prince^[1] argue that World War II German military assessment centers first recognized the value of measuring performance from multiple perspectives. By contrast, Edwards and Ewen^[2] contend that the assessment center was developed by British military intelligence in the early 1940s and was the forerunner to multisource assessments. Despite these disagreements, one point that most scholars can agree on is 360-degree performance appraisal has historical roots within a military context.

During the 1950s and 1960s, this trend continued in the United States within the military service academies. At the United States Naval Academy at Annapolis, the midshipmen used a multisource process called “peer grease” to evaluate the leadership skills of their classmates. The peer grease process, which is still used today, has a significant impact on class standing and postgraduate assignments.

In the corporate world during the 1960s and 1970s, multisource systems were used in job evaluation committees, hiring and internal selection panels (including executive selection), promotion boards, and talent selection. Organizations like Bank of America, United Airlines, Bell Labs, Disney, Federal Express, Nestle, and RCA experimented with multisource feedback in a variety of measurement situations.

Despite the fact that multisource feedback systems had been used for several decades, the use of 360-degree instruments for individual performance assessment was rare prior to the 1980s. In previous decades, performance appraisals were typically conducted in a top-down, supervisor-driven framework that was consistent with the hierarchical structure of most organizations. In the mid-1980s, organizations like duPont, Fidelity Bank, Arizona State University, and Monsanto began experimenting with multisource assessment

for development and appraisal. However, 360-degree appraisal did not replace supervisory decisions regarding performance measurement. At most, they were used as an aid in the performance measurement process.

But things have changed. Since the early 1990s, 360-degree appraisal techniques have gained a wide base of acceptance, and the idea has emerged that 360-degree feedback can be useful for performance appraisal and decision-making purposes.

THE PROCESS

360-Degree performance appraisal is employed to determine performance evaluation and/or to facilitate employee development. The specific design and implementation procedures are usually guided by the organization's goals. Within 360-degree systems, employees assess themselves, and in turn are evaluated by a number of pertinent "stakeholders." Typically, the person to be assessed reports whom he or she works with directly. This list would include subordinates, peers, and superiors, and may include customers/clients and suppliers (Fig. 1).

The supervisor responsible for the evaluation compiles a list of individuals based upon this information and upon her or his own knowledge of the employee's work responsibilities. The supervisor may

also approach others for suggestions concerning whom to contact. Ideally, the supervisor and all those who work directly with the employee rate the employee along a number of predetermined criteria (e.g., employee aptitude, task expectations, interpersonal skills, leadership skills). Usually, these criteria are considered in light of specific workplace behavior rather than on more abstract traits. In addition, the employee rates himself or herself according to the same standardized criteria. Sometimes raters provide information anonymously. Sometimes this rating is done with the aid of five- or seven-point Likert-type scales, sometimes it is done with narratives, and sometimes with both. Ratings are then analyzed in terms of relative rank (especially extreme rankings) and disparities in ratings. Extremely low scores, and disparities between raters, are used to identify areas to be addressed by the employee. Of particular note are great disparities between employee self-ratings and those of others.

After analysis, the ratings are presented to the employee. Typically, the supervisor and employee discuss the ratings, consider differences in ratings, decide upon goals to be pursued, and develop an "action plan" to help the employee meet her or his goals. The action plan may entail participating in professional development programs such as workshops and training programs; it may entail personal reflection and intentional behavior modification, or it may entail a combination of the two.

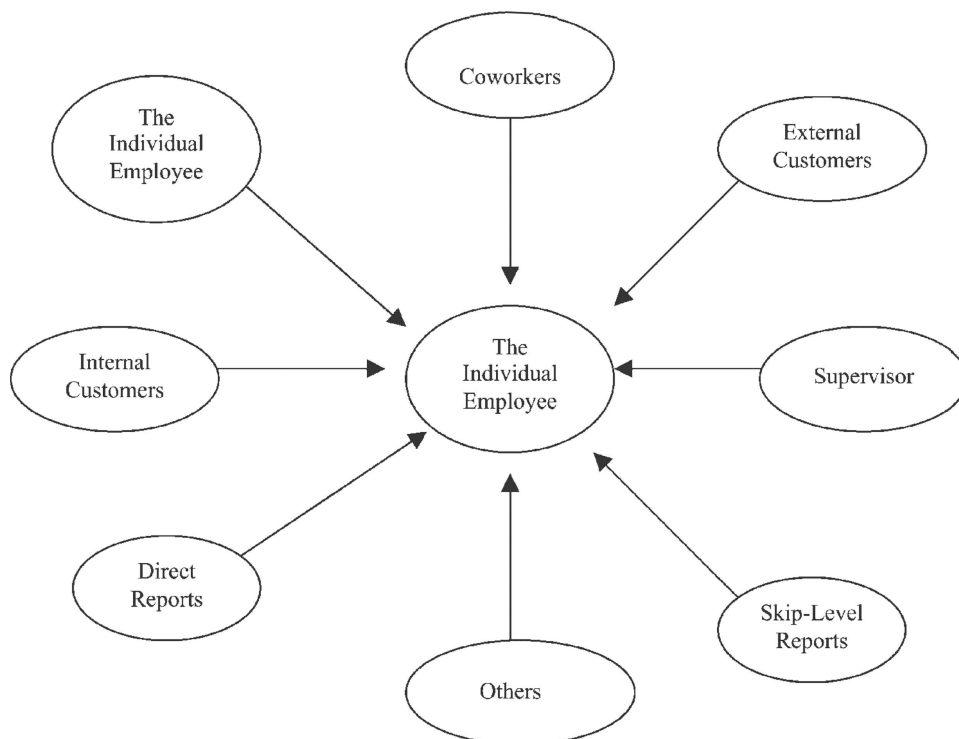


Fig. 1 360-Degree performance appraisal. (From Ref.^[2].)

POSSIBLE PROBLEMS WITH 360-DEGREE SYSTEMS

Despite the many benefits of utilizing a 360-degree approach, several authors have cautioned that like many other performance appraisal systems, 360-degree methodology is not without its problems. Boyd and Kyle^[3] proposed that 360-degree systems need to be designed and implemented by employees instead of controlled by management. An employee controlled process would reduce a number of psychological contract violations, increase organizational justice, and facilitate a democratic ethos in the workplace. Barnes^[4] pointed out that there is a potential for problems with subordinate assessments where employees desire to “get the boss” or may alternatively “scratch the back” of a manager for expected future favors. Also, the additional time and cost might be an administrative burden. Moreover, Antonioni^[5] raises other concerns. He proposed that some organizations may be jumping into implementing 360-degree appraisals without clearly defining the mission and scope of the appraisals. To prevent potential pitfalls of the system, the organization must be prepared to design a 360-degree appraisal process that supports specific missions and goals that are central to the organization. Perhaps most importantly, organization members must decide if the 360-degree approach is to be used for employee development or assessment (or both).

In addition to these concerns, Edwards and Ewen^[2] note a series of common pitfalls with 360-degree methods. The number one reason they fail is that organizations attempt to assimilate the 360-degree method within a traditional survey research scheme. In traditional survey research, investigators attempt to maximize data collection with as many items/questions as possible and with large sample sizes. In the case of 360-degree appraisal, creating measurement instruments with many items will substantially increase non-response errors. In addition, large sample sizes are not typically possible considering that perhaps eight or nine sources will rate an employee's performance. As such, statistical procedures that rely on large sample sizes in order to ensure statistical validity might not be appropriate. For example, if only eight or nine ratings are used, a simple calculation like the mean (average) might contain measurement error. In addition, organizations must consider other issues like safeguarding the process from unintentional respondent rating errors, communicating to and training organizational members on design, implementation, and evaluation of the system, and culture shock that occurs with any system that creates “change.” In short, 360-degree systems need to be designed and implemented with a great deal of thought.

360-DEGREE SYSTEMS IN THE PUBLIC SECTOR AND ITS FUTURE IN THE FIELD

It appears that over 90% of Fortune 1000 firms use multisource feedback systems for employee development or evaluation, yet its use in the public sector seems to be much more limited.^[6,7] It is difficult to assess the actual use of 360-degree performance appraisal systems in the public sector given the paucity of empirical reporting on the topic; however, a 2002 study by Glendinning^[8] indicates that 360-degree feedback has not yet been warmly received. His survey findings of 64 private and public northeast regional human resource professionals showed that 33% were currently incorporating 360-degree methods into their performance appraisal systems. Another study conducted in the public health sector showed that no local public health departments were utilizing 360-degree methods when searching six computerized search engines relating to the field of public health.^[9] The fact that the rates of 360-degree utilization appear small does not mean in reality that it is fractionally employed. It may mean that reporting of its use is lagging behind its actual implementation.

This may be the case considering that the call for 360-degree implementation has been strong in the public management and public administration literature. Since the mid- to late 1990s, many have called for the assimilation of 360-degree performance appraisal into the public sector.^[10,11] The pressure for change in the field of public personnel management has rarely been more intense. As Hays and Kearney^[12] note, “fiscal crises, a brooding and hypercritical public, and stunningly rapid movement in the personnel system's environment have exacerbated the field's traditional problems and fueled renewed calls for reform.” 360-Degree performance appraisal is but one method in the mix of calls for reform.

Much of the literature that challenges traditional performance appraisal methods is normative. Empirical studies, which assess strategy, design, implementation, and evaluation of contemporary performance appraisal methods in public management, are much less visible. Hays and Kearney reported on one study that empirically evaluated modernization within public sector human resource management, and prospects of its future direction. Included in their analysis was a focus on past and contemporary performance appraisal methods. They conducted their analysis by surveying 295 members of the American Society for Public Administration's (ASPA) Section on Personnel and Labor Relations (SPALR), and personnel directors in federal, state, and local government listed in the 1997 membership directory of the International Personnel Management Association (IPMA). Performance appraisal ranked first in importance and its

widespread use was predicted to continue. In addition, they argued that “new performance-assessment techniques such as peer evaluations, client evaluations, appraisals by subordinates, team-based evaluations, and the all inclusive 360-degree approach are anticipated to leap ahead of such traditional approaches as rating scales and checklists.” Based on their findings, a bright future for 360-degree performance appraisal is anticipated in the next decades within public personnel management.

Although the future looks bright for 360-degree implementation, there are challenges that lay ahead. Public management scholars and practitioners will have to study and evaluate the 360-degree method within the context of the public sector with an eye on the fact that it likely will not automatically transfer from the private sector. In particular, 360-degree methods will potentially contend with union, civil service, and legislative interests that pervade in the issue networks associated with public personnel management.

REFERENCES

1. Fleenor, J.W.; Prince, J.M. *Using 360-Degree Feedback in Organizations: An Annotated Bibliography*; Center for Creative Leadership: Greensboro, NC, 1997.
2. Edwards, M.R.; Ewen, A.J. *360° Feedback: The Powerful New Model for Employee Assessment and Performance Improvement*; American Management Association: New York, NY, 1996.
3. Boyd, N.M.; Kyle, K. Expanding the view of performance appraisal by introducing social justice concerns. *Adm. Theory Prax.* **2004**, *26* (3), 249–278.
4. Barnes, P. Making appraisal work in the new millennium. *Manage. Serv.* **1997**, *41*, 14–16.
5. Antonioni, D. Designing an effective 360-degree appraisal feedback process. *Org. Dyn.* **1996**, *25*, 24–38.
6. Atwater, L.E.; Waldman, D. Accountability in 360-degree feedback. *Hum. Resour. Mag.* **1998**, *43* (6), 96–101.
7. Ghorpade, J. Managing five paradoxes of 360-degree feedback. *Acad. Manage. Exec.* **2000**, *14* (1), 140–150.
8. Glendinning, P.M. Performance management: pariah or messiah. *Public Person. Manage.* **2002**, *31* (2), 161–178.
9. Swain, G.R.; Schubot, D.B.; Thomas, V.; Baker, B.K.; Foldy, S.L.; Greaves, W.W.; Monteagudo, M. Three hundred sixty degree feedback: program implementation in a local health department. *J. Public Health Manage. Pract.* **2004**, *10* (3), 266–271.
10. Roberts, G.E. Employee performance appraisal system participation: a technique that works. *Public Person. Manage.* **2003**, *32* (1), 89–98.
11. Roberts, G.E. Perspectives on enduring and emerging issues in performance appraisal. *Public Person. Manage.* **1998**, *27* (3), 301–320.
12. Hays, S.W.; Kearney, R.C. Anticipated changes in human resource management: views from the field. *Public Adm. Rev.* **2001**, *61* (5), 585–597.

Accountability and Ethics: Reconsidering the Relationships

Melvin J. Dubnick

*Departments of Political Science and Public Administration, Rutgers University,
Newark, New Jersey, U.S.A.*

INTRODUCTION

The relationship between “accountability” and “ethics” has long been a concern among students of public administration. Accountability has traditionally been regarded as the means used to control and direct administrative behavior by requiring “answerability” to some external authority. It has deep roots in American constitutional history, and can be linked to the principles implicit in the Magna Carta as well as to our system of checks and balances. In public administration, ethics has most often been associated with standards of responsible behavior and integrity.

Issues about the relationship between the two were central to the Friedrich–Finer debate of the early 1940s—one of two debates that set the intellectual agenda for the field of public administration during the Cold War era (the other being the debate between Herbert Simon and Dwight Waldo).^[1] In a 1940 essay,^[2] Friedrich argued that the traditional means (e.g., oversight and control) for holding administrators accountable were ineffective and unnecessary. It was reasonable, he contended, to defer to the judgment of administrators whose sense of professional responsibility and loyalty could be trusted when they carried out public policy in the national interest. In response, Finer^[3] reasserted the widely held view that, despite the greater sense of professional responsibility among today’s administrators, democracy still requires enhanced public control and direction of administrative agents.

That exchange was—and remains—the classic expression of the conventional relationship between ethics and accountability in public administration.^[4–7] Fundamental to that view is the assumption that the commitment of modern administrators to conduct themselves responsibly (i.e., ethically, in accord with “democratic morality”) was not sufficient to ensure that the will of the people would be carried out. Accountability, in the form of external (i.e., democratic) constraints and controls, was also necessary. Accountability mechanisms were required to render the decisions and behavior of public officials responsible, not merely in the legal, political, or bureaucratic senses of that term, but also morally.^[8–10] Ethical

behavior, in short, required the presence of external accountability mechanisms in their various forms.

Does accountability foster ethical or morally responsible behavior? Despite considerable scholarship devoted to the examination of efforts to control the bureaucracy through various accountability mechanisms,^[11–20] the existence or effectiveness of the accountability–ethics relationship has yet to be systematically examined. A growing body of work on accountability implies that it has a significant impact on administrative behavior,^[21–31] but none of these studies directly address how accountability affects the ethical standards and strategies adopted by administrators.

BARRIERS TO ANALYSIS

The barriers to such an analysis have been conceptual and methodological. Conceptually, both “accountability” and “ethical behavior” have lacked the “sharpness” and clarity^[32] required for analytical purposes. The concept of accountability has been characterized as expansive^[33] and chameleon-like^[34] as applied in both theory and practice, and the appropriate meaning of ethical behavior has long been the subject of heated debate between utilitarians and deontologists (among others).^[35,36] The lack of progress in conceptualization, however, can be attributed to the institutional bias and normative ambiguity fostered by our methodological approaches to both terms. Accountability and ethical behavior are rarely perceived in behavioral terms outside the institutional contexts that preoccupy students of public administration. As a result, accountability becomes associated with certain institutional forms of oversight and ethical behavior becomes tangled in discussions of codes and legalisms. Also important has been the inability of analysts to put aside the normative “baggage” that accompanies the use of such value-laden terms as “accountable” and “ethical” when describing human behavior.^[32]

We can overcome these barriers by adopting a “middle range” or “concrete theory” perspective^[37–40] that allows us to reconceptualize both accountability and ethical behavior as “social mechanisms”—that

is, in Robert Merton's terms "social processes having designated consequences for designated parts of the social structure."^[41] Applying this approach, we assume a sociological (rather than institutional) view of accountability. The sociological view stresses the forms and functions of accountability as processes (mechanisms) that affect social actors as *situated pressures for account-giving behavior*. In this article, we focus on four general types of mechanisms that demand account-giving responses: answerability; blameworthiness; liability; and attributability.

Similarly, we regard ethical behaviors as social mechanisms constituted as norms and standards of behavior generated as partial responses to the pressures created by accountability mechanisms. In contrast to both Kantian and utilitarian views of ethics that rely on the existence of a priori knowledge or some universal standard, we assume a more naturalistic^[42] and pragmatic perspective that stresses the functional role of ethics in dealing with social dilemmas.^[43-45] Our focus is on the development and nurturing of ethical behaviors and strategies in response to major forms of sociological accountability.

THE SOCIOLOGICAL PERSPECTIVE

Three Approaches to Accountability

Analytically, the concept of accountability can be approached in three ways: historically; institutionally; and sociologically. Historically, the term itself has distinctly Anglican and feudalistic roots, and has only recently emerged as a universal standard of governance.^[46] Mechanisms similar to the modern concept of accountability, however, do have ancient roots in Athenian democracy.^[47,48] More than historical curiosities, both the Anglican and Athenian views of accountability offer insights into modern accountability and provide standards by which to assess the various approaches that emerge from sociological analyses. For our purposes, however, we focus attention on those views derived from the contemporary study of social relationships.

Institutional perspectives approach accountability as formalized means of feedback and control established with governance structures of states and corporate entities.^[49] This view covers a range of phenomena, from constitutions to financial reporting standards, and has been the dominant perspective in the study of bureaucratic accountability.^[4,6,19]

Sociological perspectives focus attention on accountability as a type of social act that is part of a larger class of social processes or mechanisms dealing with the need to repair or overcome damaged relationships resulting from "unanticipated or untoward

behavior."^[50] That group of mechanisms includes confessing to the action or seeking forgiveness,^[51] subjecting oneself to punishment or retribution,^[52,53] engaging in "good works" to make amends for the damage done or seeking the restoration of one's good name,^[54] and associated behaviors that stand as substitutes for (or complements of) giving an account of one's behavior.

Giving of Accounts

Specifically, the social "giving of accounts" occurs when one must justify or provide excuses for the action that resulted in the faux pas or untoward act that is the focus of attention.^[50] One is essentially not denying either active or passive involvement in the particular indiscretion, but is engaging in an effort to explain why one ought not to be held either fully or partly responsible. Justification involves accepting responsibility for the act, but denying the immorality or untowardness of the behavior's consequences. Excuses, in contrast, do not deny the negative consequences of the acts in question, but argue that the account giver had no choice but to act badly and was therefore not responsible.^[50,55] In this sense, an individual is "accountable" to the extent that there exists an expectation^[56] within the community that they would provide an explanation for any act regarded as worthy of account giving.

While standing as a distinct type among the class of associate mechanisms related to repairing damaged relationships, account giving may also serve as a necessary complement to the other forms. The expectation to provide a justification or excuse for one's actions may include a demand for an accounting, even if the individual confesses to the act, apologizes, and submits to punishment or undertakes restitution. Consider the case of Scott Waddle, commander of the submarine U.S.S. Greeneville that collided with a Japanese vessel, the "Ehime Maru," south of Hawaii on February 9, 2001. The incident, which led to nine deaths, was the subject of a military court of inquiry; by April, Commander Waddle had accepted "full responsibility and accountability" for the accident and made several widely publicized apologies. But in this instance, and others like it, the public admission of guilt and expression of regret and apologies did not preclude the requirement for an accounting of what took place. Thus, with each public appearance, Commander Waddle had to relive the incident and provide an account of why he acted as he did.

In contrast, there are some acts of omission or commission considered so despicable that no amount of excuses or justifications would be acceptable. For example, Adolph Eichmann, Albert Speer, and other Nazis offered defenses of their actions based on

Table 1 Types of accountability

Expectations	Structured	Emergent
Related to persons	Type A: Answerability	Type B: Blameworthiness
Related to situations/events	Type C: Liability	Type D: Attributability

accounts that they were just “following orders,” carrying out their civic and legal “duties,” or (in the case of Speer) unaware of what was taking place despite their prominent roles in the German war effort.^[57,58] No account giving in the form of excuses or justifications was perceived as acceptable.

Typology of Account-Giving Mechanisms

With its focus on actions taken in response to social expectations, the sociological perspective can generate a useful typology of account-giving mechanisms. Expectations relevant to providing an explanation for one's behavior can be viewed along two dimensions. First, the expectations can be related to specific actors or can focus on situations. Second, the expectations setting or environment can be highly structured and relatively stable or emergent and subject to fluctuation. Table 1 illustrates the resulting forms of socially relevant accountability emerging from relating these two dimensions.

In the following sections, I describe the distinctive features of the resulting four types of account-giving mechanisms and discuss the relevant literature associated with each. I also briefly consider the implications of each for generating ethical behavior mechanisms among those being held accountable. As we see, accountability and ethics are closely related both analytically and empirically.

TYPE A: ACCOUNTABILITY-AS-ANSWERABILITY

Demanding Answers

Role-specific expectations for account giving are those found in highly structured social relationships where the tasks and obligations of individuals are either clearly articulated (i.e., formalized) or so “institutionalized” (in the sense of “regularized”) that they are perceived as inherent to the position a person is occupying.^[59,60] We label this form of accountability “answerability” to stress the idea that persons who fall in this category are expected to respond to calls for giving an account upon demand. In short, that expectation comes with their role. This conceptualization

of accountability is among the most commonly used, and has been central to contemporary studies of democratic and organizational accountability.^[22,61]

Within democratic theory,^[62,63] accountability-as-answerability indicates a relationship between the governors and the governed, involving mechanisms requiring the former to inform the latter of actions taken on their behalf (i.e., in their “interest”), and allowing the latter to judge and take action against the former based on that information. The specific mechanisms involved range from accountability through elections^[64] and mediated procedures (e.g., a “vote of no confidence” in parliamentary systems)^[65] to the creation and strengthening of “horizontal” institutions (e.g., courts and ombudsman offices).^[49] Accountability-as-answerability has also played a central role in classical and contemporary organization theory, where the central problem has been to develop mechanisms (e.g., chain of command, limited span of control) and related strategies for creating and sustaining cooperative social systems.^[66–69]

Answerability Research

There have been several streams of research associated with accountability-as-answerability mechanisms, each applying a distinct model of the social relationship involved and generating a different picture of the ethical implications of answerability. One stream models answerability as a relationship between accountable individuals and the audience to which they must provide the account. In a series of articles published in the 1980s, Tellock and his colleagues found that being told one must provide a justification for a decision had a significant influence on the individual's approach to each decision.^[70–73] Moreover, individuals put in such situations were motivated to seek the approval of their perceived audiences, leading Tetlock to develop an “accountability theory” emphasizing the view of answerable individuals as “intuitive politicians... whose primary goal is to protect their social identities in the eyes of the key constituencies to whom they feel accountable.”^[74]

In contrast, a principal agent theory stream provides a contractual context within which answerability mechanisms and relationships are used to overcome the inherent “moral hazard” and “selection” problems

principals face when relying on an agent.^[75–78] Here, the answerability mechanisms are regarded as structural responses to dilemmas created by the effort of rational principals attempting to control rational agents, each pursuing their own self-interest.

A third research stream, that of the “new institutionalism,” regards organized behavior as dominated by rules and institutional norms, and would treat answerability as part of that rule structure.^[79–81] The accountable individual is a rule follower whose actions are governed by what March and Olsen have termed, the “logic of appropriateness.” Under this logic, accountability-as-answerability is dominated by three questions: what kind of situation is this?; who am I (i.e., what is my identity or role)?; and what is the rule I should follow given my identity and the situation?^[82]

Answerability Ethics

The three perspectives on accountability-as-answerability present us with alternative models of the accountable actor, and thus imply three models of ethical behavior mechanisms. The ethics of Tetlock’s intuitive politician are driven by the psychological need to be accepted by the “audience”—in the case of elected officials, the electorate or the people, and in the case of a public administrator, the client, supervisor, or stakeholder. This can result in an approach that judges the ethicality of actions on the basis of their strategic value.^[83] The ethics of the rational self-interested agent is shaped by individuals’ calculations of what actions will best suit their interests, regardless of whether those interests are focused on the accumulation of power, the accumulation of wealth, or some other goal.^[84,85] The rule followers’ ethics are primarily shaped by role expectations and the individual’s social identity—a fact that leads them to select the most “appropriate” action among alternatives based on assessments of what best fits that identity and set of expectations.

What these three ethical roles share is a reliance on exogenous signals or structures. The intuitive politician is constantly monitoring the audience for signals about appropriate behavior, whereas the rational agent is engaged in ongoing calculations about that behavior that will provide the best payoff. The rule follower, in the meantime, is always engaged in balancing the expectations associated with their social identity with the demands of the situation. All three look outward as they make their ethical choices, and all three reflect a concern for the consequences of their actions. In that sense, Type A accountability tends to make administrators rely on various forms of ethical mechanisms stressing the consequences of their actions vis-à-vis

the expectations of those to whom they are answerable. This stress on consequences will, in turn, lead to the adoption of ethical strategies drawn from teleological schools of ethics,^[86] providing logical support to the various models of bureaucratic behavior applied or implied in many positive, political, and organizational theories. Among positive theories of bureaucratic behavior, mechanisms rooted in utilitarian ethics (act, rule, and motive) are the assumed position for administrator actions. Similarly, political theories of bureaucratic behavior emphasize mechanisms that enhance the influence and power of agencies, whereas organization theories focus on the striving for autonomy, security, and resources. Implied in these models is a set of ethical strategies closely linked to Type A accountability.

TYPE B: ACCOUNTABILITY-AS-BLAMEWORTHINESS

Status-Based Accountability

Although accountability is most often thought of in terms of answerability, it is not always applied to persons in specific and defined roles. Blameworthiness as a form of accountability involves a shift in focus from specific roles and contractual obligations as an agent to one’s perceived social status and membership in some group that has status. Whereas role identifies what tasks and functions you perform, status implies power and influence within the organization.^[87] Put briefly, one is held accountable not because of one’s tasks or formal responsibilities, but because of one’s relative social position or identification with a certain group. Thus, you are held accountable because you are regarded as socially, if not organizationally, blameworthy.^[88]

In its most extreme form, Type B accountability involves “scapegoating”—that is, imposing blameworthiness on a group simply by characterizing it as the source of the undesirable condition.^[89,90] It is, in Peter Gay’s words, the construction of a “convenient Other” upon whom one can visit the blame for any problem—and against whom we can target our anger and aggression.^[91]

But accountability-as-blameworthiness can also take less extreme forms. In military organizations, for example, those at the highest echelons of any unit are deemed blameworthy and accountable for events that occur on their “watch.” This form of accountability differs from mere answerability, which is related to whether one is performing one’s specific role. Here, the primary issue is one’s status as the commanding officer that is most important.

A relevant case is that of Rear Admiral Husband E. Kimmel, Commander-in-Chief of the Pacific Fleet at

the time of the Japanese surprise attack on Pearl Harbor on December 7, 1941. There is no question that the devastation wrought by Japan's forces was causally responsible for the damages suffered on that "day of infamy." Yet, both Admiral Kimmel and Lieutenant General Walter Short, who was Commanding General of the local Army contingent, were officially blamed and sanctioned for the state of "unreadiness" at Pearl Harbor on the morning of the infamous attack, and their careers were ended despite their potential value as commanders in the war that followed.^[92] Similarly, in a statement issued in July 1997, then Secretary of Defense William Cohen held Brigadier General Terry Schwalier "personally accountable" for the lax force protection that led to the death of 19 American service personnel at the Khobar Towers bombing a year earlier.^[93] The facts in both cases indicated that there was more blame to go around than that visited on Kimball, Short, or Schwalier, but that is not the way accountability operates in the military. "In the military," observed Judith Shklar, "responsibility has to be personalized at the highest level of the organization since its system of command is built on the principles of obedience and reliance . . . We cannot afford to be philosophically discriminating when our security depends on maintaining the principles of hierarchical responsibility for victory and defeat, especially the latter" (p. 63).^[94]

Blameworthiness in Public Service Cultures

While accountability-as-blameworthiness is an accepted part of military culture [and can be found in some organizations that model themselves after the military, (e.g. some police forces)], it is rarely evident in other formal organization contexts. However, it is a pervasive presence in the extraorganizational contexts of public service. Our government and its agents operate in a culture where officials are implicitly held blameworthy for possible misfortunes, regardless of causal responsibility.^[88,94] A hurricane is an act of nature, but government meteorologists will be held accountable for forecasting its exact point of impact as well as expected intensity, and local emergency preparedness offices will be blamed if proper steps were not taken to deal with the storm's impact. In the aftermath of the September 11, 2001 attacks on the World Trade Center and the Pentagon, issues were raised about the competence and culpability of the U.S. intelligence community as well as security at America's airports. Within 2 weeks of those tragic events, the media began to castigate the policies of the U.S. Immigration and Naturalization Service for its inability to track aliens who remained in the country illegally. Like other public sector agencies, they were blameworthy;

unlike others, they were called into account for their implied role in facilitating the devastating attack.

What is involved in this form of accountability comes close to what is traditionally termed "moral responsibility." As Marion Smiley notes, in both its Christian and Kantian forms, moral responsibility implied the existence of some higher or external authority. Under those forms, blameworthiness was a matter of whether one was in a position to violate the norms of those universal standards that constituted morality. But in discussing accountability-as-blameworthiness, we are focused on the social construction of blameworthiness and the social practice of blaming.^[95] We are blameworthy as a result of our status within a community (e.g., the mayor) or organization (e.g., the general), not due to any specific task responsibilities or actual authority. We are blameworthy because we are members of a group whose members are regarded as blamable (e.g., bureaucrats, politicians), not because anyone in that group could really be an effective causal agent.

Type B Ethical Responses

The ethics of those facing Type B accountability is necessarily pragmatic in the broadest sense of that term and can take three forms beyond the mere passive acceptance of one's fate as a blameworthy actor. First, accountability-as-blameworthiness can be approached instrumentally. Faced with the social reality of blameworthiness, the accountable individual should use the situation to deal with the challenges within that context. The heads of agencies that have been subjected to severe criticism for past performance can use the opportunity to strengthen the organizations' future capacities. In response to their experiences as field-level soldiers in Vietnam, the officers who rose through the military's ranks from the mid-1970s to the 1980s developed ethical strategies that literally transformed the role of the armed forces in U.S. policy as well as the way the military was perceived by the public and perceived itself.^[96] An equally dramatic transformation of the ethical and operational norms of the Federal Emergency Management Agency (FEMA) took place under James Lee Witt, who assumed his position in April 1993 in the wake of severe criticism for FEMA's handling of recent disaster relief efforts.

A second approach is to assume a commitment to a deontological moral standard^[97] that would meet with widespread approval within the community or organization. The call for public administrators to focus on "service" rather than "steering,"^[98] to engage in the pursuit of the "public interest,"^[10,99] to adopt a professional standard of social equity,^[100,101] or to

“refund” itself in the regime values of the constitution^[102,103] are part of this effort to fill the moral void implied by the condition of blameworthiness.

The third approach is to engage in an “identity shift” by dissociating oneself from the blameworthy status or group. Under the view of social constructivism, just as certain groups are regarded as blameworthy, so other groups are inherently praiseworthy. In lieu of a tag such as “bureaucrat” or government employee, one can stress a professional identity such as city planner, park ranger, or law enforcement official. These identity shifts, in turn, have implications for the type of accountability and ethical standards to which one is subject. Of course, as social constructions, the blameworthiness of these identities can vary from place to place and over time. Police officers in a small midwestern town will face a different set of community expectations than their peers in New York or Los Angeles; similarly, the status and perception of an FBI agent in the 1990s was quite different from that of the 1950s.

Common among these three Type B responses is the establishment of credibility for public administrators as autonomous ethical actors. The post-Vietnam era commanders focused on establishing and articulating basic rules for the use of military force—rules that sometimes led them to openly resist putting U.S. personnel in harm’s way when other options were available. The military professionalism represented by the likes of George Patton were replaced by the model provided by Colin Powell and his cohort who served and suffered as field officers during the Vietnam War. The call for the public service professional to give priority to general ethical commitments is also a step toward asserting an ethically autonomous position, as is the effort to create a distinct professional identity among the public service professions. Each of these is a means toward the realization of Carl Friedrich’s model of the responsible administrator.

TYPE C: ACCOUNTABILITY-AS-LIABILITY

Law-Like Situations

In contrast to accountabilities based on the expectations related to persons, Types C and D accountabilities focus on the structure of the situation in which actors operate. Although posited as distinct types, accountability-as-liability and accountability-as-attributability would more appropriately be described as forms found at opposite ends of a continuum. At one end are highly structured contexts in which expectations are well defined and clear to all; at the other extreme are situations so vaguely defined that they

are almost devoid of stable expectations. Our current focus, Type C accountability, falls on the “highly structured” section of that range.

As implied in its name, accountability-as-liability is closely associated with a legalist view of the world in which actions are guided and assessed according to rules that carry sanctions for non-compliance.^[104] Unlike answerability (Type A), where expectations for accountability derive from the actor’s institutional role or organizational position, liability stresses the requirements of the structured situation. The nature of that structure can vary from the strictures of legal requirements backed by state sanctions to implicit contracts between two parties that carry with them rewards for compliance and/or sanctions for violations. At the core of these mechanisms is the idea that law, in a general sense, is an effective way to establish and maintain order and control.^[105,106]

In a strictly formal sense, accountability-as-liability would be limited to shaping one’s actions according to the “letter of the law”—that is, doing what is required by law and avoiding those actions prescribed under law. For Type C accountability, however, the concept of law is much broader, taking on the characteristics of both an institutional setting (i.e., the “rule of law”)^[107,108] and a context for social relationships.^[109–112]

Issues of Liability

The relationships between Type C and other forms of accountability are quite complex for two related reasons. First, ours is an “organizational society” where individuals are constantly identified by their role or status. It is rare to find a structured situation where one is completely without some role identity. Accountability-as-liability, however, tends to focus on the act and its consequences rather than on the actors or their status-derived blameworthiness. In its most stringent form, Type C accountability seeks out the actors who “did the deed” (who were directly, causally responsible) rather than the individual “in charge” or playing some other relevant role. Thus, an individual in a supervisory role/status would not be held accountable *unless* there was both evidence and willingness to link that person to the untoward activity.

For example, in the case of the infamous 1969 My Lai massacre, the U.S. Army believed that it had to treat the event as a legal case; in so doing, it found its ability to hold people accountable severely limited. Taking the legal liability route, the Army charged Lieutenant William Calley with the premeditated murder of at least four civilians. Several others in command positions at the scene of the massacre (which involved at least 175 to 200 deaths) were initially

charged, but it was later decided to drop those charges in all but one additional case due to “insufficient evidence”—a standard reflecting the legal nature of the accountability approach being applied. The exception was Captain Ernest Medina, Calley’s immediate superior officer, who was also brought to trial and charged with murder. However, there was a difference. The murder charges against him were not for specific acts of premeditated killing (although evidence indicates that he killed at least four civilians at My Lai); rather, he was charged with 102 murders under a provision of military law that allowed prosecution of those with command responsibility. He was court-martialed and acquitted of those charges. Thus, although the military could have (and did in many cases) hold many more persons blameworthy under its Type B accountability approach (and subject to organizational sanctions), it held only Calley legally liable for four of approximately 200 deaths and sentenced him to 20 years in prison. By focusing on liability, in other words, and not pursuing specific counts of premeditated murder against Medina and others, the military was unable to fulfill the need for accountability-as-liability.

This leads to the second issue associated with Type C accountability—the problem of collective responsibility.^[113] Just as the legal system has difficulty dealing with collective responsibility,^[114] liability-based accountability systems are constantly challenged by the “many hands” problem.^[115] As Dennis Thompson noted, typical solutions to this problem include applying “hierarchical” or “collective” forms of responsibility, and he finds both wanting. In their place, he advocates applying “personal responsibility” approach based on some weak causal connection between an individual and the event. In terms of the present framework, that would mean shifting from Type C to either Type A or Type B accountability—a move that raises issues regarding the integrity and fairness of any sanctions that might be imposed. Such an approach does not deal with the complex situations where even a weak causal link cannot be proven. In such cases, even the legal system has had to modify its standards by allowing liability to rest almost entirely on an individual possible “fault” rather than cause.^[116]

As with its legal equivalent, the inherent problems of accountability-as-liability are many.^[117–120] Nevertheless, it is an approach to accountability that is extremely important in public administration. A central factor shaping Type C accountability for public administrators has been the judicial treatment of two issues—administrative discretion and administrative immunity. Historical shifts in the judiciary’s handling of administrative discretion and delegation of authority cases^[121,122] have altered the jurisdictions and tasks of administrative agencies to such a degree (some

have argued) that we have witnessed a de facto constitutional transformation of the American republic.^[123] The courts’ willingness to lift the cover of sovereign immunity over the past several decades^[124,125] has been as significant in its impact on accountability, exposing administrators to legal actions and making it urgent for them to become more “constitutionally competent.”^[126]

As significant has been the general impact on public administration of the legalistic mindset fostered by these developments. Almost every area of public management has shown signs of deference to the “lawyer’s” perspective, and an abandonment of the administrator’s view of the world.^[127] The very act of entering the public service places one in a distinctive legal relationship to the state,^[128] and efforts to assure public employee integrity through Type C approaches have been pervasive and often counterproductive for the day-to-day operations of government.^[129]

Law-Like Ethics

The ethics related to Type C accountability have their roots in distinct and conflicting concepts of law. Two approaches emerge from debates among philosophers regarding the nature of law, while a third is derived from the analogical reasoning used in the practice of law.

In the philosophy of law, a major division exists between those who assume laws to be human artifacts without any inherent moral value (the *positivist* school) and those who assess laws in terms of their relation to a higher law standard (the *natural law* school). Each school generates numerous ethical positions, but for our purposes we offer them in caricature as the positivist *ethics of obligation* and the naturalist *ethics of conscience*. As applied by those who are held accountable, the ethics of obligation calls for adherence to the explicit rules that define a situation. Here, we find the ethics of neutral competence that has been so central to the norms of American public administration.^[130–133] The ethics of conscience, in contrast, is manifest in efforts to have public administrators adhere to some “higher standard” when engaged in the enforcement or implementation of the law. Here, again, we find various standards from “regime values” and “public interest” to “social equity” and “justice-as-fairness.”^[100,134–136]

A third ethical approach relevant to Type C accountability comes from the practice of law. Sunstein^[137] offered the term “analogical reasoning” to describe the type of thinking that takes place in legal practice. It is reasoning that focuses on applying laws and rules to specific cases based on how that law or rule has been applied in the past to similar (although

not identical) cases. Sunstein highlighted four features of analogical reasoning. First, it is based on the norm of “principled consistency”—that is, the need to provide some principle to “harmonize seemingly disparate outcomes.” Second, there should be a focus on the details of the case, letting the facts shape the decision. Third, analogical reasoning avoids the application of any “deep or comprehensive theory” that might overwhelm the facts of the case. Finally, this form of reasoning produces and applies standards that should be at a “low or intermediate level of abstraction.” In public administration, the ethics of “constitutional competence” advocated by Rosenbloom and Carroll^[125,126] operates at this analogical level by focusing on the need for public service employees to become familiar with their place and responsibilities within the U.S. constitutional system through relevant cases.

The common thread holding these Type C ethical strategies together is their commitment to determining what is ethical through a reasoning process—one guided by rules, the second by a higher standard, and the third through analogy. In contrast to the stress on consequences (generated by Type A account giving) and blame avoidance (through Type B), the ethical approaches of Type C accountability rely on ethical behavior mechanisms that call for an actor to engage in a certain type of reasoning that fosters consideration of procedural requirements, legal standards, or precedent.^[138]

TYPE D: ACCOUNTABILITY-AS-ATTRIBUTABILITY

Relevance of Non-Work Expectations

Accountability-as-attributability brings into consideration those arenas of social life where the rules and roles of public administration are *seemingly* irrelevant, but are potentially quite important in the environment of accountability of public sector workers. It involves those contexts where the fact that an individual holds a certain position in an organization, or that one is subject to rules and constraints in the workplace, is of little or no consequence *most of the time*. It is the context of the “non-work domain” where one lives a “private life” distinct from the public life of the government job, or so it seems. In reality, accountability pressures spillover into the private lives of all employees, both public and private.

In the non-work domain, one is subject to accountability standards that are broader in scope, more diverse, and constantly subject to change over time and from place to place. The sources of those standards are other social actors in the domain who regard

an action or behavior as requiring an accounting by some actor. In short, they *attribute* an action or behavior to an individual and would *expect* an accounting if they were in a position to demand one. This attributing behavior is regarded by social psychologists as part of our standards means for making sense of, and dealing with, the world.^[139–141] From the perspective of “attribution theory” research, this form of attribution is subject to ideological, cultural, and stereotypical bias,^[142,143] as well as situational determinism.^[144] In addition, attributions tend to be unstable and malleable by changing conditions such as information or expectations.^[72,145,146] The shortcomings of such attributions, however, have not diminished their relevance.

Being of “Suspect” Character

The link between specific instances of attribution and our Type D accountability emerges when those attributions lead to the perception of individuals as being of suspect character given the actions attributed to them. Thus, an individual seen emerging in a disheveled state daily from a local bar might be regarded as a “drunk,” and someone cited by police for speeding might be characterized as a reckless driver. An attributed behavior becomes a character attribute, and it is the fact that we are subject to such characterizations that can make us Type D accountable. Individuals whose actions in private life are found to be questionable or potentially embarrassing to their employer are likely to find themselves being held accountable for those actions, despite their irrelevance to the employee’s tasks or functions.

This is especially true for those who hold public office. In the private sector, the relevance of Type D accountability pressures is likely to depend on numerous factors, including (but not limited to) one’s position and status in the organization and the nature of the non-work behavior that triggers the concern. Is the individual a highly visible employee whose offense affects the public image of the business? Was the behavior so offensive that the presence of the individual will result in significant losses in sales or productivity unless action is taken?

Exposure to this form of accountability is unavoidable, and yet students of management have paid little attention to it.^[147,148] Scholars have long acknowledged the intermingling of social and personal factors. They were central to the work of the human relations school and played an important role in Barnard’s classic analysis of executive management.^[66,149] Interest in the relationship of family life to work has produced a stream of research in sociology and organization studies,^[150,151] including the examination of strategies developed to deal with the boundaries between work

and non-work.^[152,153] There is also research focused on the non-work lives of public sector workers, in particular.^[154–156] But, with few exceptions,^[157] little of this research considers the link between personal life and accountability.

Uniqueness of Public Office Holding

The rules that protect one's private life from public intrusions do not necessarily apply to those in public offices. Although modern society might proffer the "principle of uniform privacy" as a desirable standard for its citizens, for public office holders it is the "principle of diminished privacy" that is more likely to be applied. There are various reasons for this, including the perceived necessity to know more about the behavior of those who wield power over us, as well as expectations that public officials must achieve a higher standard of conduct in their private affairs than ordinary citizens.^[158]

Of course, accountability-as-attributability is not limited to those who hold public office. What kind of accountability is involved? Some takes the form of formal constraints on the private, non-work lives of government employees. Hatch Acts, conflict-of-interest legislation, and related policies are designed explicitly to deal with the perceived dangers of having civil servants too actively engaged in politics or private business enterprises.^[128,159–161] Other examples are found in managerial strategies to offset the potentially adverse impact of private preferences. The possibility that employees or their families might grow too attached to local communities ("going native") has led some agencies to establish personnel transfer policies to minimize the influence of non-work factors.^[162] Concern for an agency's reputation or public image is also used as a rationale for holding employees accountable for their private behavior and lifestyles. Private behavior has been regarded as grounds for adverse personnel actions, even if the behavior had little or nothing to do with the position of workplace.^[163] As significant, the courts have held that the "perception" of questionable private behavior can be used to take action against a public employee.^[164]

Type D Ethical Strategies

The ethical responses to Type D accountability pressures are likely to require that the individual accept the linkages between one's personal and professional/work lives. Two assumptions are at work here. First is the assumption that individuals accepting a public office do so with the expectation that they will be subject to diminished privacy and greater scrutiny of their behavior. Second is the fact that, in today's complex

social world, efforts to carve out a private space free from the intrusions of work or public life is all but impossible. At least some aspects of the behavior attributed to individuals in their personal lives are relevant to how they will be viewed in their working lives, and vice versa. This fact of modern social life means that the ethical demands of work life often overlap or intrude into the ethics of personal life, a phenomenon that has been at the center of many of the most popular and influential studies of American social dynamics since the 1950s.^[165–168] The result is the blurring of the lines between one's personal and work life ethics.

Thus, the individual must constantly contend with an existential conflict between ethical demands generated by one's role expectations, perceived blameworthiness or liability, and those related to one's standing as an autonomous private individual or citizen. In some cases, the solution would be an ethical position that subordinates at least some of one's personal autonomy to the demands emerging from Types A, B, or C accountability. Accepting the limits on one's citizenship rights is the price paid under certain Hatch Act provisions, just as forest rangers accept the order to move their families to the next assignment station. As significant is the choice to subordinate one's private tastes and behavior to the demands imposed by one's job or status.^[169,170]

A more positive stance has been articulated by Terry Cooper in his calls for a "citizenship ethics," which he later presents as the "responsible administrator" model.^[171,172] Subordination of the personal is replaced with a commitment to continuously reflect on one's life as a public administrator. Implied is the idea that one can achieve self-actualization through such an ethical stance, an idea that traces back to ancient ethics and its emphasis on ethics as a means for achieving the good (e.g., happiness) in oneself (in contrast to the modern focus on achieving the good for others).^[173]

At a more general level, the response to Type D accountability might be the adoption of a virtue or character ethics approach. In an environment where standards and rules are ambiguous or unstable, it would be reasonable to assume that a consistent adherence to some higher standards of personal conduct (e.g., to be trustworthy, honest, benevolent, caring) would provide some protection from the whims of accountability-as-attribution. Some have articulated this in terms of being a person of "good character."^[174] It can also involve the a commitment to act virtuously on and off the job as the situation requires,^[175–179] or at least to avoid falling prey to the "ordinary vices" that plagued us in both our private and public existence.^[180]

The ethical path selected in response to Type D accountability pressures has implications for the

Table 2 Accountability–ethics relationships

Accountability type		Ethical strategies based on	
A: Answerability	Intuitive politician	Rational agent	Rule follower
B: Blameworthiness	Instrumental	Higher moral standard	Identity shift
C: Liability	Positivist obligation	Naturalist conscience	Analogical reasoning competence
D: Attributability	Subordination	Citizenship ethics	Virtue/character ethics

general orientation of the individual toward the other forms of accountability. Deference to the intrusions of accountability-as-attributability is most closely associated with the norm of political neutrality that characterizes mainstream American public administration. A “citizenship ethics” is likely to complement the effort to promote competency-centered ethics, while implied in the virtue/character ethics approaches is a view of public administrators as more actively engaged autonomous ethical agents who understand their moral obligations.^[181]

CONCLUSION

Among the various issues raised by the Friedrich–Finer debate, those focused on accountability–ethics relationships are at once the most significant and least explored. Despite the lack of explicit attention to that relationship in the public administration literature, there are obviously considerable resources that can be brought to bear on this topic. Through the reconceptualization of the key terms in that relationship, this essay opens the door to a more thorough examination of the relevant issues.

Such an examination would begin with a further elaboration and empirical validation of the associations summarized in Table 2. If those posited relationships are validated, the implications for our understanding of administrative behavior can be significant. By focusing on accountability and ethical norms and standards as social mechanisms, we can start to reexplore the “black box” of “contingent” administrative behavior that has been so central to public administration scholarship since Herbert Simon began to model the administrative decision maker in the 1940s. Relying on social mechanisms in lieu of models and positive theory assumptions about bureaucratic behavior, we are able to consider various forms of reasoning and “reasons” that help to shape administrative choices and behavior.^[37,182]

There are also potential implications for those concerned with institutional design and administrative reform. Reformers who advocate greater accountability, for instance, typically put forward suggestions

that create or enhance answerability mechanisms. A more fully elaborated “theory” of the relationship between accountability and ethical strategies would help us to assess the potential impact of alternative forms of answerability and to explain the success or failure of various reforms. The same logic could be useful in highlighting the source of current problems. What might otherwise be perceived as an arbitrary abuse of discretionary authority, for example, might be due to the absence of (or an indifference to) Type C accountability mechanisms.

REFERENCES

1. Dubnick, M.J. *Demons, Spirits, and Elephants: Reflections on the Failure of Public Administration Theory*; American Political Science Association: Atlanta, GA, September 1–4.
2. Friedrich, C.J. Public policy and the nature of administrative responsibility. In *Public Policy: A Yearbook of the Graduate School of Public Administration*; Friedrich, C.J., Mason, E.S., Eds.; Harvard University Press: Cambridge, MA, 1940; 3–24.
3. Finer, H. Administrative responsibility in democratic government. *Public Adm. Rev.* **1941**, *1* (4), 335–350, Summer.
4. Burke, J.P. *Bureaucratic Responsibility*; Johns Hopkins University Press: Baltimore, MD, 1986.
5. Harmon, M.M. *Responsibility as Paradox: A Critique of Rational Discourse on Government*; Sage: Thousand Oaks, CA, 1995.
6. Gruber, J.E. *Controlling Bureaucracies: Dilemmas in Democratic Governance*; University of California Press: Berkeley, CA, 1987.
7. McSwite, O.C. *Legitimacy in Public Administration: A Discourse Analysis*; Sage Publications: Thousand Oaks, CA, 1997.
8. Appleby, P.H. *Morality and Administration in Democratic Government*; Louisiana University Press: Baton Rouge, LA, 1952.
9. Gilbert, C.E. The framework of administrative responsibility. *J. Pol.* **1959**, *21*, 373–407, August 3.
10. Marx, F.M. Administrative ethics and the rule of law. *Am. Polit. Sci. Rev.* **1949**, *43* (6), 1119–1144.
11. Aberbach, J.D. *Keeping a Watchful Eye: The Politics of Congressional Oversight*; The Brookings Institution: Washington, D.C., 1990.

12. Balla, S.J. Administrative procedures and political control of the bureaucracy. *Am. Polit. Sci. Rev.* **1998**, 92 (3), 363–373, September.
13. Calvert, R.L.; McCubbins, M.D.; Weingast, B.R. A theory of political control and agency discretion. *Am. J. Polit. Sci.* **1989**, 33 (3), 588–611.
14. Harris, J.P. *Congressional Control of Administration*; Brookings Institution: Washington, D.C., 1964.
15. Hood, C.; James, O.; Scott, C. Regulation of government: has it increased, is it increasing, should it be diminished? *Public Adm.* **2000**, 78 (2), 283–304.
16. Kettl, D.F. Micromanagement: congressional control and bureaucratic risk. In *Agenda for Excellence: Public Service in America*; Ingraham, P.W., Kettl, D.F., Eds.; Chatham House: Chatham, NJ, 1992; Chapter 5.
17. Light, P.C. *Monitoring Government: Inspectors General and the Search for Accountability*; Brookings Institution: Washington, D.C., 1993.
18. McCubbins, M.D.; Schwartz, T. Congressional oversight overlooked: police patrols versus fire alarms. *Am. J. Polit. Sci.* **1984**, 28 (1), 165–179, February.
19. Rosen, B. *Holding Government Bureaucracies Accountable*, 2nd Ed.; Praeger Publishers: New York, 1989.
20. Wood, B.D.; Waterman, R.W. *Bureaucratic Dynamics: The Role of Bureaucracy in a Democracy*; Westview Press: Boulder, CO, 1994.
21. Behn, R.D. *Rethinking Democratic Accountability*; Brookings Institution Press: Washington, D.C., 2001.
22. Caiden, G.E. The problem of ensuring the public accountability of public officials. In *Public Service Accountability: A Comparative Perspective*; Jabbara, J.G., Dwivedi, O.P., Eds.; Kumarian Press: West Hartford, CT, 1988; 17–38.
23. Deleon, L. Accountability in a 'reinvented' government. *Public Adm.* **1998**, 76, 539–558, Autumn.
24. Dubnick, M.J.; Romzek, B.S. *American Public Administration: Politics and the Management of Expectations*; Macmillan: New York, 1991.
25. Dubnick, M.J.; Romzek, B.S. Accountability and the centrality of expectations in american public administration. In *Research in Public Administration*; Perry, J.L., Ed.; JAI PRESS: Greenwich, CT, 1993; 37–78, Chapter 2.
26. Kearns, K.P. *Managing for Accountability: Preserving the Public Trust in Public and Non-profit Organizations*; Jossey-Bass: San Francisco, CA, 1996.
27. Romzek, B.S. Accountability of congressional staff. *J. Public Adm. Res. Theory* **2000**, 10 (2), 413–446.
28. Romzek, B.S.; Dubnick, M.J. Accountability in the public sector: lessons from the challenger tragedy. *Public Administration Review* **1987**, 47 (3), 227–238.
29. Romzek, B.S.; Dubnick, M.J. Issues of accountability in flexible personnel systems. In *New Paradigms for Government: Issues for the Changing Public Service*; Ingraham, P.W., Romzek, B.S., Eds.; Jossey-Bass: San Francisco, CA, 1994; 263–294.
30. Romzek, B.S.; Dubnick, M.J. Accountability. In *Defining Public Administration: Selections from the International Encyclopedia of Public Policy and Administration*; Shafritz, J.M., Ed.; Westview Press: Boulder, CO, 2000; 382–395.
31. Romzek, B.S.; Ingraham, P.W. Cross pressures of accountability: initiative, command, and failure in the ron brown plane crash. *Public Adm. Rev.* **2000**, 60 (3), 240–253.
32. Kaplan, A. *The Conduct of Inquiry: Methodology for Behavioral Science*; Chandler Publishing Co.: Scranton, PA, 1964.
33. Mulgan, R. 'Accountability': an ever-expanding concept? *Public Adm.* **2000**, 78 (3), 555–573.
34. Sinclair, A. The chameleon of accountability: forms and discourses. *Account. Organ. Soc.* **1995**, 20 (2/3), 219–237.
35. Beu, D.; Buckley, M.R. The hypothesized relationship between accountability and ethical behavior. *J. Bus. Ethics* **2001**, 34 (2), 57–73.
36. Garofalo, C.; Geuras, D. *Ethics in the Public Service: The Moral Mind at Work*; Georgetown University Press: Washington, D.C., 1999.
37. Elster, J. *Nuts and Bolts for the Social Sciences*; Cambridge University Press: New York, 1989.
38. Elster, J. A plea for mechanisms. In *Social Mechanisms: An Analytical Approach to Social Theory*; Hedström, P., Swedberg, R., Eds.; Cambridge University Press: Cambridge, 1998; 45–73.
39. Hedström, P.; Swedberg, R. Social mechanisms: an introductory essay. In *Social Mechanisms: An Analytical Approach to Social Theory*; Hedström, P., Swedberg, R., Eds.; Cambridge University Press: Cambridge, 1998; 1–31.
40. Lane, R. *Political Science in Theory and Practice: The 'Politics' Model*; M.E. Sharpe: Armonk, NY, 1997.
41. Merton, R.K. *Social Theory and Social Structure*; Free Press: New York, 1968; 1968 enlarged.
42. Harman, G. *The Nature of Morality: An Introduction to Ethics*; Oxford University Press: New York, 1977.
43. LaFollette, H. Pragmatic ethics. In *The Blackwell Guide to Ethical Theory*; LaFollette, H., Ed.; Blackwell: Oxford, UK, 2000; 400–419.
44. Flanagan, O. Ethics naturalized: ethics as human ecology. In *Mind and Morals: Essays on Cognitive Science and Ethics*; May, L., Friedman, M., Clark, A., Eds.; MIT Press: Cambridge, MA, 1996; 19–43, Chapter 2.
45. Putnam, R.A. The moral impulse. In *The Revival of Pragmatism: New Essays on Social Thought, Law, and Culture*; Dickstein, M., Ed.; Duke University Press: Durham, NC, 1998; 62–71.
46. Dubnick, M. Clarifying accountability: an ethical theory framework. In *Public Sector Ethics: Finding and Implementing Values*; Charles Sampford, N.P., Bois, C.-A., Eds.; The Federation Press/Routledge: Leichhardt, NSW, Australia, 1998; 68–81, Chapter 5.
47. Roberts, J.T. *Accountability in Athenian Government*; University of Wisconsin Press: Madison, WI, 1982.
48. Elster, J. Accountability in athenian politics. In *Democracy, Accountability, and Representation*; Przeworski, A., Stokes, S.C., Manin, B., Eds.; Cambridge University Press: Cambridge, UK, 1999; 253–278.

49. Schedler, A., Diamond, L.J., Plattner, M.F., Eds.; *The Self-Restraining State: Power and Accountability in New Democracies*; Lynne Rienner Publishers: Boulder, CO, 1999.
50. Scott, M.B.; Lyman, S.M. Accounts. *Am. Sociol. Rev.* **1968**, *33* (1), 46–62, February.
51. Tavuchis, N. *Mea Culpa: A Sociology of Apology and Reconciliation*; Stanford University Press: Stanford, CA, 1991.
52. Minow, M. *Between Vengeance and Forgiveness: Facing History after Genocide and Mass Violence*; Beacon Press: Boston, 1998.
53. Borneman, J. *Settling Accounts: Violence, Justice, and Accountability in Postsocialist Europe*; Princeton University Press: Princeton, NJ, 1997.
54. Benoit, W.L. *Accounts, Excuses, and Apologies: A Theory of Image Restoration Strategies*; State University of New York Press: Albany, 1995.
55. Sykes, G.M.; Matza, D. Techniques of neutralization: a theory of delinquency. *Am. Sociol. Rev.* **1957**, *22* (6), 664–670.
56. Cava, A.; West, J.; Berman, E. Ethical decision-making in business and government: an analysis of formal and informal strategies. *Spectrum* **1995**, *68* (2), 28–36.
57. Arendt, H. *Eichmann in Jerusalem: A Report on the Banality of Evil*; Penguin Books: New York, 1976, revised.
58. Sereny, G. *Albert Speer: His Battle with Truth*; Knopf: New York, 1995.
59. Turner, R.H. The role and the person. *Am. J. Sociol.* **1978**, *84* (1), 1–23 July.
60. Biddle, B.J. Recent developments in role theory. *Annu. Rev. Sociology* **1986**, *12*, 67–92.
61. Schedler, A. Conceptualizing accountability. In *The Self-Restraining State: Power and Accountability in New Democracies*; Schedler, A., Diamond, L.J., Plattner, M.F., Eds.; Lynne Rienner Publishers: Boulder, CO, 1999; 13–28.
62. Held, D. *Models of Democracy*, 2nd Ed.; Stanford University Press: Stanford, CA, 1996.
63. Pitkin, H.F., Ed.; *Representation*; Atherton Press: New York, 1969.
64. Fearon, J.D. Electoral accountability and the control of politicians: selecting good types versus sanctioning poor performance. In *Democracy, Accountability, and Representation*; Przeworski, A., Stokes, S.C., Manin, B., Eds.; Cambridge University Press: Cambridge, UK, 1999; 55–97.
65. Laver, M.; Shepsle, K.A. Government accountability in parliamentary democracy. In *Democracy, Accountability, and Representation*; Przeworski, A., Stokes, S.C., Manin, B., Eds.; Cambridge University Press: Cambridge, UK, 1999; 279–296.
66. Barnard, C.I. *The Functions of the Executive*; Harvard University Press: Cambridge, MA, 1968.
67. March, J.G.; Simon, H.A. *Organizations*; John Wiley and Sons: New York, 1958.
68. McMahon, C. *Authority and Democracy: A General Theory of Government and Management*; Princeton University Press: Princeton, NJ, 1994.
69. McGregor, D. *The Human Side of Enterprise*; McGraw-Hill: New York, 1960.
70. Tetlock, P.E. Accountability and the perseverance of first impressions. *Soc. Psychol. Q.* **1983**, *46* (4), 285–292, December.
71. Tetlock, P.E. Accountability and complexity of thought. *J. Pers. Soc. Psychol.* **1983**, *45* (1), 74–83.
72. Tetlock, P.E. Accountability: a social check on the fundamental attribution error. *Soc. Psychol. Q.* **1985**, *48* (3), 227–236, September.
73. Tetlock, P.E.; Kim, J.I. Accountability and judgment processes in a personality prediction task. *J. Pers. Soc. Psychol.* **1987**, *52* (4), 700–709.
74. Tetlock, P.E.; Skitka, L.; Boettger, R. Social and cognitive strategies for coping with accountability: conformity, complexity, and bolstering. *J. Pers. Soc. Psychol.* **1989**, *57* (4), 632–640.
75. Miller, G.J. *Managerial Dilemmas: The Political Economy of Hierarchy*; Cambridge University Press: Cambridge, UK, 1992.
76. Eisenhardt, K.M. Agency theory: an assessment and review. *Acad. Manage. Rev.* **1989**, *14* (1), 57.
77. Waterman, R.W.; Meier, K.J. Principal-agent models: an expansion? *J. Public Adm. Res. Theory* **1998**, *8* (2), 173–202.
78. Moe, T.M. The politics of structural choice: toward a theory of public bureaucracy. In *Organization Theory: From Chester Barnard to the Present and Beyond Expanded*; Williamson, O.E., Ed.; Oxford University Press: New York, 1995; 116–153.
79. March, J.G.; Olsen, J.P. The new institutionalism: organizational factors in political life. *Am. Polit. Sci. Rev.* **1984**, *78* (3), 734–749, September.
80. March, J.G.; Olsen, J.P. *Democratic Governance*; Free Press: New York, 1995.
81. March, J.G.; Olsen, J.P. *Rediscovering Institutions: The Organizational Basis of Politics*; Free Press: New York, 1989.
82. March, J.G. *A Primer on Decision Making: How Decisions Happen*; Free Press: New York, 1994.
83. Paul, J.; Strbiak, C.A. The ethics of strategic ambiguity. *J. Bus. Commun.* **1997**, *34* (2), 149–159.
84. Bøhren, Øyvind. The agent's ethics in the principal-agent model. *J. Bus. Ethics* **1998**, *17*, 745–755.
85. Brehm, J.; Gates, S. *Working, Shirking, and Sabotage: Bureaucratic Response to a Democratic Public*; University of Michigan Press: Ann Arbor, 1997.
86. Scheffler, S. *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Considerations*; Clarendon Press: Oxford, UK, 1994, revised.
87. Stryker, S.; Macke, A.S. Status inconsistency and role conflict. *Annu. Rev. Sociology* **1978**, *4*, 57–90.
88. Dubnick, M.J. Public service ethics and the cultures of blame, Fifth International Conference of Ethics in the Public Service, Brisbane, Australia, Aug 5–9, 1996.
89. Douglas, T. *Scapegoats: Transferring Blame*; Routledge: London, UK, 1995.
90. Girard, R. *The Scapegoat*; Johns Hopkins University Press: Baltimore, 1986.

91. Gay, P. *The Cultivation of Hatred: The Bourgeois Experience, Victoria to Freud*; W. W. Norton and Co.: New York, 1993.
92. Cohen, E.A.; Gooch, J. *Military Misfortunes: The Anatomy of Failure in War*; Vintage Books: New York, 1991.
93. Cohen, W.S. *Personal Accountability for Force Protection at Khobar Towers*, Department of Defense, Office of the Secretary of Defense, Washington, D.C., July 31; <http://www.defenselink.mil/pubs/khobar/index.html>.
94. Shklar, J.N. *The Faces of Injustice*; Yale University Press: New Haven, CT, 1990.
95. Smiley, M. *Moral Responsibility and the Boundaries of Community: Power and Accountability From a Pragmatic Point of View*; University of Chicago Press: Chicago, 1992.
96. Woodward, B. *The Commanders*; Simon and Schuster: New York, 1991.
97. Chandler, R.C. Deontological dimensions of administrative ethics. In *Handbook of Administrative Ethics*; Cooper, T.L., Ed.; Marcel Dekker: New York, 1994; 147-156.
98. Denhardt, R.B.; Denhardt, J.V. The new public service: serving rather than steering. *Public Adm. Rev.* **2000**, *60* (6), 549-559.
99. Huddleston, M.W. The carter civil service reforms: some implications for political theory and public administration. *Polit. Sci. Q.* **1981-1982**, *96* (4), 607-621.
100. Frederickson, H.G. *The Spirit of Public Administration*; Jossey-Bass Publishers: San Francisco, 1997.
101. Marini, F., Ed.; *Toward a New Public Administration: The Minnowbrook Perspective*; Chandler Publishing: Scranton, PA, 1971.
102. Wamsley, G.L.; Bacher, R.N.; Goodsell, C.T.; Kronenberg, P.S.; Rohr, J.A.; Stivers, C.M.; White, O.F.; Wolf, J.F. *Refounding Public Administration*; Sage Publications: Newbury Park, CA, 1990.
103. Wamsley, G.L., Wolf, J.F., Eds.; *Refounding Democratic Public Administration: Modern Paradoxes, Postmodern Challenges*; Sage Publications: Thousand Oaks, CA, 1996.
104. Shklar, J.N. *Legalism: Law, Morals, and Political Trials*; Harvard University Press: Cambridge, MA, 1986.
105. Pound, R. *Social Control Through Law*; Transaction Publishers: New Brunswick, 1997.
106. Schuck, P.H. *The Limits of Law: Essays on Democratic Governance*; Westview Press: Boulder, CO, 1999.
107. Ingram, P. Maintaining the rule of law. *Philos. Q.* **1985**, *35* (141), 359-381.
108. von Hayek, F.A. *The Constitution of Liberty*; Regnery: Chicago, 1972.
109. Fuller, S.R.; Edelman, L.B.; Matusik, S.F. Legal readings: employee interpretation and mobilization of law. *Acad. Manage. Rev.* **2000**, *25* (1), 200-216.
110. Minow, M. *Making All the Difference: Inclusion, Exclusion, and American Law*; Cornell University Press: Ithaca, 1990.
111. Minow, M. *Not Only for Myself: Identity, Politics and the Law*; The New Press: New York, 1997.
112. Weber, M. *Economy and Society: An Outline of Interpretive Sociology, Volume 1*; University of California Press: Berkeley, CA, 1978.
113. French, P.A., Ed.; *Individual and Collective Responsibility* 2nd; Schenkman Books: Rochester, VT, 1997, rev.
114. Stone, C.D. *Where the Law Ends: The Social Control of Corporate Behavior*; Harper Torchbooks: New York, 1975.
115. Thompson, D.F. Moral responsibility of public officials: the problem of many hands. *Am. Polit. Sci. Rev.* **1980**, *74*, 905-916, December 4.
116. Thomson, J.J. Remarks on causation and liability. *Philos. Public Aff.* **1984**, *13* (2), 101-133.
117. Huber, P.W. *Liability: The Legal Revolution and Its Consequences*; Basic Books: New York, 1988.
118. Levmore, S.X., Ed.; *Foundations of Tort Law*; Oxford University Press: New York, 1994.
119. Lieberman, J.K. *The Litigious Society*; Basic Books: New York, 1981.
120. Howard, P.K. *The Death of Common Sense: How Law is Suffocating America*; Random House: New York, 1994.
121. Horwitz, R. Judicial review of regulatory decisions: the changing criteria. *Polit. Sci. Q.* **1994**, *109* (1), 133-169.
122. Stewart, R.B. The reformation of american administrative law. *Harvard Law Rev.* **1975**, *88*, 1669-1813.
123. Lowi, T.J. *The End of Liberalism: The Second Republic of the United States*, 2nd Ed.; W. W. Norton and Co.: New York, 1979.
124. Rosenbloom, D.H. *Public Administration and Law: Bench V. Bureau in the United States*; Marcel Dekker: New York, 1983.
125. Rosenbloom, D.H. Public administrators' liability: bench v. bureau in the contemporary administrative state. *Public Adm. Q.* **1987**, *10* (4), 373-386, Winter.
126. Rosenbloom, D.H.; Carroll, J.D.; Carroll, J.D. *Constitutional Competence for Public Managers: Cases and Commentary*; F. E. Peacock Publishers: Itasca, IL, 2000.
127. Dimock, M.E. *Law and Dynamic Administration*; Praeger Publishers: New York, 1980.
128. Rosenbloom, D.H. *Federal Service and the Constitution: The Development of the Public Employment Relationship*; Cornell University Press: Ithaca, NY, 1971.
129. Anechiarico, F.; Jacobs, J.B. *The Pursuit of Absolute Integrity: How Corruption Control Makes Government Ineffective*; University of Chicago Press: Chicago, 1996.
130. Kaufman, H. Emerging conflicts in the doctrines of public administration. *Am. Polit. Sci. Rev.* **1956**, *50* (4), 1057-1073, December.
131. Lockard, D. The city manager, administrative theory and political power. *Polit. Sci. Q.* **1962**, *77* (2), 224-236.
132. Finer, H. Better government personnel: america's next frontier. *Polit. Sci. Q.* **1936**, *LI* (4), 569-599.

133. Rourke, F.E. Responsiveness and neutral competence in american bureaucracy. *Public Adm. Rev.* **1992**, 52 (6), 539–546, November/December.
134. Marshall, G.S.; Choudhury, E. Public administration and the public interest: re-presenting a lost concept. *Am. Behav. Sci.* **1997**, 41 (1), 119–131, September.
135. Rohr, J.A. *Public Service, Ethics, and Constitutional Practice*; University Press of Kansas: Lawrence, KS, 1998.
136. Rohr, J.A. *Ethics for Bureaucrats: An Essay on Law and Values*, 2nd Ed.; Marcel Dekker: New York, 1989.
137. Sunstein, C.R. *Legal Reasoning and Political Conflict*; Oxford University Press: New York, 1996.
138. Raz, J. *Practical Reason and Norms*; Princeton University Press: Princeton, NJ, 1990.
139. Forsyth, D.R. The functions of attributions. *Soc. Psychol. Q.* **1980**, 43 (2), 184–189, June.
140. Crittenden, K.S. Sociological aspects of attribution. *Annu. Rev. Sociology* **1983**, 9, 425–446.
141. Howard, J.A. A social cognitive conception of social structure. *Soc. Psychol. Q.* **1994**, 57 (3), 210–227, September.
142. Howard, J.A.; Pike, K.C. Ideological investment in cognitive processing: the influence of social statuses on attribution. *Soc. Psychol. Q.* **1986**, 49 (2), 154–167.
143. Al-Zahrani, S.S.A.; Kaplowitz, S.A. Attributional biases in individualistic and collectivistic cultures: a comparison of americans with saudis. *Soc. Psychol. Q.* **1993**, 56 (3), 223–233.
144. Holloway, S.D.; Fuller, B. Situational determinants of causal attributions: the case of working mothers. *Soc. Psychol. Q.* **1983**, 46 (2), 131–140.
145. Yarkin-Levin, K. Anticipated interaction, attribution, and social interaction. *Soc. Psychol. Q.* **1983**, 46 (4), 302–311.
146. Tetlock, P.E. The influence of self-presentation goals on attributional reports. *Soc. Psychol. Q.* **1981**, 44 (4), 300–311, December.
147. Cozzetto, D.A.; Pedeliski, T.B. Privacy and the workplace. *Rev. Public Pers. Adm.* **1996**, 16 (2), 21–31.
148. Cozzetto, D.A.; Pedeliski, T.B. Privacy and the workplace: technology and public employment. *Public Pers. Manage.* **1997**, 26 (4), 515–527.
149. Roethlisberger, F.J. *Management and Morale*; Harvard University Press: Cambridge, MA, 1941.
150. Cohen, A. An examination of the relationships between work commitment and non-work domains. *Human Relat.* **1995**, 48 (3), 239–263.
151. Bielby, D.D. Commitment to work and family. *Annu. Rev. Sociology* **1992**, 18, 281–302.
152. Kanter, R.M. *Men and Women of the Corporation*; Basic Books: New York, 1977/1993.
153. Kirchmeyer, C. Managing the work–non-work boundary: an assessment of organizational responses. *Human Relat.* **1995**, 48 (5), 515–536.
154. Romzek, B.S. Personal consequences of employee commitment. *Acad. Manage. J.* **1989**, 32 (3), 649–661.
155. Johnson, C.M.; Duerst-Lahti, G. Private lives and public work: professional careers in state civil service. *Rev. Public Pers. Adm.* **1991**, 12 (1), 14–32.
156. Romzek, B.S. Work and non-work psychological involvements: the search for linkage. *Adm. Soc.* **1985**, 17 (2), 257–281.
157. Dworkin, T.M. It's my life—leave me alone: off-the-job employee associational privacy rights. *Am. Bus. Law J.* **1997**, 35 (1), 47–103.
158. Thompson, D.F. *Political Ethics and Public Office*; Harvard University Press: Cambridge, MA, 1987.
159. Thruber, K.T. Big, little, littler: synthesizing hatch act-based political activity legislation research. *Rev. Public Pers. Adm.* **1993**, 13 (1), 38–51.
160. Roberts, R.N. Conflict of interest regulation, employees' rights, and the constitution. *Public Adm. Q.* **1992**, 16 (3), 344–367 Fall.
161. Rosenbloom, D.H. The public employment relationship and the supreme court in the 1980s. *Rev. Public Pers. Adm.* **1988**, 8 (2), 49–65.
162. Kaufman, H. *The Forest Ranger: A Study in Administrative Behavior*; Johns Hopkins University Press: Baltimore, MD, 1967.
163. Lewis, G.B. Lifting the ban on gays in the civil service: federal policy toward gay and lesbian employees since the cold war. *Public Adm. Rev.* **1997**, 57 (5), 387–395.
164. Stark, A. The appearance of official impropriety and the concept of political crime. *Ethics* **1995**, 105 (2), 326–351.
165. Bellah, R.N.; Madsen, R.; Sullivan, W.M.; Swidler, A.; Tipton, S.M. *Habits of the Heart: Individualism and Commitment in American Life*; University of California Press: Berkeley, CA, 1985.
166. Bellah, R.N.; Madsen, R.; Sullivan, W.M.; Swidler, A.; Tipton, S.M. *The Good Society*; Alfred A. Knopf: New York, 1991.
167. Whyte, W.H. *The Organization Man*; Doubleday Anchor: Garden City, NY, 1956.
168. Coleman, J.S. *The Asymmetric Society*; Syracuse University Press: Syracuse, NY, 1982.
169. Hummel, R.P. *The Bureaucratic Experience: A Critique of Life in the Modern Organization*, 4th Ed.; St. Martin's Press: New York, 1994.
170. Denhardt, R.B. *In the Shadow of Organization*; The Regents Press of Kansas: Lawrence, KS, 1981.
171. Cooper, T.L. *An Ethic of Citizenship for Public Administration*; Prentice Hall: Englewood Cliffs, NJ, 1991.
172. Cooper, T.L. *The Responsible Administrator: An Approach to Ethics for the Administrative Role*, 4th Ed.; Jossey-Bass Publishers: San Francisco, 1998.
173. Annas, J. Ancient ethics and modern morality. *Philos. Perspect.* **1992**, 6, 119–136.
174. Luke, J.S. Character and conduct in the public service: a review of historical perspectives and a definition for the twenty-first century. In *Handbook of Administrative Ethics*; Cooper, T.L., Ed.; Marcel Dekker: New York, 1994; 391–412.

175. Hart, D.K. The virtuous citizen, the honorable bureaucrat, and 'public' administration. *Public Adm. Rev.* **1984**, 44 (Special Issue), 111–120, March.
176. Hart, D.K.; Smith, P.A. Fame, fame-worthiness, and the public service. *Adm. Soc.* **1988**, 20 (2), 131–151, August.
177. Frederickson, H.G.; Hart, D.K. The public service and the patriotism of benevolence. *Public Adm. Rev.* **1985**, 45 (5), 547–553.
178. Frederickson, H.G.; Hart, D.K. Patriotism, benevolence, and public administration. In *The Spirit of Public Administration*; Jossey-Bass Publishers: San Francisco, 1997; 195–208.
179. Hart, D.K. Administration and the ethics of virtue: in all things, choose first for good character and then for technical expertise. In *Handbook of Administrative Ethics*; Cooper, T.L., Ed.; Marcel Dekker: New York, 1994; 107–123.
180. Shklar, J.N. *Ordinary Vices*; Belknap/Harvard University Press: Cambridge, MA, 1984.
181. Bovens, M. *The Quest for Responsibility: Accountability and Citizenship in Complex Organisations*; Cambridge University Press: Cambridge, England, 1998.
182. Hedström, P., Swedberg, R., Eds.; *Social Mechanisms: An Analytical Approach to Social Theory*; Cambridge University Press: Cambridge, 1998.

Accounting and Reporting for Private Non-Profit Organizations—Balancing Economic Efficiency with Social Mission

John F. Sacco

*Department of Public and International Affairs, George Mason University,
Fairfax, Virginia, U.S.A.*

Walter Vance

General Accounting Office, Washington, District of Columbia, U.S.A.

INTRODUCTION

This chapter discusses how the value of economic efficiency, which is typically associated with private business accounting and operations, is increasingly being used to measure private non-profit accounting and operations financial performance. Since the mid-1990s, private non-profit organizations under the Financial Accounting Standards Board (FASB) have begun to follow the full cost accrual and consolidation model that is associated with measuring economic efficiency. This change suggests that the public choice philosophy (business efficiency in government and non-profits) has cast its shadow over the traditional progressive philosophy (social mission) that historically was the model of operations in private non-profit organizations. Specifically, the use of full costing for every project, taking a hard look at projects that are not “breaking even,” obtaining more outputs for less input, and putting extra emphasis on “earned income” in the form of donations are now all associated with the new non-profit environment. Whether the pendulum will swing back to having social mission instead of economic efficiency being the primary criteria by which private non-profits are judged is a critical question given the importance of the work that non-profits undertake. With issues like environmental degradation, poverty, and international unrest in the forefront of the news, more pressure is placed on private non-profit organizations to address these issues. How should the increasingly important private non-profit sector account, measure, and report success?

DEFINITION

Private non-profit organizations are not affiliated with a government, even though they may receive grants or aid from different levels of government. Much of

their revenue comes from voluntary contributions or earnings for services provided—not taxes.

The range of functions provided by private non-profit entities is wide. The traditional private non-profit is a charity, such as Catholic Relief Services or the District of Columbia Capital Area Food Bank. They provide services to the needy. Trade associations are private non-profit organizations but they serve their members as opposed to the public at large. The American Bankers Association is a trade organization. Like trade associations, business leagues such as local chambers of commerce are private non-profit entities. Even political action groups that lobby for legislation can fall in the private non-profit category. Private non-profits are corporations and as such must obtain corporation status from a state government. To obtain tax-exempt status they must seek approval from the Internal Revenue Service (IRS). Many private non-profits must report to the state in which they were incorporated and to the IRS. IRS form 990 is the usual way in which private non-profits provide information to the IRS on a yearly basis.

THE ROAD TO FASB AND THE EMPHASIS ON COMPETITION

A private non-profit does not have to follow FASB accounting standards. However, they cannot obtain a clean (unqualified) audit without fairly expressing their financial statements in accordance with generally accepted accounting principles (GAAP) as set by FASB. Even obtaining a bank loan may require financial reporting in accord with FASB standards. Although FASB is a private operation, it has permission from the Securities and Exchange Commission (SEC) to write accounting rules for business and provide non-profit entities. When private non-profit organizations use FASB accounting standards, they are

subject to all FASB requirements. However, some FASB rules were written specifically for private non-profit organizations. Four of the FASB standards that are specific to non-profit entities will be discussed in detail.

The financial reporting model imposed by FASB is the accrual and consolidation model with some selective use of fair value accounting (discussed later). For the most part this accrual and consolidation model emphasizes economic efficiency. All costs, regardless of whether cash has changed hands must be matched against revenue to determine whether the cost (effort) generated adequate income to break even or show a surplus. The Statement of Activities for a private non-profits (similar to the income statement of a business) differentiates among program expenses, administration (often called management and general), and fund-raising. If private non-profit organizations are formed to serve a social mission in the community, then a high percentage of expenses going to fund-raising and administration might contradict the social mission orientation. With the Statement of Financial Position (the balance sheet in business), money restricted for certain purposes can be distinguished from money that is unrestricted. While all the money must benefit the social mission as defined by the charter, some monies may be restricted to address special aspects of the social mission.

Use of the Economic Efficiency Criteria and the Accrual and Consolidation Model

The inclusion of private non-profit organizations into the FASB fold has a history behind it. FASB became concerned about the many sources of accounting rules for private non-profit organizations. Specifically, FASB felt that users of financial reports were getting inconsistent information. In some cases, FASB and others felt the accounting and reporting rules for these private non-profit organizations were too flexible and allowed the organizations to provide information that was not sufficiently candid.

Now that FASB is making the accounting and reporting standards for private non-profit organizations, this means private non-profit organizations must recognize revenue when earned and account for all costs necessary to earn those revenues in that period. If a pledge is made late in the fiscal year with the cash anticipated during the next period, then it is revenue in the year when pledged, not when the cash is received. The work, the phone calls, the web site, and the direct mailing have been done to earn the contribution during the period when the pledge was made. Thus, the pledge is considered earned revenue. On the expense side, if employees have pension benefits or other accrued

compensated absences (e.g., sick leave and vacation) the cost of those (usually some present value of the future payment) must be included during the period when they were promised even though the money will not be paid until a later date.

Consolidation is also part of the FASB approach. Separate funds to distinguish current donations from endowments are no longer used for external reporting. The total of all revenues (in consolidated format, not fund format) is reported. If the private non-profit has a sizable endowment, that endowment is part of the consolidated assets and revenues. Thus, the private non-profit may look very wealthy even in years when cash donations are low and the non-profit is facing liquidity problems. In the past, those sizable endowments could be placed in a separate fund and not counted as part of the total wealth (now called net assets) of the non-profit.

SFAS 116—ACCOUNTING FOR CONTRIBUTIONS RECEIVED AND CONTRIBUTIONS MADE

In examining Statement of Financial Accounting Standards (SFAS) 116, issued under due process by the Financial Accounting Standards Board, it is important to understand that the standard applies to private non-profit organizations that receive contributions as well as organizations or individuals that make the contribution. If a company makes a contribution to a private non-profit organizations, the timing of when (i.e., the basis) to incur the expense for the company and when to recognize the revenue for the non-profit both come from SFAS 116.

The main goal of SFAS 116 is to make sure that contributions are appropriately and consistently recorded, using the accrual logic for revenue recognition.

With respect to contributions, SFAS 116 has several broad categories that help determine when revenue should be recognized. Fig. 1 provides the categories.

Example of Revenue

Assume a private non-profit organization has the following transactions and events.

What amount can be considered revenue and in which of these categories?

1. One party contributes a \$1000 check with no conditions attached.
2. A company promises to contribute \$10,000 contingent on the private non-profit collecting \$5000 specifically to match the promise made by the company.

Categories	Revenue?
Unconditional contributions	Recognized in period received at fair market value
Conditional promises	Recognized when conditions are substantially met because future uncertainty is attached to the conditions
Services	Only if a nonfinancial asset is created and special skills ordinarily contracted are used

Fig. 1 When to recognize revenue.

- Another individual pledges \$2000 with the promise to make the donation in this current fiscal period.
- An electrician donates his time to the organization by installing wiring. The electrician's time is valued at \$500. If the electrician had not done this skilled work, the private non-profit would have had to purchase the service.
- One company, aware of the \$5000 matching requirement of the other company, promises to give \$1500 if the full \$5000 matching requirement is reached.
- A local government contracts with the private non-profit organization. The organization receives \$1200 for work to be performed in the next period.
- An individual donates \$1000 but restricts its use to a certain program.

Fig. 2 shows the results.

No conditions were applied to the \$4000 revenue consisting of two \$1000 cash contributions (Nos. 1 and 7) and one \$2000 (No. 3) pledge contribution. The \$11,500 of conditional promises (Nos. 2 and 5) would be placed in the notes to the financial statements if the conditions were not met this accounting and reporting period. For service (No. 4) notice how the revenue (\$500) from service is offset by an expense (\$500). The money received (No. 6) from the local government contract is not revenue; rather, it is a liability because the non-profit still owes the work to the local government. Having the cash is insufficient under the accrual logic to declare something a revenue. The revenue must be earned. In this case of the contract, the non-profit must do the work later and thus has a liability or future sacrifice often called unearned

revenue. In addition to liabilities, other distinctions are important in the revenue recognition logic.

Classifying Revenue into Net Assets— Permanently Restricted, Temporarily Restricted and Unrestricted

When revenue is earned by a non-profit it needs to be placed in one of three categories as designated by the donor.

- Permanently restricted—Oftentimes, the corpus or original amount cannot be spent. It is permanently restricted. Only the interest or gains earned might be unrestricted and available for current expenses.
- Temporarily restricted—This can only be spent on a certain program, e.g., health program, or cannot be spent until a later period, e.g., 3 months from now. Other temporary restrictions are possible.
- Unrestricted—Can be spent in any legal manner related to the mission.

SFAS 117—FINANCIAL STATEMENTS OF NON-PROFIT ORGANIZATIONS

The Three Required Financial Statements Under SFAS 117 are:

- Statement of financial position (balance sheet).
- Statement of activities (the term income is not used because non-profit organizations carry out activities to benefit the community not earn a profit).
- Statements of cash flow (shows the sources of cash, cash payments, and increase or decrease in cash balance, or more technically, cash and cash equivalents).

Categories	Revenue	Not revenue	Liability	Expense
Unconditional contributions	\$4,000			
Conditional promises		\$11,500		
Service as revenue	\$500			\$500
Work to be performed			\$1,200	
Total	\$4,500	\$11,500		\$500

Fig. 2 What are the amounts in the appropriate categories?

Assume that the non-profit has finished the fiscal year that went from 7/1/x0 to 6/30/x1.

Statement of Financial Position

In this first statement presented, the statement of financial position (or balance sheet) reflects the accounting equation (assets = liabilities + net assets) in that assets and liabilities are shown. Note that net assets are used in place of equity (the label to connote business ownership) and further that net assets are divided into permanently, temporary, and unrestricted categories. Net assets and the three categories are used to capture the not-for-profit nature and the types of restrictions that go with donations. Remember, also, the balance sheet is for a point in time and shows ability to pay short- and long-term obligations from the asset pool. It also shows the ability of the entity to take advantage of emerging opportunities by comparing things such as liquid assets with short-term obligations (Fig. 3).

The asset section does not have any nomenclature that is overly complex. It is arranged in terms of liquidity, from the most liquid, cash, to the least liquid, property, plant, and equipment, and long term investments. Contributions receivable parallel accounts

receivable typical of private sector business operations. Liabilities are ordered from those that need to be paid the soonest to those that need to be paid later.

In the analysis of this balance sheet, total assets, \$29,000, exceed liabilities, \$13,000, by \$16,000, suggesting a reasonably healthy financial status. For instance, the cash in unrestricted, \$2000, is sufficient to cover the accounts payable (which total \$500). A closer look reveals some possible problems. As is often the case with non-profit entities, donor restrictions can limit flexibility. The entity has a note payable that might come due soon. The note payable is \$5000, whereas the excess of unrestricted cash over accounts payable is only \$1500. Thus, the non-profit may have to borrow. An examination of the net assets section shows the type and extent of restrictions. Of the total difference between liabilities and assets, \$16,000, only \$6000 is not permanently restricted. The rest, \$10,000, is permanently restricted, which severely limits its ability to be used in the short term.

Statement of Activities

For the statement of activities, observe how the categories, unrestricted, temporarily restricted, and

	Private nonprofit organization name		
	Statement of financial position		
		6/30/x1	
Assets			
Cash – unrestricted		\$2,000	
Cash – temporarily restricted		3,500	
Interest receivable		50	
Contributions receivable		1,450	
Property, plant, equipment (net of depreciation)		12,000	
Long-term investments		10,000	
Total		\$29,000	
Liabilities and net assets			
Accounts payable		\$500	
Notes payable		5,000	
Long-term debt		7,500	
Total		\$13,000	
Net assets			
Unrestricted		\$2,000	
Temporarily restricted		4,000	
Permanently restricted		10,000	
Total net assets		\$16,000	
Total liabilities and net assets		\$29,000	

Fig. 3 Statement of financial position (balance sheet).

Revenue, gains, and other support	Unrestricted	Temporarily restricted	Permanently restricted	Total
Contributions	\$75,000	5,000	10,000	\$90,000
Income from investments	500			500
net unrealized and				
realized gains	500			500
Net assets released				
from restrictions				
satisfaction of program				
restrictions	1,000	-1,000		
Total	\$77,000	4,000	10,000	\$91,000
Expenses and losses				
Program A	35,000			
Program B	20,000			
Management and general	15,000			
Fund raising	5,000			
Total	\$75,000			\$75,000
Change in net assets	2,000	4,000	10,000	16,000
Beginning	0	0	0	0
End	\$2,000	4,000	10,000	\$16,000

Fig. 4 Private non-profit organization name; statement of activities; for the period 7/1/x0 to 6/30/x1.

permanently restricted are included, just as they are in the balance sheet (Fig. 4). Notice too, the bottom line is not profit or loss but change in net assets. Unlike the statement of financial position, the statement of activities is for a period of time, not a point in time. In this case the statement answers the question, “What has been the financial success (revenues matched against expenses) for the period 7/1/x0 to 6/30/x1?”

Overall, as presented in the total column, net assets for the non-profit have increased by \$16,000 for the period. Total revenue was \$91,000 and total expenses were \$75,000. However, when the total column is dissected, only \$6000 (\$2000 from unrestricted and \$4000 from temporarily restricted) of the change in net assets is not permanently restricted. Most of the change in net assets (\$10,000) comes from the permanently restricted category. Overall, the organization brought in more money than it spent (on an accrual basis) but a significant amount of the financial success is permanently restricted.

Statement of Cash Flow

The statement of cash flow is designed to show where the cash came from and where it went. For instance, if a private non-profit entity gets most of its cash from borrowing or grants, then future survival

may be in question. Will the grants continue and will contributions be sufficient to repay amounts borrowed and any associated interest payments? Raising cash via heavy borrowing will show in the statement of cash flow and such information can be vital to outsider readers of the statements. The statement of cash flow also shows whether the amount of cash changed (grew, stayed the same, or dropped) during the period.

SFAS 124—ACCOUNTING FOR CERTAIN INVESTMENTS

Non-profit entities often receive stocks and bonds as donations. SFAS 124 applies to equity securities (i.e., purchase or donation of stocks) with readily determinable market value and all investments in debt securities (e.g., purchase or donation of bonds). With SFAS 124 comes the interjection of “fair value” accounting and reporting (as opposed to historical cost accounting and reporting, which is used for most items such as property, plant, and equipment) and other rules related to investing in securities. On the surface, fair value accounting is not overly complex. Often called “mark to market,” it means that even those increases or declines in the value of securities not sold (so called “paper changes”) must be recognized at the end of the period. If the value goes up during the reporting

period, that is an unrealized holding gain. It goes on the statement of activities as a part of revenue. If the value falls, even without a sale, that constitutes an unrealized holding loss. It too goes on the statement of activities as loss subtracted from revenues. Unrealized gains and losses would also affect the balance sheet value of the investment. Presumably, outside users of financial statements (e.g., donors) are better informed about the financial performance and quality of management with the application of fair value accounting for securities.

The simplicity of “fair value” reporting stops at the conceptual level and becomes much more complex in implementation. The use of fair value depends on an array of circumstances, including for stocks, the amount of a company that a private non-profit owns. When a non-profit holds a large stock endowment in one company, the accounting becomes complex. A private non-profit may own 30% of a company from an endowment. If so, the accounting becomes even more complex and can likely move away from the fair value approach to an approach called the equity method where the non-profit shows the earnings or losses of the company as part of their own revenue and value. For instance, if a non-profit owns 30% of a company and the company loses \$90,000 dollars, then the non-profit shows a \$30,000 loss.

SFAS 136—TRANSFER OF ASSETS TO A NON-PROFIT

The complete title of this SFAS (136) indicates its focus. The full title is “Transfer of Assets to a Not-for-Profit Organization or Charitable Trust that Raises or Holds Contributions for Others.” As might be expected from the title, this SFAS is designed to answer the question about how a non-profit reports a contribution when the donor specifies another entity to ultimately or potentially receive the donation. SFAS 136 is of particular interest to federated fund-raising organizations. The United Way is an example of a non-profit that will be affected by this standard. It often collects donations and contributions that will be transferred to another non-profit organizations.

In the terminology of SFAS 136, the unit receiving the assets is the recipient while the unit that will or can ultimately get the assets is the beneficiary. The contributor is the donor. It is the donor’s specifications that affect the answer to how the recipient and beneficiary account for the donation. As with other accounting rules, SFAS 136 can become more complicated as is the case when the recipient and beneficiary are economically interrelated (one is a subsidiary of the other). Then both can share a stake in the donation.

SOP 98-2—ACCOUNTING COSTS OF ACTIVITIES OF NOT-FOR-PROFIT ORGANIZATIONS AND STATE GOVERNMENT ENTITIES THAT INCLUDE FUND-RAISING

People who donate assets to non-profit organizations as well as auditors who render opinions on non-profit financial statements have been concerned that amounts spent on fund-raising can be underreported to make it look like the non-profit is putting most of its expenses in mission-oriented programs. As a result, the American Institute of Certified Public Accountants wrote a statement of position (SOP) to clarify accounting and reporting for fund-raising, SOP 98-2.

Generically, the accounting term used to deal with this issue of allocating costs when more than one type of product is generated from the same process is “joint costs.” In business, a typical example is allocating a portion of a cost on one cut of meat when there is only one expense for the entire carving process. In a non-profit, an example is allocating the costs of postage, envelopes, labor, and machinery when educational or program information and fund-raising are included in the same mailing.

If an effort includes both fund-raising information and another activity such as the educational aspect of the social mission, allocation of joint cost to both program and fund-raising expenses can be used only if certain criteria are met. The activity must call on the audience to do something about the social mission. If the letter, for example, is directed toward environmental cleanup, then the letter must call for specific action such as attending a Saturday morning neighborhood cleanup. Simply saying that clean neighborhoods are important is insufficient for allowing the cost to be counted as program expenses as opposed to fund-raising expenses. If a specific, mission-oriented action is requested, an allocation between fund-raising and program expenses can be made.

Even the call for action may not be sufficient to allocate costs between fund-raising and program expenses. The audience selected should be people who believe in the mission. If the audience is selected because of past contributions, then all costs go to fund-raising. If the outside firm developing the campaign receives compensation on the basis of the assets donated, then all costs must be assigned as fund-raising costs.

CONCLUSION

Traditionally, private non-profits were considered to be driven by a social mission. In the mid-1990s, the accounting rule-making body for publicly traded companies, FASB, became a part of the life of private non-profits. As a result, private non-profit

organizations have a significant economic efficiency criterion to meet, and they did so, in part, by following the accrual and consolidation model (complemented by fair value) that FASB place on them. The new model demands more output for less input and hard decisions about projects that fail to break even—not exactly a “kinder, gentler” type of accounting!

Private non-profit entities are still adjusting to the dual forces of the competitive mentality of global markets and the world of social upheaval. In short, charity is in a new realm of competing in the market place for money and even for delivering services. Many for-profit agencies wish to sell the same services non-profits provide. The demands on private non-profit operations are considerable. They have their social mission to pursue, they must compete perhaps more

so than in the past, and they are considered to be an important part of maintaining world social order.

BIBLIOGRAPHY

1. Larkin, R.F.; DiTommaso, M. *Wiley Not-for-Profit GAAP 2003*; John Wiley & Sons Inc.: New York, 2002.
2. McLaughlin, T.A. *Streetsmart Financial Basics for Non-profit Managers*, 2nd Ed.; John Wiley & Sons Inc.: New York, 2002.
3. McMillian, E.J. *Model Accounting and Financial Policies & Procedure Handbooks*, Revised Edition Ed.; American Society of Association Executives: Washington, DC, 1999.
4. Salomon, L.M. *America's Nonprofit Sector: A Primer*, 2nd Ed.; The Foundation Center: New York, 1999.

Accounting, Financial

Mozaffar Khan

University of Toronto, Toronto, Ontario, Canada

INTRODUCTION

It is said that to truly understand a culture one must be proficient in its language. Financial accounting is the language of business. Just as linguistic constructs (or words) are used to articulate our unobservable thoughts to listeners, so also are financial accounting constructs used to communicate summary information to a firm's external stakeholders about the economic activities of the firm. Similarly, just as linguistic rules guide how we meaningfully string together words, so also do financial accounting rules govern what, when, where, and how information is summarized and transmitted to outsiders to best convey the firm's economic fundamentals.

The set of financial accounting rules is commonly referred to as generally accepted accounting principles (GAAP). There are two kinds of GAAP in the United States—GAAP for business and GAAP for government. The latter defines accounting standards for the various federal, state, and local governments and agencies. The primary rule-making body for business accounting standards in the United States is the Financial Accounting Standards Board (FASB), which derives its authority from the Securities and Exchange Commission. The primary governmental accounting standard setter in the United States is the Governmental Accounting Standards Board (GASB). Both the FASB and the GASB are arms of the Financial Accounting Foundation.

FINANCIAL ACCOUNTING FOR BUSINESS

The FASB's Conceptual Framework of Accounting is a collection of overarching objectives and principles that informs its promulgation of rules and serves to impose internal consistency on the set of rules. According to this framework, a primary objective of financial statement reporting is to provide information to investors, creditors, and other external stakeholders that is useful in helping them assess the amounts, timing, and uncertainty of the firm's future cash flows. Four of the primary desired qualitative characteristics of useful financial statement information are held to be

relevance, reliability, comparability, and consistency. Relevant information is that which is timely, facilitates prediction, and updates prior information. Reliable information is that which is faithfully represented, neutral, verifiable, and conservative. Comparability requires that the financial statements be presented in a uniform format, using standard accounting rules, across companies. Consistency requires that a given company have accounting policies, choices of accounting method, and format of financial statement presentation that are consistent over time.

The Business Financial Statement Reporting Environment

Considerable complexity is introduced into the financial statement reporting environment by the fact that it consists of multiple parties whose interests are not necessarily aligned with each other. This problem is addressed in the principal-agent theory, which examines the conflict of interest between the principal or owner of a business and the agent or manager hired to run the business. Typically, because the agent is involved in the day-to-day operation of the firm and the principal is not, the agent has more information than the principal about the future prospects of the firm. This information asymmetry gives agents the opportunity to further their own interests at the expense of those of the principal. In the financial reporting context, managers (who are the agents) may seek to release only self-serving information to investors and creditors (who are the principals). This is known as the adverse selection or hidden information problem, and its existence hinders the achievement of the financial reporting objective assumed in the Conceptual Framework of Accounting (the provision of useful information to investors and creditors). Thus, accounting rules have to be designed to mitigate the adverse selection problem.

Another principal-agent problem arises from the fact that efforts exerted by agents in discharging their responsibilities are not fully and directly observable by principals, thereby giving agents the opportunity to shirk. This is known as the moral hazard or hidden effort problem. Because firm performance is driven by

various factors other than managerial effort, it is not possible for the owner to fully infer managerial effort after observing firm performance *ex post*. Consider a firm that has high performance in a given year despite low managerial effort and low performance in another year despite high managerial effort. If managerial compensation is tied to firm performance, the manager gains *ex post* at the owner's expense in the first instance and the owner gains *ex post* at the manager's expense in the second instance. *Ex ante* however, neither party knows who is going to gain. The point here is that it is important to both the owner and the manager that performance measures are highly correlated with managerial effort so that effort may be reasonably inferred from performance. The implication for financial reporting is that accounting rules governing performance measurement should be designed to mitigate the moral hazard problem.

Both adverse selection and moral hazard result from information asymmetry. This asymmetry cannot be completely eliminated and can only be reduced at some cost. For example, an owner may devise elaborate monitoring and reporting schemes, but these would be neither perfect nor inexpensive. Financial accounting rules are therefore also designed to balance reporting costs with reporting benefits.

Business Financial Accounting Research

Financial accounting theory is rooted in economic theory. Theoretical financial accounting research involves building economic models that help understand, explain, predict, and control the financial accounting environment. These models may also guide the design of new accounting rules.

Empirical financial accounting research uses econometric methods to test theoretical models. It also tests whether new accounting rules better capture a firm's economic fundamentals. Typically, accounting researchers assume that capital markets are semi-strong form efficient, that is, that all publicly available information is reflected in the current stock price. If a new accounting rule allows financial statements to portray more accurately the underlying economic position of the firm then it should be useful to investors and creditors in making rational investment decisions. Thus, the establishment of the new rule should bring new information to the market and, given market efficiency, should lead to revised market expectations of the amounts, timing and uncertainty of future cash flows. These revised expectations can lead to any of three testable effects: they may lead to abnormal stock returns, abnormal trading volume and/or abnormal stock price volatility. The presence of any of these effects then is

held to be evidence of the decision usefulness of the new accounting rule.

Selected Topics in Business Financial Accounting Research

Voluntary disclosure

This literature examines why managers issue forecasts of quarterly or annual earnings per share (EPS) and otherwise reveal to the securities markets private information about the firm's future prospects when they are not mandated to do so by SEC disclosure rules. Four suggested motives and reasons for voluntary disclosure are legal liability, adverse selection, expectations adjustment, and transactions cost. In the following discussion, news refers to private EPS information relative to current public expectations of future EPS (e.g., bad news refers to private management information that actual EPS will be substantially below what the market expects, while no news refers to information that actual EPS will be very close to expected EPS, either from above or below). Under the legal liability argument, managers who suppress bad news may be successfully sued by investors for the loss suffered once actual performance is realized and disclosed in the future, and the stock price plummets. Thus, managers are expected to preemptively disclose only bad news prior to the mandated earnings announcement date to mitigate their legal liability. Under the adverse selection hypothesis, managers preemptively disclose only self-serving or good news to immediately boost the price of their firm's stock. The expectations adjustment hypothesis posits that managers preemptively disclose both good and bad news, but not no news, to "correct" the market's expectations. The assumption is that the market does not like negative surprises (or bad news) on earnings announcement dates and that the firm is better off revealing positive surprises (or good news) as early as possible. Finally, the transactions cost hypothesis argues that a firm's managers will disclose all private information to investors to lower investors' information costs in acquiring that firm's stocks and to increase investor confidence in the firm. This in turn will translate into a lower cost of capital for the firm.

The empirical evidence is mixed as to which motive best explains voluntary disclosure, but the transactions cost hypothesis appears to be the weakest of the four. For more information, see Refs.^[1-3]

Positive accounting theory (PAT)

GAAP often allows managers some discretion in their choice of accounting method (first in, first out vs. last

in, last out for inventory, straight line vs. accelerated method for depreciation, etc.). Given discretion, why do firms make the choices they do? Note that an accounting policy simply determines how an economic event is portrayed. It does not change the economics of the event. Inspired by positive economics and motivated by agency theory, PAT attempts to explain firms' accounting policy choices. In other words, it attempts to explain why certain economic events are accounted for in the manner they are. It argues that a firm can be viewed as a web of contracts between multiple parties (managers, investors, creditors, regulators, consumer groups), with each party seeking to achieve its interests over, and perhaps at the expense of, those of others.

For example, consider the conflicts in the manager-stockholder relationship. If managers are compensated based on reported earnings, then they will choose income-increasing accounting policies so that their compensation will increase despite the fact that there has been no change in the economic value of the owners' equity in the firm. Another conflict arises when outside investors attempt a takeover of the firm. Although their offer may be in the best interests of the current stockholders, it may be rebuffed by managers if they fear that they will be displaced by the new owners.

Consider, too, the conflict in the creditor-stockholder relationship. Managers, acting on behalf of stockholders, may attempt to circumvent financial limits stipulated in the debt indenture to avoid triggering a technical default. For example, if the indenture calls for the firm to maintain a certain minimum level of return on assets, then managers may opportunistically make accounting policy choices that increase reported income or decrease average total assets when they are in danger of falling below the specified minimum.

Finally, firms are also believed to have a latent social contract with society. If an industry or firm finds itself reaping excess or abnormally high profits, then it may attract the scrutiny of governmental regulators or consumer groups. For example, if firms in the oil industry rake in excess profits when the price of gasoline is high, then they may be suspected of price gouging. Positive accounting theory predicts that such firms will attempt to depict earnings as being lower than they are and will thus make accounting choices that decrease reported earnings to avert attention.

Empirical research has found evidence of accounting choices broadly consistent with those predicted by PAT. For more information, see Refs.^[4-6].

Earnings management (EM)

This literature examines whether managers opportunistically manipulate reported earnings up or down.

For example, if current earnings performance is poor, then managers may "borrow" earnings from future periods to increase current reported earnings to acceptable levels, or if current earnings are exceptional, then they may "store" some earnings to use in future periods. The motive for this may be to meet market expectations or to maximize their bonus. Another example of EM is a "big bath"—a new chief executive officer (CEO) may have the management team aggressively record write-offs and other charges in the current year so that good performance is easier to achieve in future years. In effect, the new CEO "cleans house," knowing that a huge loss in the current year will be blamed on the predecessor and good performance in future years will be self-attributed. Earnings management may also occur when a domestic industry petitions the federal government for import relief. This industry may manage its reported earnings down to bolster its claims of damage from foreign competitors and their products.

This literature assumes that managers believe that capital markets are fixated on the reported earnings figure and that the market is unable to fully detect EM. If managers believed otherwise, then they would not engage in EM because doing so would be futile. Clearly, managers would like to manipulate earnings surreptitiously. The literature therefore generally assumes that EM occurs through the manipulation of accrual accounts because this is the least easily detected.

Empirical researchers have attempted to identify specific instances of EM to prove that the phenomenon exists in reality. For any given firm, only its total accruals are observable or identifiable. The managed portion of total accruals cannot be identified from financial statements. Typically, the researcher will specify a statistical model that generates the expected level of "normal" total accruals in any given year. The difference between expected accruals and actual total accruals is the "abnormal" accrual for that year. Abnormal accruals are held to represent the manipulated portion of total accruals. This decomposition of observable total accruals into its unobservable normal and abnormal components is heavily dependent on the mechanism that is assumed to generate "normal" accruals. Thus, findings of abnormal accruals constitute evidence of EM only insofar as the normal accrual-generating model is correctly specified. For more information, see Refs.^[7] and ^[8].

Anomalies

As mentioned previously, a key maintained hypothesis in financial accounting research is that capital markets are efficient. The anomalies literature attempts to debunk this hypothesis by investigating instances where market efficiency is apparently violated, that is, where

public information that is marginally useful in predicting a firm's future performance is apparently not incorporated in stock prices. The typical approach adopted by researchers is to use some public information to develop a trading strategy that yields excess or risk-adjusted returns. The implication then is that this information could not have been incorporated in stock prices and that the markets are inefficient. For example, certain financial statement information may be used to predict which firms will experience robust performance and which will experience poor performance. Next, a long position may be taken in the first set of firms and an off-setting short position in the second set of firms, under the assumption that the stock price will rise in response to strong realized but unanticipated performance and decline if performance turns out to be poor. If markets are efficient, then future performance implied by all current public information should be anticipated in the current stock price and there should be no stock price reaction once performance is realized as anticipated. However, if the market is not efficient with respect to some information, then the previous trading strategy should yield positive net returns.

Another example of an anomaly is the postearnings announcement drift. It is observed that the stock price of firms that announce robust earnings performance tends to drift upward for an appreciable period of time after the earnings announcement date, and a similar downward drift is observed for firms announcing poor performance. The anomaly here is that this phenomenon is a documented empirical regularity; that is, there continues to persist this period of time over which price movements are predictable. Consider a firm announcing poor current earnings. If investors are aware of the drift phenomenon, then they should immediately short the stock once earnings are announced. The stock price will drop immediately as a result, and no drift will occur thereafter. However, this does not happen. The drift phenomenon continues to be observed.

A third example of an anomaly would be an abnormal stock return response to an accounting method change. In efficient markets, an accounting method change should generate no abnormal returns because it merely changes the way an economic event is accounted for, or depicted, and does not change the underlying event itself.

One explanation for these anomalies is that researchers who have developed hypothetical trading strategies to exploit apparent market inefficiencies have not taken transactions costs or short-selling constraints into account. The costs incurred in searching for mispriced securities and buying or selling them can be substantial, and may offset the marginal benefit of engaging in such activity. Another explanation is

that risk factors may not be fully controlled for. For example, market efficiency would call for abnormal returns in response to an accounting method change if this change also leads to a change in the firm's risk.

It is important to note that market inefficiency does not imply investor irrationality. An ostensible profit opportunity may not be exploited if the mispriced stock is available only in limited quantities or if the transactions costs are prohibitive even though the stock is available in unlimited quantities.

Although significant evidence of apparent market inefficiency has been tendered, it has not been as unsailable as to warrant abandonment of the assumption of market efficiency, especially in the absence of refutable alternative hypotheses. For more information, see Ref.^[9].

Earnings response coefficient (ERC)

In an efficient market, a firm's stock price changes only in response to revised investor expectations about the amounts, timing, and uncertainty of its future cash flows. One stimulus for expectation revision is unanticipated earnings. Faced with an earnings surprise, investors try to assess how long they can expect the unanticipated portion of earnings (termed the earnings innovation) to persist. This persistence assessment is then fed into an asset valuation model, such as the discounted cash flow (DCF) model, to yield a new price. The earnings persistence is a parameter of the firm's time series of earnings. Typically, a pure random walk is held to be a reasonable description of a firm's time series of annual earnings, implying that all earnings innovations are permanent. Given this, and assuming a one-to-one link between earnings and cash flows, the DCF model simply prices a stock with a dollar earnings innovation as the current stock price, plus a dollar, plus the value of a dollar in perpetuity.

The ERC is the marginal stock price change in response to a \$1 earnings innovation. For example, given the previous assumptions, and a discount rate of 10%, the ERC should be $1 + 1/0.1 = 11$. Empirical ERC estimates have been substantially lower, however, spawning a large literature attempting to explain this attenuation. One explanation is that a portion of the supposedly unanticipated earnings was actually previously anticipated by market participants, so that the earnings innovation or surprise is measured with error. Attenuation due to measurement error in the earnings surprise can be mitigated by using a reverse regression econometric design. Another explanation for the observed attenuation is that a portion of the unanticipated earnings may be transitory; that is, it may have resulted from one-time gains or losses rather than from continuing operations. However, empirical ERC estimates have been shown to remain attenuated,

even when transitory items are removed from reported earnings. Finally, it has been argued that the marginal investor response to losses is lower than that to profits because losses have less information content than profits about the future prospects of the firm. Losses are less informative than profits because losses are not expected to occur in perpetuity. They can also be less informative as a result of the application of the conservatism principle of accounting, under which managers are quicker to recognize foreseeable losses than to recognize foreseeable gains. Reported gains are a lot more certain than reported losses. For more information, see Refs.^[10–13].

FINANCIAL ACCOUNTING FOR GOVERNMENT

The goals and accounting environments of government and business differ. These factors lead to differences between GAAP for government and GAAP for business.

Although the goals of business are typically to maximize return on investment, the goals of government as recognized by GASB are twofold—accountability to citizenry and promoting the common welfare. These give rise to the external financial reporting objectives of the GASB. One objective is to demonstrate accountability. This implies providing information that is useful in assessing accountability, including such information as whether there was intergenerational shifting of resources. For example, will some current-year services be funded partly by future-year citizens, or are some future-year benefits funded by current-year citizens? Another objective is to report on current operating results by revealing the sources and uses of revenues. A third objective is to report on the entity's financial condition in terms of its long-term financial viability. Finally, a fourth objective is to report on long-lived physical and non-financial assets, and their future service potential.

The Governmental Financial Accounting Environment

In setting accounting standards, the GASB is mindful of certain characteristics that are peculiar to the governmental accounting environment. One salient characteristic of this environment is that decision making within a governmental entity is far more political, because of its public nature, than decision making within a corporate entity. In addition, governments have to balance the needs of diverse constituents who often have conflicting notions of the common good. Another characteristic is the use of fund accounting by governments. A fund is its own fiscal and accounting

entity, set up to ensure availability of resources for a specific purpose. The monies from a given fund can only be applied toward the designated purpose of that fund. Usually, the General Fund is the only unrestricted fund in a government entity. Although separate financial reporting for each fund is no longer required, each fund still has to maintain its own books. Yet another characteristic is that governmental budgets are legal documents in that they often carry the force of law. They cannot be altered once passed. They are also public documents because they are products of the public intent as revealed through public hearings and dialog. This is in contrast to corporate budgets that have the discretionary flexibility to respond to changing conditions through the fiscal year and that are secret for competitive reasons.

A fourth characteristic is that some entity may be affiliated with a governmental entity, yet offer private services. The conflict here is whether to follow government or business GAAP. For such entities, the compliance hierarchy is GAAP for government first, then GAAP for business, followed by other sources of accounting standards such as experts, common practice, and the accounting literature. A fifth characteristic is that resources are often transferred from one level of government to another, requiring intergovernmental accountability. For more information, see Refs.^[14] and ^[15].

Governmental Accounting Research

Research in this area examines diverse questions, such as performance measurement, the impact of incentives, the impact of politics and the tenure of elected officials, discretionary accounting practices, and the accounting and auditing impact of intergovernmental revenue sharing, to name a few.

A prominent issue in governmental accounting today is whether to move toward accrual-based accounting as is required for businesses by business GAAP. Governmental accounting today typically follows a modified accrual method, thereby hiding such future liabilities as past service (or pension) costs.

CONCLUSION

Financial accounting is a vast and vibrant field. Accounting practice (both for business and for government) is constantly evolving to meet changing needs, and accounting research is gaining rigor in theory as well as sophistication in empirical methods. It is impossible to understand the culture and environment of financial management without understanding financial accounting.

REFERENCES

1. Ajinkya, B.B.; Gift, M.J. Corporate managers' earnings forecasts and symmetrical adjustments of market expectations. *J. Acc. Res.* **1984**, *22* (2), 425–444.
2. Lang, M.; Russell, L. Cross-sectional determinants of analyst ratings of corporate disclosures. *J. Acc. Res.* **1993**, *31* (2), 246–271.
3. Scott, W.R. *Financial Accounting Theory*, 2nd Ed.; Prentice-Hall Canada Inc.: Scarborough, Ontario, 2000.
4. Press, E.G.; Weintrop, J.B. Accounting-based constraints in public and private debt agreements: Their association with leverage and impact on accounting choice. *J. Account. Econ.* **1990**, *12* (1–3), 65–95.
5. Watts, R.L.; Zimmerman, J. *Positive Accounting Theory*; Prentice Hall: Englewood Cliffs, NJ, 1986.
6. Zmijewski, M.E.; Hagerman, R.L. An Income Strategy Approach To The Positive Theory Of Accounting Standard Setting/Choice. *J. Account. Econ.* **1981**, *3* (2), 129–149.
7. Dechow, P.M.; Sloan, R.G.; Sweeney, A.P. Detecting earnings management. *Account. Rev.* **1995**, *70* (2), 193–225.
8. Jones, J.J. Earnings management during import relief investigations. *J. Acc. Res.* **1991**, *29* (2), 193–228.
9. Ou, J.A.; Penman, H.P. Financial statement analysis and the prediction of stock returns. *J. Account. Econ.* **1989**, *11* (4), 295–329.
10. Collins, D.W.; Kothari, S.P. An analysis of intertemporal and cross-sectional determinants of earnings response coefficients. *J. Account. Econ.* **1989**, *11* (2/3), 143–181.
11. Das, S.; Lev, B. Non-linearity in the returns-earnings relation: Tests of alternative specifications and explanations. *Contemp. Account. Res.* **1994**, *11* (1-II), 353–379.
12. Easton, P.D.; Zmijewski, M.E. Cross-sectional variation in the stock market response to accounting earnings announcements. *J. Account. Econ.* **1989**, *11* (2/3), 117–141.
13. Hayn, C. The information content of losses. *J. Account. Econ.* **1995**, *20* (2), 125–153.
14. Chan, J.L. The governmental environment: characteristics and influences on governmental accounting and financial reporting. In *Handbook of Governmental Accounting and Finance*; Apostolou, N.G., Crumbley, D.L., Eds.; John Wiley & Sons: 1988.
15. Johnson, W.L.; Goodwin, G.L. State and local government accounting principles. In *Handbook of Governmental Accounting and Finance*; Apostolou, N.G., Crumbley, D.L., Eds.; John Wiley & Sons, 1988.

BIBLIOGRAPHY

1. Ball, R.; Brown, P. An empirical evaluation of accounting income numbers. *J. Acc. Res.* **1968**, *6* (2), 159–178.
2. Beaver, W.; Lambert, R.; Morse, D. The information content of security prices. *J. Account. Econ.* **1980**, *2* (1), 3–28.
3. Skinner, D.J. Why firms voluntarily disclose bad news. *J. Acc. Res.* **1994**, *32* (1), 38–60.

Achieving Productivity through Budgeting

Gerald J. Miller

Department of Public Administration, Rutgers University, Newark, New Jersey, U.S.A.

Donijo Robbins

School of Public and Nonprofit Administration, Grand Valley State University, Grand Rapids, Michigan, U.S.A.

INTRODUCTION

Budgets serve many purposes, all of which can promote productivity. The budget serves as a plan, a means of control, a motivation device, and a process for accountability. At its narrowest, a budget is a plan that estimates proposed expenditures for a given period, usually for a year or a “fiscal” year, and relates planned spending to an estimate of revenues needed to finance the plan. The budget also acts as a control device, directing that only the amount specified in the budget be spent in the way the budget dictates. In this sense, budgets control the amount of scarce resources fed into the governmental system, as well as the size and nature of the programs within the budget. Budgets offer incentives, implying a contract in which a set amount of performance or results is exchanged for a certain amount of resources. Finally, budgets improve the accountability of government to citizens by making budget decision making public. That is, government budgets reveal to the public the amount of money spent, with what results, and often why it was spent.

BUDGETING AND PRODUCTIVITY

Budget designers create a process that forces managers to consider different programs, often in competition within and between government departments. Managers commonly employ a simple economic idea—that any alternative must be judged in terms of other alternatives. Budgeting for productivity involves analysis, and proponents of analysis argue that explicit or systematic examination of alternative policies allows the public to hold its public officials accountable for performance.

Budget designers have productivity concerns in mind. In narrow terms, productivity may refer to economic efficiency, a surplus of the goods or services produced over the resources used in their production. Closely akin to this idea are two others: achieving the highest levels of effectiveness for a given investment

and choosing the least costly investment that will yield a given level of performance. These often translate through budgets into managerial plans, controls, and incentives for more innovative programs and greater performance.

Another view of productivity holds that budgets should steer government agencies to spending on projects with large, durable returns on investment. Budgets can favor productivity by encouraging choice and favoring alternative investments that have larger returns for the public resources committed to projects, agencies, and activities.

Political leaders make ultimate budget decisions. Political decisions are often based on the geographic distribution of spending and log rolling in which political leaders trade support for each other's favored spending plans. Compromise-driven politics tend to dampen any resolve to increase the effectiveness or efficiency in government programs. Much more emphasis gets placed on maintaining the existing distribution of budget resources, thereby making innovative choices seem irrelevant to the process.

Given all the effort to make the budgeting process favor more productive programs, politics still overshadows economic efficiency as the basis for budget decisions. Thurmaier's research supports this.^[1] He suggested that “conflicts between political and economic imperatives are handled subtly in practice, but it appears that economic and technical calculations are subordinated to prior political determinations.”

Since the era of modern budgeting began in the early twentieth century, government budgets have undergone various reforms. First, the line item budget was a rudimentary way of recording and balancing expenditures and revenues; as such, the line item budget produced accountability. Early performance budgeting was a method of going one step beyond the line item budget by classifying items by function and determining the efficiency of the budgeted expenditures. Functional spending revealed how much revenue was needed to produce a certain level of program production. Next, there was program budgeting, which combined

planning and budgeting, and promoted the economic analysis of allocation choices. Zero base budgeting decentralized decision making in an effort to redirect funds from existing programs to new programs. Finally, performance base budgeting attempts to tie the allocation of resources in budgets to the improvement in program performance, the achievement of results, and the success of efforts to realize strategic outcomes.

The success of these reforms has not been impressive. Few have lasted beyond the administration of the executive who instituted them, or they have lasted in only a single department. Although many contextual factors may have led subsequent administrations to seek other methods of budgeting, most reforms have failed to combine a sharp sensibility for both productivity and politics. One former analyst for the legislative branch observed that the “failure of [reform] efforts . . . are largely a result of an inability to account for politics. No set of budget techniques can substitute for political decisions about who wins and who loses.”^[2]

PRODUCTIVITY-ORIENTED SPENDING DECISIONS

If the conceptual linkage between productivity and budgeting is stronger than that revealed in the practice of budgeting, efforts attempt to narrow the difference. These greater efforts have rested in the major and minor modifications of productivity improvement and budget practices; productivity-friendly and budget-friendly reforms may bridge the gap between the concepts and the practices of productivity improvement and budgeting. Productivity-friendly elements refer to reducing the budget process’ outright exclusion of productivity goals. Budget-friendly elements empower productivity programs to work inside budget routines.

Productivity-Friendly Budget Elements

Budgeting serves many purposes for decision makers. Lauth^[3] warned of the pitfalls of using budgeting as the handmaiden for productivity improvement. The budget is simply too cumbersome to help, he said. Rather, the budget can be made less hostile and more friendly to the ideas underlying productivity improvement. Three major groups of efforts narrow the differences between management and politics: technical modifications, support for measurement, and incentives.

Technical modifications

Small, technical modifications could overcome some of the barriers between budgeting and productivity. For example, many opportunities for cost savings through

transfers among budget categories and line items within the fiscal year are forbidden without legislative approval. To illustrate, purchases or repairs may be deferred due to this restriction, forcing more costly ones later. Budgeted personnel lines are seldom able to be converted to any other object class or vice versa. Thus, consultants may be allowed for many times the expense of a fully trained, permanent employee. Finally, central finance offices often pull back from productive line units salary savings due to attrition, unpaid leads, or other temporary absences. The department is shackled to lower-level performance—fewer people to do the same amount of work—without the opportunity to use the same salary savings to reward employees for what might be a temporary bulge in workload.

Support for measurement

Productivity and budgets often stand at arms length when the performance-to-budget dollar measurements became complicated or numerous. In truth, agreement about valid measures of public sector work performance only exist on the input side, such as dollars given to school districts. Many blame time limits, training gaps, and primitive computing capability for the lack of development in the other areas, whereas others blame the incoherence in the organization segregation and information flows between program evaluation and budgeting.

In one of the clearest insights into the operation of budget offices, Grizzle^[4] dispelled much doubt about the ability to solve these problems. She found that many efforts to create time and develop information would involve negligible amounts of the central budget office’s work year and little cost.

Her view is that measurement problems could be less considerable and more mythical than believed, if measurement were merely a technical problem, not a problem that becomes politicized. When measurement is a political problem, it becomes one of several behavior problems for which budget officers blame incentives.

Incentives

The motivation to perform productively gets caught in the cross-currents produced by budgets. There are, first of all, incentives to misrepresent information in budgeting, thereby confusing the measurement problem further.^[5] Organization, program, and job survival become the ultimate goals when decision makers outside the organization threaten budget reallocations. The incentive for change and innovation has often come from making those inside the organization both responsible for and rewarded for reallocation.

Just as there are incentives to misrepresent, there are incentives to work in less than the most productive ways. Incentive pay is the best known antidote. The availability of funds and careful design of incentives for individuals is important. Group incentive plans are another variation. As reported by the U.S. General Accounting Office in 1983,^[6] for example, North Carolina state employees could propose programs for which they would be rewarded as a group, with a quarter of the savings, if greater productivity were realized. The state budget office chose those for which documentation might provide clear assessments of savings. Strong political support for such incentive programs must exist. In the North Carolina case, Lauth^[3] reported, the legislature terminated the program in less than 10 years because the program did less to reduce costs than it did to induce group incentives.

Budget-Friendly Productivity Elements

At another level, the productivity program can profit by major changes in the budget system. The budget system itself has been changed, in this case to one of two major systems, an expenditure control budget and a target base budget. Short of wholesale system change, retained savings plans remove the biggest disincentive to productivity improvement, the transfer of budget surpluses to the central treasury at the end of the fiscal year. Still short of a savings plan, calculating mechanisms may be added to budget systems. Finally, research on policy and program outcomes suggests that more effort in this direction may hold considerable promise.

First developed in California cities in the 1970s, expenditure control budgeting (ECB) has five elements, according to Herzik.^[7] First, the system relies on a base budget figure adjusted for population and prices level changes. Second, ECB base budgets assume existing service levels and require explicit approval of changes. Third, departments create retained savings plans, developed through productivity plan-induced savings, to finance changes in service levels or increases in programming. Fourth, central budget offices permit departments to carry forward savings realized in a fiscal year. Fifth, ECB relies on line item control of spending.

The most important aspects of ECB are selection of a base year and maintenance of department head interest in service delivery innovation. Base years may be different for agencies because, at any time, some have unique expenditures, such as capital projects, and others do not. Careful scrutiny prevents one department's base budget from being inflated, creating unfairness.

Second, a department normally having considerable political clout and able to gain bigger budgets to meet service demands may feel unfairly treated under ECB. Because all expansion must be paid out of savings, the department resists. Top management action to force attention to innovation consists of making the sole route to expansion the production of savings.

Target base budgeting^[8,9] relies on excellent revenue forecasting to establish a target. Given that accuracy, central budget offices allocate a target figure to departments within which existing service delivery costs must fit. New programs must pass muster separately in budget requests.

The existing services or target budget, even though revenue driven, force priority setting and often innovations. Rubin pointed out that the targets are assigned as a percentage of the base budget for a department. The targets "may be less than the department's base, by a percent that varies from year to year, and may vary from department to department."^[8] The target is protected, however, in that it is often an appropriated lump sum, with review of productivity or performance goals and not spending proposals, allowing various trade-offs and transfers within departments, between lines, between programs, and even between operating and capital budgets.

The unprotected or new budget requests to expand services to compete with those of other departments. The list of a department's projects is submitted in priority order. All lists get substantial scrutiny from budget examiners, Rubin stated. These lists force detailed scrutiny to the margin, making target base budgeting rational but not comprehensive.

Although ECB explicitly mandates retained savings and target base budgeting creates savings that may be reallocated, such savings plans may exist independently of a budget system. Klay^[10] reviewed these savings plans and described two types—conditional appropriations and discretionary savings plans.

Klay^[11] proposed that legislatures give state agencies direct incentives to innovate by appropriating the agency's funds in two parts—one unconditionally and the other conditionally. The legislature would authorize a part of an agency's program on the condition that the agency generate savings in its unconditionally appropriated program. If the agency is given appropriations for normal service delivery and the highest-priority program expansions in its unconditional appropriation, the legislature might also allow the agency to use savings on lower-priority expansions. The legislature might actually premise supplemental appropriations, especially when revenues are underestimated, on the generation of savings.

To retain some leverage over agencies, especially when the agencies that generate savings are not agencies that central budget offices or legislatures consider

to be well managed, finance officers might distribute savings, at their discretion, when productivity goals are met. Klay^[10] argued that also giving these same well-managed agencies their savings would provide substantial incentives. Discretionary savings plans might also be used in cases where several agencies competed in supplying essentially the same or similar services. Many pointed out the productivity potential of interagency competition, especially among agencies such as city governments, school boards, county governments, and independent districts, which perform some similar services. State oversight agencies, which have some control over the subnational governments, might retain discretion to reward those that produce the service most efficiently or that develop the greatest level of innovation. The ability to use retained savings might provide an incentive, and merger of service providers might aid in gaining effort toward either efficiency or innovation.

Ultimately, budget-friendly productivity elements must include outcome measures. Providing equitable incentives across agencies may instill a will to innovate, be efficient, or attain goals. Attitudes may portray the perceived needs of association members, organization customers, or agency clients. Allocation, however, demands an unbiased measure of achievement to which incentive programs may urge effort and against which attitudes may be viewed and balanced. To force budget-friendly performance decisions, therefore, there must exist outcome measures and a system to monitor work with these measures and evaluate performance itself.

Fisk^[12] warned of the difficulty in finding outcome measures and suggested three basic types of measures, operational, immediate, and consequential. Operational measures are those that describe the internal efficiency of an agency, including, for example, the number of reports produced, audits completed, and equipment downtime. Such measures help in the day-to-day management of an agency, but, strictly speaking, they are not productivity measures, as are immediate and consequential measures.

Immediate measurements are those that depict the final organizational output divided by the resources used to produce it or its inputs. For example, tons of garbage collected per employee hour provides a ratio, a technical efficiency measure. The shortcomings of these measures relates to their lack of power to explain why an organization collects garbage—is garbage collection a public health matter or a issue of aesthetics?—or how well we are doing it. Knowing that we are collecting garbage for public health reasons tells managers what priority to assign kinds of garbage and, therefore, where to go to collect it.

A consequential measure is one Fisk said that^[12] “addresses the issue of a program’s impact on society

and whether the program makes optimum use of resources to achieve its goals.” Knowing that an organization collects garbage to protect the health of the public can provide an avenue to explore, as above, the kinds of garbage to collect and where to go to collect the garbage. Other examples of consequential measures include deaths prevented per employee hour for a fire department and jobs created per employee hour for an economic development project. These measures directly relate organization members, customers, and clients to the organization’s work.

Although immediate and consequential measures form the bulk of traditional performance measurement in government agencies, relating them is an effort of more recent vintage. Swiss^[13] used the concept of “output chains” to show that budget-friendly productivity elements can be both valid and highly relevant to managers. That is, chains relate the question, “Did we perform the job without wasting resources?” to the other question, “Did the job achieve the desired result?” In our fire department example, we would establish first the number of fires fought (internal efficiency), then the cost per fire fought or fire fought per employee hour (immediate measures), and finally deaths prevented per fire or per employee hour (the number of reasonably foreseeable deaths without any fire action minus the actual number of deaths divided by the number of employee hours).

Finally, the measure of input used in immediate and consequential measures has traditionally been related to labor productivity. Labor costs dominate budgets, are relatively easy to calculate, are readily available in existing datasets, and have private sector data available for comparison. Simply, labor productivity reveals hours paid.

Because budgets, organization structures, and systems of work are based implicitly on output chains, making these chains explicit demands considerable work but relates technical efficiency measures based on labor productivity to immediate measures to consequential measures and then, most important, provides feedback. Feedback systems that use the outcome measures have budget consequences. Simply, the system should reveal the overall performance of a given program or organization and answer the questions raised about whether results are achieved.

What the measures emphasize may or may not jibe with the organization’s goals, with fairness, or with the budget’s emphasis. Consider lessons for social welfare entitlement programs learned from one past program initiative^[14,15] and the budgeting process connected with it. Measures were developed for overpayments and ineligible recipients. The system was monitored and state agencies administering the program sanctioned when their rates substantially exceeded federal standards. Yet, the system did not ferret out those

underpaid or those who were eligible but not receiving benefits. Truly, the organization's goals were minimized and unfairness increased, and the initiative forced displacement of the long-term economic goals of budget policy.

The initiative also splintered efforts to evaluate the program and guide its development in a budget-friendly way. First, the effort to evaluate the program went in two directions: one political to buttress the claim that the program had eliminated welfare cheats, and the other economic to justify the use of transfer payments to provide income floors—a safety net—for the young, the disabled, and the poor. Second, the initiative failed to guide budget-friendly productivity efforts. The initiative, as many who oppose it now look back and see, failed to discourage dependency. It did nothing to offer alternatives to transfer payments, more or less emphasizing greater spending in the future as transfer payment dependency reached future generations.^[16] Moreover, the initiative's implicit budget policy became cost centered. Cost-centered budgeting, generally, rewards projects that cost little, whatever their productivity value. Cost-centered budgeting leads eventually to zero-sum gaming where increased spending in one area means decreased spending in others and where deferred maintenance (keeping destitute or needy individuals whole) supports new projects.

CONCLUSION

A budget is a plan, a set of controls, a potentially large set of incentives, and a means of holding governments accountable for both what work is done and how much the work costs. Productivity and budgets would seem to be self-reinforcing and dynamic in encouraging innovation and performance. Many attempts have occurred over the years to change the format of the budget to make the conceptual linkage between productivity and budgeting stronger in practice. Many agree that these reforms succeeded in some ways, but failed to survive the politics of budgeting. Further, less ambitious efforts also exist to produce a politically acceptable connection. These efforts include approaches that lead budgeting to include productivity-friendly elements and ideas by change budgets to focus on productivity concerns.

REFERENCES

1. Thurmaier, K. Decisive decision making in the executive budget process: analyzing the political and economic propensities of central budget bureau analysts. *Public Adm. Rev.* **1995**, 55, 448–460.
2. Joyce, P.G. Using performance measures for federal budgeting: proposals and prospects. *Public Budg. Finance* **1993**, 13 (4), 3–17.
3. Lauth, T.P. Budgeting and productivity in state government: not integrated but friendly. In *Performance-Based Budgeting*; Miller, G.J., Hildreth, W.B., Rabin, J., Eds.; Westview: Boulder, CO, 2001; 191–202.
4. Grizzle, G. Linking performance to funding decisions: What is the budgeter's role? *Performance-Based Budgeting*; Miller, G.J., Hildreth, W.B., Rabin, J., Eds.; Westview: Boulder, CO, 2001; 203–214.
5. Jones, L.R.; Euske, K.J. Strategic misrepresentation in budgeting. *J. Public Adm. Res. Theory* **1991**, 1, 437–460.
6. U.S. General accounting office. *Increased Use of Productivity Measurement Can Help Control Government Costs*; U.S. Government Printing Office: Washington, D.C., 1983.
7. Herzik, E.B. Improving budgetary management and fostering innovation: Expenditure control budgeting. *Public Prod. Manage. Rev.* **1991**, 14, 237–248.
8. Rubin, I.S. Budgeting for our times: Target base budgeting. *Public Budg. Finance* **1991**, 11, 5–14.
9. Wenz, T.; Nolan, A. Budgeting for the future: Target base budgeting. *Public Budg. Finance* **1982**, 2, 88–91.
10. Klay, W.E. Management through budgetary incentives. In *Performance-Based Budgeting*; Miller, G.J., Hildreth, W.B., Rabin, J., Eds.; Westview: Boulder, CO, 2001; 215–227.
11. Klay, W.E. A legislative tool to encourage agency efficiency. *Public Prod. Rev.* **1978**, 3 (1), 23–31.
12. Fisk, D. *Measuring Productivity in State and Local Government*; Bulletin 2166, U.S. Department of Labor, Bureau of Labor Statistics, U.S. Government Printing Office: Washington, D.C., 1983; 2 pp.
13. Swiss, J.E. *Public Management Systems: Monitoring and Managing Government Performance*; Prentice-Hall: Englewood Cliffs, NJ, 1991.
14. Danziger, S. *Approaching the Limit: Early Lessons from Welfare Reform*; Joint Center for Poverty Research Conference: Chicago, 2000.
15. Randall, R. Presidential power versus bureaucratic intransigence: The influence of the Nixon administration on welfare policy. *Am. Polit. Sci. Rev.* **1979**, 73, 798–800.
16. DeParle, J. To Moynihan, welfare dependency signals new ill. *N.Y. Times* December 9, **1991**, 141, A13.

Acquiring Resources through Price Negotiation: A Public Sector Approach

Rupert G. Rhodd

Department of Economics, Florida Atlantic University, Davie, Florida, U.S.A.

INTRODUCTION

Commercial negotiation has been with us for a long time, some accounts going as far back as when ships first sailed to China to buy silk and spices. In the pre-Industrial Revolution era the small scale of manufacturing operations meant that the entrepreneurs had to be skilled in everything including the purchasing of raw materials. If entrepreneurs were able to determine the quality of the inputs and, at the same time negotiate a price to minimize the average cost of inputs, the greater the demand for their product and the higher the profit margin. With the coming of the Industrial Revolution and large scale production there was a role for purchasing agents. In this period negotiation focusing on quality, delivery, and service “as industrial sellers customarily set prices at or near cost plus 10%.”^[1] The focus of negotiation shifted in the 1930s when sales people were granted the authority to set price at any level. Today, negotiation focuses solely on securing the best possible price.

For any organization, achieving long- and short-term goals and objectives depends on a host of economic relationships including the demand and supply of resources, which determine the price of resources. With increasing population and demand for government services, and with budgetary constraints, negotiating the optimal price and quantity of resources that are required to deliver government services has become an important activity of government agencies. This is recognized by the increased role and prominence of procurement departments/personnel in government agencies. Through negotiation, the purpose of these departments/personnel is “to secure the best possible long or short-term agreement for the organization, consistent with the concept of lowest total cost.”^[1]

Negotiation, as an important component of the organization’s strategy to acquire resources, can be considered a subtopic of game theory and can be analyzed using a similar approach as the broader topics of game theory. Alfred Chandler^[2] defines strategy as the determination of basic long-term goals and objectives, and the adoption of courses of action to achieve these goals.^[2] Strategy, which includes negotiation,

is important to the organization’s success, to the achieving of its long-term goals.

In a general way, a model of negotiation is an attempt to model in a specific way the interactions of competing utility or wealth maximizers, as it takes into account the strategies of competing players. Negotiation is therefore concerned with the analysis of strategic interaction in which the decision maker is assumed to interact with others in the environment, this causing the optimal decision to be affected by the action of others.^[3] Although it uses the same players found in any market (buyers and sellers), it extends the analysis through which price is determined by including actual interaction such as asymmetric information and haggling over the price. As in any game of strategy, there may be *cooperation* in which the players use contracts and are able to plan long-term strategies. In other situations there could be *non-cooperation* especially if enforcing the contract is difficult.

A model of negotiation could be included under the broader topic of “game theory” in which payoff functions and strategy sets are assigned to the participants, and the various outcomes are noted when particular strategies are chosen to maximize the payoff. “Game theory is concerned with the actions of individuals who are conscious that their actions affect each other.”^[4] Game theory is used mostly in situations where private decision makers are seeking to maximize wealth in the marketplace, not in situations where purchases are made by government agencies as they are not assumed to be “maximizers” of wealth or profit. However, because purchases by government agencies include contracts, are budget restricted, take place under various forms of competition, and oftentimes include some haggling, there is scope for analyzing government procurement through some form of game theoretic model using similar assumptions. The analysis presented below will not be a “pure form of game theory.” It will seek to specify conditions under which government procurement takes place and theorize as to where the final price will settle; closer to the seller’s offering or closer to the buyer’s suggestion.

VERTICAL INTEGRATION VS. OUTSOURCING

Vertical Integration

Public and private firms, seeking to acquire goods, generally do so through vertical integration or outsourcing. Whereas a vertical chain of production is the coming together of firms at various stages in the production of a good, outsourcing is that situation in which a firm gets an input to deliver a good or service, or the firm acquires the good or service to be delivered from an external source. With vertical integration, benefits such as: 1) reduction in transaction and coordination costs; 2) continuity of supply; 3) the non-sharing of proprietary rights; and 4) greater control over the quality of inputs are often realized.

Organizing production through vertical integration means that firms are organized into a business unit. The implication from this is that measuring the performance of individual firms requires that a “transfer price” be established for goods and services exchanged. With a transfer price, total profits can be reallocated among firms in the business unit and this could impact the business unit’s overall profit. Supplying goods through vertical integration may therefore not encourage least cost production by an individual firm because of subsidies by more profitable firms within the integrated business unit.

Transfer of goods between firms in a vertically integrated business unit can take place as follows. If there is a competitive external market for the good in question, the product can be transferred at the “external market price.” If there is no external market or if for some reason the market price does not truly measure the opportunity cost of producing the good, the “marginal production cost” could be used to determine the transfer price. With the marginal production cost being the cost to produce the last unit, this therefore represents the value of resources foregone to produce the last unit. Some firms in an integrated unit have also made use of “full-cost transfer prices” because it is felt that marginal cost of production focuses on variable cost and omits fixed cost.^a This method is simple, is easily implemented, and is the most popular of the pricing mechanism used by firms in an integrated unit.^[3]

Price negotiation can assist in the transfer of goods between firms in an integrated unit and also in the acquisition of goods through outsourcing. In an integrated unit, the price at which goods are transferred is aptly labeled the “negotiated transfer price” because it is determined by negotiation between the units.

A negotiated price between firms in a business unit is expected to maximize the combined profits of the negotiating firms. The selling firm will not negotiate a price below its production cost, and the acquiring firm will not pay a price above that for which it can buy the product elsewhere. The reference to purchasing a good at a price not higher than that for which it can be purchased elsewhere indicates that the market does play an indirect role, and serves more as a reference point for the determination of a negotiated transfer price. Because it is possible for two firms in an integrated unit to negotiate a transfer price without at the same time agreeing on the quantity to be transferred at that price, there is no guarantee that the negotiated price will maximize the business unit’s value. There is also the possibility of a long, drawn-out, and time-consuming process which when converted to a monetary value could increase the cost of acquiring goods and services.

Outsourcing

For the public sector in the United States, goods and services are acquired mostly through outsourcing, which is generally defined as obtaining goods and services from outside rather than providing them in-house. There are many possible reasons why the public sector may have decided to acquire goods and services through outsourcing. Among them are: 1) heightened competition between supply firms and the relatively low cost of goods and services; 2) flexible production techniques and the willingness of producers to satisfy government needs; 3) the short tenure of government and the disruption that would be caused when the leadership/ruling party changes; and 4) improved communications and the relative ease with which goods and services can be obtained from outside agencies.

When firms seek to acquire goods and services through outsourcing, the cost of goods and services are determined by market conditions, or the price is determined through negotiation, especially where there are long-term contracts.^b Buying goods in the competitive market could be advantageous as compared to a non-competitive or negotiated situation because it could easily be argued that as competitive firms do not make surplus profit over the long-run period, the market-determined price tends to be lower than a negotiated price. However, even with the potential benefits from acquiring goods at market price, procurement officers in the public sector have used negotiation with long-term contracts and a few vendors to

^aWhereas variable costs (e.g., direct labor costs and commissions to salespeople) increase with output, fixed costs (e.g., lease agreement and administrative expenses) remain constant when output increases.

^bLong-term contracts could be in the form of long-term supply and distribution contracts, franchise contracts, leasing contracts, or strategic alliances.

acquire goods and services. There seems to be the feeling that through negotiation there is more control over price, quality, and delivery. Although this may be true for quality and delivery, the negotiated price is more dependent on the skillfulness of the parties “at the negotiating table” and conditions in the market.

GOVERNMENT PROCUREMENT AND MARKET CONDITIONS

Negotiation in general differs from a ball game or a war where only one side wins and the other side loses. In successful negotiations, both sides win something, giving rise to a “win-win” situation. When procurement officers or purchasing agents seek to acquire goods and services through negotiation, the objectives of negotiation are: 1) to obtain the quality specified; 2) to obtain a fair and reasonable price; and 3) to get the vendor to perform the contract on time. Although all three objectives are important, budgetary restrictions on public sector agencies cause most attention to be paid to obtaining goods and services at a fair and reasonable price. Mention is often made of securing goods and services at the “right price,” that which is fair to both buyer and seller.^[5] And even with this definition, the “right price” is not static because firms are able (within limits) to adjust their asking and offering price, which could vary with market conditions.

In the United States we find the three forms of competition that are discussed in any elementary microeconomics textbook. At one extreme, there is the idealist form of perfect competition characterized by “atomistic” competition in which a large number of sellers trade a homogeneous good. This form of competition is also characterized by the availability of low cost of accurate information and the ability of firms to freely enter and leave the industry.

At the other extreme of the competitive spectrum is monopoly, where one firm controls the supply and hence the price of the product. Some reasons for the establishment of monopolies include the control of specific assets, production requiring large output and the realization of economies of scale, the availability of excess capacity and the ability to increase production at will, pre-commitment contracts, licenses and patents, and pioneering brand advantages.

Between the two extremes are conditions of imperfect competition where the number of sellers of a heterogeneous or homogenous good can be large or small. Under this market form, the supplier has some control over “brand” price. Studies have shown that in the United States, most goods are traded under conditions where there is some freedom to adjust price,

and this would imply conditions of imperfect competition.^c We accept the conclusions of these studies as true, but we also believe that tightly budgeted expenditures and the encumbering of funds for future expenditure cause the government sector to secure goods under varying conditions of competition. To get the biggest “bang for the buck,” government procurement is forced into markets where the price will be “right” or most beneficial to the agency.

Pricing of Goods

One of the tips given for conducting successful negotiation is “do your homework.” For government agents, this includes knowledge of the product and market. Regarding the procurement of goods, one would be more inclined to believe that if quality and quantity are easily ascertained, pricing issues involving government procurement would be at a minimum. If this is so, the situation boils down to whether pricing should be based on full cost, marginal cost, or some method to benefit the government agency as well as the firm supplying the good.

In the most competitive market, substitute goods are differentiated by design, wrapping, or other such features. The market has a large number of sellers with the individual seller forced to sell at “near equal” prices. Furthermore, to remain in this market, suppliers must be very efficient. We could therefore infer from this that suppliers in this market will sell to government agencies at the lowest possible price, that the goods will be of the highest value, and that price is a true indicator of quality. Furthermore, if government procurement involves “large dollars” and contracts that can cover multiple years, firms selling to government agencies will endeavor to have a long relationship with the agencies by selling at a lower price. Based on the above, government procurement agents seem to have some amount of buying power.

Intense competition forces manufacturers to make their products intrinsically different. This gives room for different negotiated prices between suppliers and the government, and components such as service and delivery are included in the price. Also, as the number of producers/sellers in the market declines, suppliers will have more power over the price at which goods are purchased. It also follows that as the products become more differentiated, more effort will be required by government agencies to determine quality and similarity of prices. If procurement personnel is limited, government agencies could be forced to accept

^cSee Ref.^[5], p. 244, for a list of these studies.

the seller's words, with the negotiated price more beneficial to the seller.

If the government seeks to buy goods from traditional sellers, procurement agents could find themselves involved in a game. This is because traditional sellers see negotiation as a game in which they offer to sell their products at a very high price, expecting the buyer to counter at a very low price. Through haggling and counter-offers, the established price is set somewhere in between the seller's high price and the buyer's lower price. If the government procurement official is expecting this response, both buyer and seller will be using various means (tricks, creative lies, and artful badgering) to negotiate in their favor. This could compromise delivery, quality, and goodwill.

Pricing of Services

Research indicates that employee's compensation as a percentage of non-capital direct expenditure is between 30% and 40% at the state and local levels, and 15% and 20% at the federal level.^[6] Economic theory proposes that labor should be paid according to its marginal revenue product, which is the marginal product of labor expressed in dollar value. This approach is only useful in the public sector where the output of labor is easily determined and where the government can determine the quality of the output. If quality and quantity are not easily determined, there is room for a negotiated wage rate. Many factors can determine the negotiated wages, these including union representation, skills of labor, demand by the public sector, and wage rate in other sectors of the economy.

There is also the additional issue of what price should the government pay labor when productivity of labor and wages in the other sectors of the economy increase faster than in the public sector. This issue is important because depending on the policy chosen, the supply of labor in the public sector could decline, efficiency could fall, and the average cost of services in the public sector could increase. Here again, negotiations are important. To reduce the above problems, the negotiated price of labor should be close to that which is offered in the more efficient private sector. Furthermore, because jobs in the public sector tend to be more secure and with the likelihood of more generous benefits, paying labor a rate close to that paid in the more efficient private sector could attract labor from the

private sector and improve the efficiency of labor in the public sector.

Even with the analysis outlined above, we understand that each procurement project is unique and complex and thus defies the use of a general rule or policy. We also believe that for each purchasing organization, the regulations, and the rules are different. These complicate the procurement process. In the end, the procurement approach that is used and the manner in which it is implemented will determine the success or failure of government's projects. Because of the dynamic nature of today's market, it is imperative that government agencies continue their vigilance on procurement procedures.

CONCLUSION

With the increasing size of government spending, and with more pressure on the public sector to provide a wider range of goods and services, negotiating the "best" price for the highest quality of goods and services is of greatest importance. To facilitate this process, government procurement officials must be well trained in negotiation, business decision-making, and economics and their departments must be adequately equipped with the latest technology to seek out suppliers. In the long run, a more informed procurement division will go far in maximizing society's benefits from public expenditure.

REFERENCES

1. Cavinato, J.; Kauffman, R. *The Purchasing Handbook: A Guide for the Purchasing and Supply Professional*; McGraw-Hill: 2000; 449, 500.
2. Chandler, A. *Strategy and Structure, Chapters in the History of the American Industrial Enterprises*; MIT Press: Cambridge, MA, 1962; 13 pp.
3. Brickley, J.; Smith, C.; Zimmerman, J. *Managerial Economics and Organizational Architecture*; McGraw-Hill: 2001; 213, 448.
4. Rasmusen, E. *Games and Information*; Blackwell, 1990; 21 pp.
5. Dobler, D.; Burt, D.; Lee, L., Jr. *Purchasing and Materials Management: Text and Cases*, 5th Ed.; McGraw-Hill: 1983; 242 pp.
6. Hyman, D. *Public Finance*; Dryden, 1993.

Activity-Based Costing in the Public Sector

Thomas Buttross

*Department of Accounting, Pennsylvania State University at Harrisburg,
Middletown, Pennsylvania, U.S.A.*

George Schmelzle

*Department of Accounting and Finance, School of Business and Management Science,
Indiana University Purdue University, Fort Wayne, Indiana, U.S.A.*

INTRODUCTION

This article is meant to provide the reader with a general understanding of activity-based costing (ABC), and its uses and benefits in the public sector. Some of the limitations of ABC are also covered.

ABC is designed to provide information for strategic decisions. Applying this model to the public sector, ABC can provide information on the cost of providing government services for strategic decisions, such as determining the affordability of providing discretionary government services (e.g., picking up yard waste); setting user fees, both internally between agencies and externally, for billable government services (e.g., water and waste water services); and determining whether to outsource government services (e.g., trash pickup).

Activity-based management (ABM) takes the ABC model to the next logical step, providing information for operational decisions (i.e., for cost management aimed at improving day-to-day operations). For example, instead of outsourcing trash pickup, the governmental entity might use the activity-based information to improve the government operation to the point where it is competitive with the private sector. The combined use of ABC and ABM is so common that many writers refer to the combined models as one—ABCM (activity-based cost management).

Activity-based budgeting (ABB) takes the model to its third logical step. Once the organization knows the cost of providing various services, it can budget the level of services it wants to provide and determine the amount of resources required to provide that level of services. Given scarce resources, the level of services is likely to exceed the available resources, even after additional sources of funds have been considered. Reductions of funds can then be logically tied to reductions of particular services. For example, if recycling pickup costs \$2 per house per pickup, the cost savings of reducing pickups from once per week to once every two weeks can be determined. Viewed from the opposite angle, the effect of cutting resources available

for recycling pickup in half on the frequency of pickups can be more clearly determined.

ADVANTAGES OF ABC

Activity-based costing has several advantages:

1. It provides understandable information—people intuitively understand activities, and how activities consume resources and raise the cost of cost objects.
2. It reduces arbitrary cost allocations by charging resources to the activities that consume them and charging activities to the cost objects that consume them, where possible.
3. It identifies activities that highlight departmental interdependencies, which can reduce functional suboptimization.
4. It provides cost driver information for managers and employees to remove costs on existing services or avoid costs on new services.
5. It provides data for other improvement initiatives (e.g., total quality management).
6. It improves budgeting by tying the budget to the actual work that is done.

CAVEATS REGARDING ABC

Activity-based costing is not a panacea; ABC models tend to address cost only, at a strategic level too general to satisfy operational needs. Operational improvement information comes from the addition of ABM and requires the addition of non-cost performance measures to the model.

Even ABCM does not lead to improvements. People must use the data to make improvements. If the ABC system indicates that governmental trash pickup is more costly than the private sector, it will continue to be more costly unless the governmental garbage

entity uses ABM information to reduce the cost of providing that service.

Activity-based costing, ABM, and ABB are not one-time efforts, which is another way of saying that they are costly in terms of time and money. If ABC indicates that the cost of providing a service is too high and that service is reengineered to be less costly, then a new ABC model needs to be constructed to determine the new cost of providing that service.

Finally, there are problems caused by generally accepted accounting principles (GAAP). Under GAAP, a cost incurred in 1 year, but related to several years, may be expensed in the year incurred. For example, a major training program for newly installed software may be expensed when incurred, making it less likely to be reflected in the cost of future activities that benefited from the training. Conversely, a cost delayed, such as maintenance, will not be reflected in the accounting records until incurred, making it unlikely to be reflected in the current activities that benefited from the use of the asset.

MODEL AND TERMINOLOGY

The CAM-I Cross

To understand ABC systems, it is instructive to examine the CAM-I Cross. The vertical portion of the cross, the cost assignment view, is ABC, whereas the horizontal portion, the process view, is ABM (Fig. 1).

As the cost assignment view shows, resources are consumed by activities and activities are consumed by cost objects. From another perspective, cost objects require the performance of certain activities, and those activities require the use of certain resources. Each part is discussed in turn.

Activities (Tasks and Processes)

Activities are at the center of ABC. They are a series of tasks (work steps) performed by an organization yielding a useful aggregation. To be useful, the aggregation needs to be neither too broad nor too narrow. For example, “provide unemployment benefits” would probably be too broad. Such aggregations of activities are called processes. “Hand blank benefits form to next person in line” would be too narrow. Such narrow parts of activities are called tasks or work steps. “Process unemployment-benefits application,” “authorize benefits payment,” and “issue benefits check” might be proper aggregations.

Because activities represent work done by the organization, they should be defined by an action verb followed by a noun, with an adjective in the middle optional. Where only one type of application is processed (or where every application processed requires the same level of resources), “process application” is an understandable activity name. Where different applications are processed, more descriptive names such as “process unemployment-benefits application” are warranted.

There are various ways to identify activities (e.g., storyboarding and activity dictionaries) but their discussion is beyond the scope of this article.

Resources

Resources are economic elements used by an organization (e.g., labor, equipment, utilities). Although the model refers to resources, the cost of resources is what appears in the model. So, while labor is a resource consumed by activities, the salaries, which represent the cost of the labor, are reflected in the ABC computations.

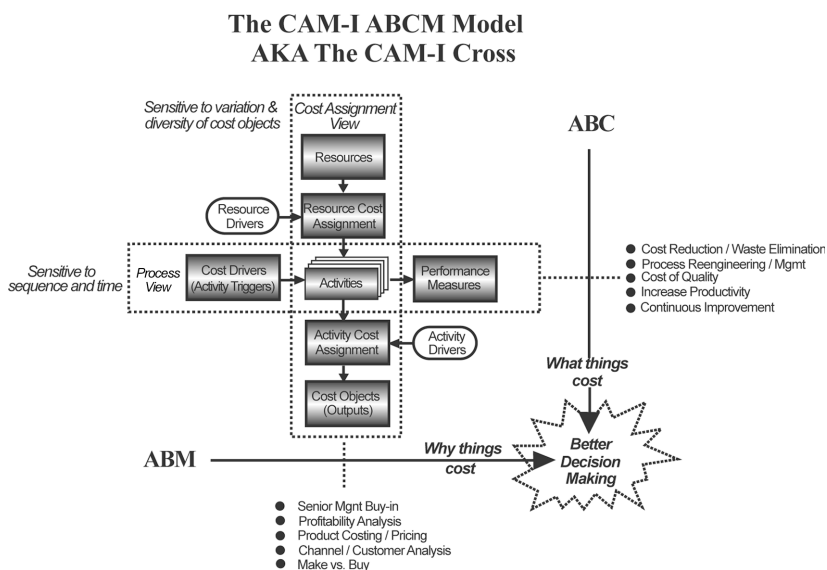


Fig. 1 The CAM-I Cross. (© 1990 CAM-I. Used with permission.)

To process unemployment benefits applications, the entity will use resources such as labor, equipment, utilities, and office supplies. The cost of these resources can be found in the accounting records and needs to be charged to the activity that consumed the resources.

Cost Objects

Cost objects are the reason that organizations perform activities. They are simply objects for which management desires a cost. Examples include products, customers (individual recipients of benefits or selected groups of benefit recipients, such as laid-off workers from a particular employer), and services.

Cost objects cause activities (e.g., provision of unemployment benefits requires the provider to process unemployment-benefits applications, authorize benefits payment, and issue benefits checks). The activities caused by a particular cost object should be charged to that cost object.

Resource Drivers

Resource drivers are used to charge the cost of resources to activities. In some cases, a resource is used for only one activity, and therefore, can be traced to the activity. If employees spend 100% of their time processing unemployment-benefits applications, then the employees' salaries can be charged in total to that activity. However, cost assignment (allocation) becomes necessary when employees spend time on multiple activities.

Resource drivers (first-stage drivers) are used to assign resource costs to activities. They represent a measure of the quantity of resources consumed by an activity. As examples, 1) labor and fringe benefits of an employee can be assigned based on the percentage of time the employee spends "processing unemployment-benefits application" and "authorizing benefits payment," and 2) equipment rent or depreciation can be assigned based on the percentage of time the equipment is used for different activities.

There are various ways to identify resource drivers (e.g., labor logs, questionnaires, interviews of employees and supervisors, observation, physical measurement for items like square footage) but their discussion is beyond the scope of this article.

Activity Drivers

Activity drivers (second-stage drivers) are factors that are used to assign activity costs to cost objects. They represent a measure of the quantity of activity cost consumed by the cost object. There should be a

cause-and-effect relationship with the cost driver. For example, assume that we are processing unemployment benefits for two companies, Company X and Company Y, and we want to assign the cost of this processing to each company. In this case, the activity driver might be "number of claims processed." If we process twice as many claims for Company X as Company Y, then twice as much cost for this activity should be charged to Company X.

Activity drivers exist at various levels: unit, batch, product or service, customer, facility, and organization sustaining are common categories.

Cost Drivers

Cost drivers are the root causes of costs. Identifying cost drivers can help an organization to improve processes or reduce costs by eliminating or reducing the non-value-added activities and by performing value-added activities more efficiently.

Again, consider the processing of unemployment claims. Assume that the agency has a policy that requires a supervisor to review each unemployment-benefits application that is processed, before payments can be made. The policy requiring the supervisor to look at each processed claim would be a cost driver. The related work step or task—the supervisor actually reviewing each claim—would be included in the activity "authorize benefits payment." By eliminating this policy, the cost of performing this activity would be reduced. Another cost driver of both "process unemployment-benefits application" and "authorize benefits payment" is the design of the unemployment-benefits application. A poorly designed document will increase the time required to input/review the information and increase the likelihood of errors, thereby increasing the processing and authorization costs. Redesigning a problematic claim form can reduce the cost of these activities.

There are multiple cost drivers associated with most activities. The cost of the activity "process unemployment-benefits application" is driven by many factors, including the number of claims processed, the completeness of the information supplied by the claimant, the training and experience of the agency employees processing the documents, the design of the form, and the time of day when processing occurs. The identification of significant cost drivers is critical to ABM.

As a final note, cost driver is not a synonym for activity driver. There is only one activity driver for a particular activity, but there are multiple cost drivers. In the case of "process unemployment-benefits application," one of the cost drivers—the number of claims processed—is likely to be used as the activity driver.

Table 1 Estimated overhead costs using record-keeping categories

Overhead only	Estimated cost
Salaries	\$150,000
Fringe benefits	30,000
Building rent	70,000
Building utilities	10,000
Small tools and supplies	15,000
Equipment depreciation	160,000
Equipment insurance	3,000
Equipment maintenance	17,000
Gasoline	45,000
Total	\$500,000

There are even situations where the chosen activity driver is not one of the cost drivers. If one remembers that cost drivers are what causes cost to appear in the activity cost pool, whereas activity drivers are a measure of the consumption of the activity cost pool by cost objects, the difference between the two types of drivers is understandable.

Performance Measures

Performance measures determine how effective an organization is at performing activities. These measures describe how effective the organization is at performing the work (the activities) in the organization, along with an indication of whether the activities that are being performed meet the needs of the organization's internal and external customers. Performance measures can be either financial or non-financial in nature. There are four categories of performance measures: time, quality, cost, and productivity.

An organization that processes unemployment claims might be interested in the "number of claims processed per hour" (a performance measure that is measuring time), the "number of claims that were error free" (a performance measure that is measuring quality), "the cost per claim filed" (a performance measure that is measuring cost), or "total number of claims filed

Table 2 Traditional cost allocation

Traditional cost allocation	Picking up leaves (4 DLH)	Cutting trees (5 DLH)
Direct labor (@ \$7 per DLH)	\$28	\$35
Service overhead (@ \$227 per DLH)	908	1,135
Totals per pickup/cutting	\$936	\$1,170

Table 3 ABC overhead activities

ABC overhead activities	Estimated costs
Schedule pickups/cuts	\$40,000
Set up equipment (for pickups/cuts)	100,000
Drive routes (and pickups/cuts)	270,000
Dispose of leaves and branches	50,000
Total	\$460,000

during the week" (a performance measure that is measuring productivity).

Also note that performance measures for one group can be cost drivers for another. Take the use of "number of claims that were error free" as a performance measure for the activity "process unemployment-benefits application." Errors in processing applications increase the cost of the subsequent activity—"authorize benefits payment"—because such errors require correction before authorization can occur. They can also decrease the performance of the "authorize benefits payment" activity because they decrease the number of authorizations that can be completed in a given amount of time. In other words, errors in "process unemployment-benefits application" are a cost driver for the downstream activities.

EXAMPLE

The following example helps to clarify the discussion.

The City of Kendallville performs two lawn services for its taxpayers. It picks up leaves after residents rake them into the street, and it cuts the branches of trees that are close to street signs and power lines. Except for a few smaller items, the same equipment is used for both services. Picking up leaves is done three times per year, whereas cutting trees is done twice per year.

Kendallville is divided into 100 neighborhoods. Thus, picking up leaves is done 300 times and cutting

Table 4 Activity data

ABC overhead activities	Picking up leaves	Cutting trees	Total
Schedule pickups/cuts ^a	30	100	130
Set up equipment (for pickups/cuts) ^b	5	80	85
Drive routes (and pickups/cuts) ^c	35,000	30,000	65,000
Dispose of leaves and branches ^d	50,000	100,000	150,000

^aNumber of trips scheduled.

^bNumber of setups.

^cNumber of miles driven.

^dPounds dumped.

Table 5 Overhead rates per unit of activity

Activity	Estimated overhead cost/ Estimated total activity	Overhead per activity
Schedule pickups/cuts	\$40,000/130 trips	\$307.69 per trip
Set up equipment (for pickups/cuts)	\$100,000/85 setups	\$1,176.47 pre setup
Drive routes (and pickups/cuts)	\$270,000/65,000 miles	\$4.15 per mile
Dispose of leaves and branches	\$50,000/150,000 pounds	\$0.33 per pound

trees is done 200 times. For each neighborhood, it takes 4 hours to pick up the leaves of the residents, and 5 hours to cut the branches. The city currently has a traditional cost system in which direct labor hours are used to assign overhead cost to these two services.

Table 1 shows the estimated overhead using record-keeping categories.

The predetermined overhead rate of \$227 per direct labor hour is determined by dividing the budgeted overhead cost of \$500,000 by the budgeted 2200 direct labor hours. (The budgeted direct labor hours represent 300 occurrences of picking up leaves times 4 hours per neighborhood plus 200 occurrences of cutting trees times 5 hours per neighborhood.)

Table 2 shows the cost of each service for one neighborhood.

The City of Kendallville is considering the implementation of an ABC system to cost these two services. The city manager has collected the following information.

First, the city determined that, of the \$500,000 in service overhead, \$10,000 of the small tools and supplies represents saw blades and such, which can be directly traced to cutting trees. In addition, \$30,000 of the salaries represents payments to an external consultant who reviews the neighborhoods and determines where tree cutting will be necessary. This \$40,000 will

be directly traced to the tree-cutting cost object (\$40,000/200 cuttings = \$200 per cutting), whereas the remaining \$560,000 will be assigned (allocated) using ABC.

Table 3 presents the activities were determined to be the cause of the remaining \$460,000 in overhead costs. Table 4 presents estimated activity data.

The following activity data was estimated (Table 4).

Because picking up leaves is determined by the seasons, scheduling of 10 neighborhoods at one time is possible. So, 300 pickups need only 30 trips scheduled. Also, because the pickups occur close together in time, few equipment changeovers are required between pickups. Tree cutting, however, is not restricted by the seasons and is done in small batches of neighborhoods, as need is identified by an external consultant.

Table 5 displays the overhead rates computed for each specific activity. With these overhead rates, Table 6 presents the assignment of overhead to services. The assigned overhead is then used to compute the service cost using ABC (Table 7).

In this example, the ABC rates of \$618 per pickup and \$1647 per cutting indicate that the traditional system rates of \$936 and \$1170, respectively, are overcosting the leaf removal service and undercosting the tree-cutting service. This information would be useful if, for example, the local utility reimburses the city for much of the tree-cutting service.

Table 6 Assigning overhead to services

Activity	Picking up leaves		Cutting trees	
Schedule pickups/cuts	30 trips × \$307.69 per trip	\$9,231	100 trips × \$307.69 per trip	\$30,769
Set up equipment (for pickups/cuts)	5 setups × \$1,176.47 per setup	5,882	80 setups × \$1,176.47 per setup	94,118
Drive routes (and pickups/cuts)	35,000 miles driven × \$4.15 per mile	145,250	30,000 miles driven × \$4.15 per mile	124,500
Dispose of leaves and branches	50,000 pounds disposed of × \$0.33 per pound	16,500	100,000 pounds disposed of × \$0.33 per pound	33,000
Totals (assigned \$459,250, instead of \$460,000 due to rounding)		\$176,863		\$282,387
Service overhead per pickup/cutting	\$176,863/300 pickups	\$590	\$282,387/200 cuttings	\$1,412

Table 7 ABC cost allocation

ABC cost allocation	Picking up leaves (4 DLH)	Cutting trees (5 DLH)
Direct labor (same as traditional above)	\$28	\$35
Directly traced service overhead (computed earlier)		200
Service overhead per pickup/cutting	590	1,412
Totals	\$618	\$1,647

CONCLUSION

With activity-based costing information, a public entity can make strategic decisions regarding what discretionary government services to offer, what user fees to charge internally as well as to consumers for billable government services, and whether to in source or out source public services. With the addition of activity-based management, the governmental entity can make decisions involving the management of its costs that will improve the efficiency of its operations. Finally, with the addition of activity-based budgeting, the public entity can determine which services it can afford to provide to members of the community based on available resources, and how much it can save from curtailing various services in times of scarcity.

BIBLIOGRAPHY

1. Brimson, J.A.; Antos, J. *Activity-Based Management for Service Industries, Government Entities, and Nonprofit Organizations*; John Wiley & Sons, Inc.: New York, 1994.
2. Brian, F. Marked improvements. *Gov. Exec.* **2000**, 32 (3), 86–88.
3. Carter, T.L.; Sedaghat, A.M.; Williams, T.D. How ABC changed the post office. *Manage. Account.* **1998**, 79 (8), 28–32.
4. Harvey, M. Indianapolis speeds away. *J. Bus. Strategy* **1998**, 19 (3), 41–46.
5. Kline, J.J. ABC, SPC, SEA: Alphabet soup or quality synergy. *Gov. Account. J.* **2000**, 49 (1), 40–45.
6. LaPlante, A.; Alter, A.E. U. S. Department of defense: activity-based costing. *Computerworld* **1994**, 28 (44), 84.
7. May, M. An activity-based approach to resource accounting and budgeting in government. *Manage. Acc. (Lond.)* **1995**, 73 (7), 27.
8. Simpson, W.K. Activity-based: costing, management and budgeting. *Gov. Account. J.* **1996**, 45 (1), 26–28.
9. Turney, P.B.B. *Common Cents: The ABC Performance Breakthrough*; Cost Technology: Beaverton, OR, 1993.
10. Williams, C.; Ward, M. Is ABCM destined for success or failure in the federal government? *Public Budg. Finance* **1999**, 19 (2), 22–36.
11. Williams, C.; Ward, M. The ABC's of government reinvention: can activity-based costing reinstall public trust in government? *Penn. Times* **1997**, 20 (12), 1.

Administration of State Lotteries

Jeff Leon Dense

*Department of Political Science, Eastern Oregon University,
La Grande, Oregon, U.S.A.*

INTRODUCTION

The most common form of legalized gambling in the U.S.A. are state lotteries. Forty states and the District of Columbia currently operate lotteries. From their humble beginnings in New Hampshire in 1964, lotteries have developed into a primary revenue source for a number of state governments, generating over \$48 billion in sales during the fiscal year 2004. More importantly, after accounting for prize distribution and administrative expenses, lotteries generated over \$15 billion in profits.^[1] State lotteries have assumed a prominent, albeit often misunderstood, role within public finance and American culture.

A series of structural developments contributed to the states jumping on the lottery bandwagon. Citizen resistance to additional taxation, the Reagan era devolution of government known as "New Federalism," and the existence of lottery operations in contiguous states, served as rationales underlying lottery adoption. Driven by the imperative of revenue maximization, a range of administrative practices, including product portfolio strategies, retailer commissions, operational efficiency, and joint ventures with other lotteries have a direct effect on a lottery's "bottom line." The proliferation of Native American and commercial casinos has supposedly created another weighty fiscal challenge. An appraisal of administrative practices and their subsequent impact on public finance, underscores the problems and prospects confronting state lotteries.

LOTTERY PRODUCT PORTFOLIO

One of the fundamental marketing decisions confronting lottery officials is which products should be offered to the public. A series of structural characteristics impinge on lottery performance and the amount of revenue the lottery is able to generate. While the payout rate for most lottery games hovers around 50%, lottery officials can manipulate several factors, including the frequency of the opportunity to gamble and the range of stakes associated with a particular game, in an attempt to maximize sales and ultimately revenues. Each lottery product has its own unique set of structural characteristics that, in tandem with other lottery

administrative practices, have the potential to provide a significant stream of revenue to state coffers.

Instant scratch-off tickets are a staple of all lottery product portfolios. Given their low technological requirements, instant tickets are especially attractive to newly created lotteries. Thematic games (e.g., Happy Holidays and Pan for Gold) and their simulated play value keep the player in suspense as long as possible, adding to their allure. It is common for lotteries to simultaneously operate a number of instant scratch-off ticket games. Instant tickets account for approximately 50% of traditional lottery sales.

Daily numbers games are a remnant of the illegal numbers games that were a longtime fixture of many urban neighborhoods in the U.S.A. The vast majority of state lotteries offer three and four digit numbers games on a daily basis. A popular variant of the daily numbers game is Lotto, which requires a player to match six numbers from a field ranging from 49 to 54. If no player correctly selects all numbers, the grand prize is rolled over to the next drawing. This rollover feature can lead to multimillion-dollar jackpots and a spate of free publicity for the lottery. A third variant of the daily numbers game is Keno, based on the casino game of the same name. Keno allows the player to choose a number of numbers from a range of 80. Keno is currently offered by 11 states.

One of the most important game portfolio selection decisions is whether to adopt Video Lottery Terminals (VLTs). Similar to slot machines, VLTs are far more similar to casino gaming products than those traditionally offered by state lotteries. The ownership of VLTs differs by jurisdiction, ranging from private sector and technology provider, to government ownership. Despite the relatively low odds of winning, the opportunity to immediately reinvest the winnings and the fast-paced action has permitted VLTs to assume a primary role in lottery revenue generation.

STRUCTURAL CHARACTERISTICS

Of all the structural characteristics of lottery products, the most important from the revenue maximization perspective is the frequency of opportunity to gamble. A skilled VLT player can play approximately 15 games

a minute in comparison with fast-draw Keno (five minutes) and bi-weekly Lotto drawings. However, state lottery officials are able to maximize revenues from more static game offerings like Instant Scratch-Off Games by altering the cost of participating in these games. While all states offer \$1 Instant tickets, there has been a recent trend toward introducing higher priced Instant tickets. Nearly all lotteries currently offer a \$10 Instant game, and eight states have begun offering \$20 tickets. While the jackpot for these higher priced Instant games is higher than their lower priced brethren, it should be noted that none of the states offering \$20 dollar Instant tickets has decided to legalize VLTs.

While Lotto and Daily Numbers games have lingered with their traditional \$1 cost, it is conceivable that given the increased competition for the gambling dollar, the cost of participating in traditional lottery games will rise in the future. There is a multiplicity of price-point strategies for lottery administrators to contemplate as they manipulate their product portfolio.

RETAILER COMMISSIONS

States employ a range of compensation schemes to reward retailers, vendors, and game operators, which have an intrinsic effect on the lottery's "bottom line." In an effort to maximize profitability, lottery officials must develop an easily accessible network of retailers to provide their products to the public. While virtually any business in the U.S.A. is eligible to sell lottery products, convenience stores and grocery stores account

for nearly three-quarters of all lottery sales. However, this retail service is not provided free of charge. The typical sales commission paid ranges from 5% to 7%. Moreover, the majority of state lotteries provide a cashing bonus to sellers who redeem winning tickets, usually 1%, providing an additional incentive for retail outlets to sell lottery products.

While there is a relatively standard formula for compensating retailers of traditional lottery products, there is wide variability in the structure of compensation for VLT operators (Table 1).

With the exception of Oregon, VLTs are operated by third parties, who are able to extract a high percentage of net sales for their services, in excess of 25% in each VLT-operating state.^[2] Increased sales should logically lead to increased profits. However, the inefficient nature of lottery operations, especially VLTs and their exorbitant commission structure, should give public officials pause as they attempt to maximize profits while providing a modicum of administrative oversight.

OPERATIONAL EFFICIENCY

While prizes paid to players provide the impetus for the public to participate, the high cost of retailer commissions and administrative expenses impede the lottery's ability to maximize revenue to government coffers and good causes like senior citizen programs, parks and recreation, and education. There are a variety of measures that can be utilized to assess the operational efficiency of state lotteries, including cents spent to generate a dollar of revenue, percentage of

Table 1 Video lottery terminal operator compensation rates

Jurisdiction	Retailer compensation percentage	Ownership and operation
Delaware	49%	Third party operators
Iowa	80%	Retailer
Louisiana	Bars and restaurants: 74% Racetracks/Off track betting: 77.5% Truck stops: 67.5%	Third party operator or retailer
Montana	85%	Third party operator or retailer
Oregon	Sliding scale, minimum of 20%	State lottery
New York	29%	Third party operator
Rhode Island	Racetrack: 27% Jai Alai: 26%	Third party operator
South Dakota	50%	Third party operator or retailer
West Virginia-Limited Video Lottery Terminal Program	57%	Third party operator or retailer
West Virginia-racetracks	43%	Retailer

(From Ref.^[2])

personal income spent on lottery products, and administrative costs as a percentage of net lottery profits (Table 2).

However, these statistical measures fail to accurately portray the cost of lottery operation to the general public. Instead, per capita administrative expenses provide a clear picture of the cost of lottery operation to the playing, and nonplaying, public. The evidence suggests that the high cost of administering traditional lottery products impedes the efficient operation of state lotteries and, ultimately, their ability to maximize revenues.

JOINT VENTURES

State lotteries have attempted to create economies of scale, and moreover, increase sales by offering games in concert with other lotteries. The most popular of these games are Powerball, played in 27 states, and Mega Millions, played in 12 states. No state offers both these games. Multimillion dollar jackpots allure the public to play these games, with jackpots in excess of \$100 million common. These games have aided lottery administrators in ensuring that a drawing for a lottery product is held daily, with drawings for

multistate games alternating with those offered by each state. The high odds of winning the grand prize (Powerball = 1:120,526,770 and Mega Millions = 1:175,711,536) mark these as the quintessential “bad bet.” However, the administration of these products is relatively efficient in comparison with other traditional lottery products, and quite profitable for those states that decide to participate in these multistate lottery games.

COMPETITION FOR THE GAMBLING DOLLAR

State lotteries once reigned supreme over the American gambling landscape. Lottery officials contend that the increased competition for the gambling dollar from Native American gaming, riverboat casinos, and Internet gambling will erode the market for state lotteries, leading to an adverse effect on public finance. Hence, public officials are never quick to embrace new forms of gambling activity that they perceive may minimize the fiscal benefits lotteries provide to state government.

One of the fundamental gaming policy issues confronting state governments is the potential effect of substitution and saturation on the state's lottery. Substitution refers to the diversion of spending on one type

Table 2 Per capita lottery operational expenses, 2004

State	Per capita expense (\$)	State	Per capita expense (\$)
District of Columbia	82.35	Indiana	16.04
Wisconsin	55.22	Texas	15.89
Massachusetts	50.07	Missouri	15.68
Rhode Island	35.25	New Mexico	14.60
Georgia	33.02	Idaho	14.58
New York	30.09	Iowa	13.75
South Carolina	26.29	Colorado	13.47
Maryland	25.70	Illinois	13.46
Connecticut	25.05	Kansas	12.43
Vermont	23.72	Minnesota	11.99
Michigan	23.04	Nebraska	11.82
Kentucky	22.47	Louisiana	10.88
New Jersey	22.31	Montana	10.74
Maine	21.90	Washington	10.70
New Hampshire	21.84	California	10.33
Delaware	21.19	Arizona	10.17
Ohio	20.11	South Dakota	9.45
Pennsylvania	19.93	Tennessee	8.97
Virginia	18.60	West Virginia	4.78
Florida	17.85	North Dakota	2.50
Oregon	16.62	Average	20.61

(From Ref.^[1].)

of gaming activity to another form of gaming. The related concept of saturation suggests that all consumer products, including gaming products, follow a product life cycle with four phases: introduction, growth, maturity, and decline. Saturation refers to a situation where sales of a lottery product, or wagering on a particular form of gaming, peak or flatten.

Evidence indicates that the introduction of commercial casinos has had a negligible impact on lottery sales and lottery-generated revenue.^[3] This finding has strident implications on the fiscal future of cash-strapped state governments throughout the U.S.A. who are contemplating an expansion of gaming activity. The reality is that there is far more substitution among lottery products than there is between lotteries and casinos. For example, the addition of a multistate lotto style game has a negative effect on sales of existing lotto games.

Given the “bottom line” orientation of state government officials, it is reasonable to infer that decision makers will argue that the proliferation of gambling is in the public interest when new forms of gambling activity provide an increase in revenues to state coffers. In every one of the nine states with commercial casinos and a state lottery, total gaming revenue (i.e., the sum of lottery and casino revenues accruing to the state) has increased since the introduction of casino gaming (Table 3).

While the decision to legalize casino gaming has long-term policy implications, public officials, lottery officials in particular, are often driven by short-term revenue considerations. A majority of states have realized double-digit average annual growth in total gaming revenues in the first five years of casinos operating in the state (Table 4).

Hence, the additional revenue provided to cash-strapped state coffers trumps the opposition of lottery officials to the legalization of additional forms of gambling activity.

Table 3 Average annual change in total gaming revenue

State	Average annual percentage increase (%)
South Dakota	56.1
Missouri	35.7
Louisiana	35.7
New Jersey	34.2
Indiana	24.1
Iowa	24.9
Connecticut	20.3
Colorado	19.9
Illinois	6.7
Michigan	2.3

First five years after casino legalization.
(From Ref.^[3])

Table 4 Average annual change in lottery-generated revenue: first five years after casino legalization

State	Average annual percentage increase (%)
South Dakota	51.1
New Jersey	25.2
Missouri	18.7
Colorado	10.6
Connecticut	3.7
Illinois	0.6
Louisiana	−0.4
Michigan	−1.4
Indiana	−1.5
Iowa	−1.5

(From Ref.^[3])

While this “bottom line” finding is significant for states suffering from budgetary stress, the majority of lottery states that have introduced casino gaming have been able to sustain a level of lottery profitability akin to that exhibited before casino legalization. The data reveal a pattern of flattening lottery-generated revenue, not a significant loss in lottery revenue, which is more than offset by the profitability of the newly legalized casino industry. This finding has significant implications for cash-strapped state officials seeking to maximize the “bang for the buck” provided by gambling activity. States operating lotteries that are considering the expansion of gambling activity should heed the lessons that can be gleaned from states that have already taken this same step in the gambling policy process.

The evidence reveals a relative absence of substitution between these competing forms of gambling activity. There are several explanations for the lack of a substitution effect between commercial casinos and state lotteries. First, it is reasonable to assume that commercial casinos and state lotteries are two separate markets differentiated by income level and consumer motive. Lottery players have lower incomes and tend to gamble to enhance their income. Conversely, casino players are “tourists,” i.e., they have higher disposable incomes and perceive casino gaming as a form of entertainment rather than an income-enhancing opportunity. Moreover, lotteries and commercial casinos are qualitatively different. From the perspective of ease of purchase, lotteries present a far more convenient gambling alternative.

CONCLUSIONS

While their dominance over the American gambling landscape has received a series of stiff challenges, state

lotteries will continue to play a significant role in public finance for the foreseeable future. There are several conceivable strategies lottery administrators can embrace to maximize future profits, including manipulating the structural characteristics of gambling (e.g., frequency to gamble) and introducing new products to their game portfolio (VLTs), privatizing operations, and increasing their operational efficiency. Given its weighty contribution to public finance, it is reasonable to assume gambling-generated revenue, and state lotteries in particular, will continue to play an important role within state government budgetary practice.

REFERENCES

1. McQueen, P. More than \$15 billion served. *Int. Gaming Wagering Business* **2005**, 26 (3), 29–32.
2. Oregon Audits Division. Oregon state lottery: video lottery compensation in the United States and Canada. **2004**; <http://www.sos.state.or.us/audits> (accessed May 2005).

3. Dense, J. State lotteries, commercial casinos, and public finance: an uneasy relationship? *Gaming Law Rev.* **1999**, 3 (5/6), 317–328.

BIBLIOGRAPHY

1. Borg, M.; Mason, P.; Shapiro, S. *The Economic Consequences of State Lotteries*; Quorum: Westport, Connecticut, 1991.
2. Clotfelder, C.; Cook, P. *Selling Hope*; Harvard University Press: Cambridge, Massachusetts, 1989.
3. Collins, P. *Gambling and the Public Interest*; Praeger: Westbury, Connecticut, 2003.
4. Dense, J. *Comparative Lottery Analysis: The Impact of Casinos on Lottery Revenues and Total Gaming Revenue*; Center for Policy Analysis: North Dartmouth, Massachusetts, 2004.
5. McGowan, R. *State Lotteries and Legalized Gambling*; Quorum: London, 1994.
6. Nibert, D. *Hitting the Lottery Jackpot*; Monthly Review Press: New York, 2000.
7. *Public Sector Gaming Study Commission*, Final Report; Florida Institute of Government: Tallahassee, 2000.

Administrative Discretion

Kenneth F. Warren

*Department of Political Science, Saint Louis University,
St. Louis, Missouri, U.S.A.*

INTRODUCTION

Administrative discretion refers to the degree of latitude or flexibility exercised by public administrators when making decisions or conducting any agency business. The chief source of administrative discretion comes from legislative bodies that have drafted vague laws. These skeletal statutes essentially allow public administrators the discretionary power to interpret laws as they see fit, as long as their discretionary interpretations do not contradict specific statutory provisions. Public administrators are normally not hesitant to exercise broad discretionary power because the reviewing courts tend to routinely defer to agency discretion (i.e., expertise), and also are specifically prohibited by statute from reviewing any agency action precluding judicial review or committed to agency discretion by law [Section 700 of the Administrative Procedure Act (APA) of 1946, as amended, United States Code: Title 5].

DISCRETIONARY BEHAVIOR IN THE CONTEXT OF FORMAL VS. INFORMAL AGENCY ACTIONS

Public administrators can follow relatively formal or informal procedures. In reality, agency actions are not wholly formal or informal, but a blend of the two. That is, formal action would consist of decision-making processes going solely by the book, whereas procedural steps are meticulously followed to reach decisions. However, because the APA or other laws or rules always permit administrators some administrative discretion when making decisions, even under the most structured or formal decision-making situations, total adherence to formal procedures would be both impossible and undesirable.

Actually, only about 10% of administrative behavior can be legitimately classified as formal.^[1] A formal agency action usually consists of those actions by administrators that fall well within prescribed (usually written) procedures, as dictated by statutes, court rulings, and/or agency rules and regulations. Formal administrative action, in contrast to informal administrative action, is characterized by the limited role an administrator's discretion can play in determining

what procedures can be followed and what decisions can be reached as the result of following specified procedures. One clear example of formal agency behavior can be seen when agencies conduct elaborate, trial-like formal evidentiary hearings where relatively strict rules of procedure are followed. Agency rule making can also be quite formal if administrators follow the elaborate provisions for making rules as required in Sections 556 and 557. Generally speaking, under the APA, informal action refers to any administrative action or inaction (e.g., a decision not to act) that does not involve a trial-like hearing. Formal rule making requires a trial-like hearing and is governed by Sections 556 and 557 of the APA, whereas informal rule making, which does not require a formal hearing, is governed by Section 553 or not by the APA at all. Likewise, adjudications that require a formal trial-like evidentiary hearing are considered formal, whereas adjudications that occur without a trial-like hearing are considered informal.

However, it would be futile to try to classify administrative actions into neat categories of either formal or informal because, in practice, administrators interchange formal and informal procedures. The point is that informal administrative actions are characterized by a broader use of administrative discretion than are formal administrative activities. As the formal-informal agency action continuum in [Fig. 1](#) illustrates, the more agency actions are controlled by administrative discretion, the more informal the agency action is. Conversely, the more formal the procedural codes (e.g., statutes, rules, court orders) that dictate agency actions, the more formal the agency action is. Of course, agency actions would never, in practice, be completely controlled by formal procedural codes or totally dominated by an administrator's discretion. The extreme ends in the continuum only represent the theoretical extremes. Although at times one may not perceive any evidence that an administrator's discretionary decisions are being modified by prevailing legal sanctions, in reality sane administrators are always cognizant of the need to submit somewhat to environmental pressures to conform to expected patterns of behavior. Systems theory, popular in the social sciences, conveys that it would be unrealistic to believe that administrators could expect to survive in

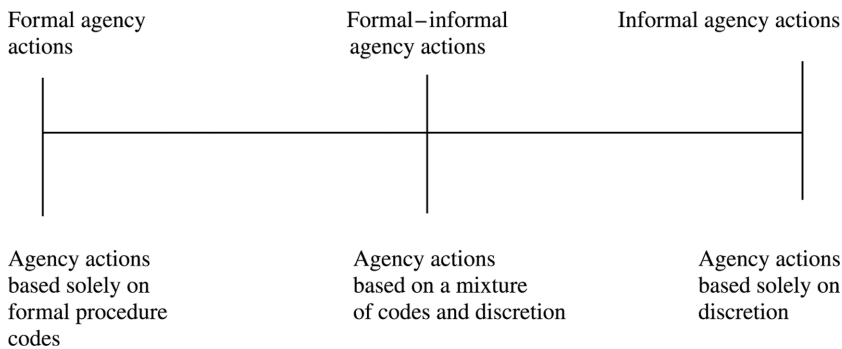


Fig. 1 Discretion's role in formal-informal agency actions.

office for long by totally ignoring prevailing legal and professional sanctions when employing their discretionary powers. An appropriate response to systemic demands would entail keeping the use of discretion within at least minimally tolerable bounds. However, this is not to suggest that administrative discretion is not abused. In fact, social scientists and legal scholars have devoted considerable attention to studying abuses of discretion and how to control them. Yet, it is maintained that the vast majority of discretionary acts by our public administrators are reasonable because they are guided by numerous legal, ethical, and professional standards of conduct forced upon them by the system in which they function. Even when reviewing court cases involving the use of discretion, one finds most allegations of discretionary abuse rest on technical legal arguments, that is, on whether an administrator acted within or outside their discretionary authority, as defined by specific statutes or rules.

EXPANDING ON THE MEANING OF ADMINISTRATIVE DISCRETION

Discretionary actions by administrators are informal administrative acts, even though they may be within an administrator's formal authority. According to Kenneth C. Davis, perhaps the most authoritative source on the concept of administrative discretion, public administrators have discretion whenever the effective limits on their power leave them free to make choices among possible courses of action or inaction.^[2] To Davis, it is important to stress, as it is implied above, that discretion is not limited to what is authorized or what is legal, but includes all that is within the *effective* limits on the officer's power.^[3] Davis added that the specific phraseology he employs to describe discretion is necessary because a good deal of discretion is illegal or of questionable legality.^[3]

Davis stated that the lifeblood of the administrative process is informal discretionary action.^[4] Further, he emphasized that informal discretionary power is enormous, especially when compared with an

administrator's quite limited formal powers.^[4] It is appropriate to mention here that political scientists, in reference to presidential powers, have for many decades acknowledged that president's informal powers are many times greater than their formal powers, as sanctioned by the Constitution, various statutes, and court decisions.^[5,6] Scholars have noted that, although vague and general, statutes and constitutional provisions give rise to discretionary power, but they do not effectively control or limit discretion. Consequently, willful administrators, including presidents, can use their discretionary authority, to make themselves far more powerful than their creators, the Constitutional framers and legislators, probably originally intended.^[1]

When examining the breadth of informal discretionary power, it is easy to understand why it is said to be the lifeblood of the administrative process. Informal discretionary powers include such positive functions as initiating, prosecuting, negotiating, settling, contracting, dealing, advising, threatening, publicizing, concealing, planning, recommending, and supervising.^[1] But the most awesome discretionary power is the omnipresent power to do nothing.^[1] That is, administrators may simply decide not to initiate, prosecute, negotiate, settle, contract, deal, and so on. These negative discretionary powers, in practice, provide administrators with more discretionary clout than positive discretionary powers, because it is far more difficult to check the former than the latter. That is, action, simply because it is action, is easier to control than inaction because checks are designed mostly to cope with abusive action, not abusive inaction.

For those interested in the problems of implementing public policies, the above has obvious implications. Through failing to act, or acting feebly, administrators can use their discretionary powers to help ensure the failure of certain duly authorized public programs under their jurisdiction that they disfavor. Although courts have ruled that administrators are not free to use their discretionary authority to block the implementation of programs authorized by Congress or to prevent persons from receiving their statutory

entitlements, strong-willed administrators may nevertheless have their way in most instances, especially because forcing administrators to comply with statutory demands through court action is so difficult.^[7] As Davis and Pierce concluded: “Judicial review of agency action can never be more than a partial solution to the problem of discretion.”^[1] “Federal agencies take millions of actions every year. It is unrealistic to expect courts to identify or deter each occasional case in which agencies abuse their discretion.”^[4,8] Besides, courts work too slowly, frequently acting after the damage has been done (after administrators have achieved their purpose).

When we think of the discretionary power of administrators, we probably think in terms of how administrative officials may employ their broad discretion to say yes or no to certain major program proposals. For example, in *D.C. Federation of Civic Associations v. Volpe*, 459 F.2d 1231 (1971), John Volpe, then Secretary of Transportation, used his discretion to approve construction of the Three Sisters Bridge, a controversial project vehemently opposed by many civic associations and environmentalists. Although Volpe’s decision essentially constituted an abuse of discretion in the eyes of Judge Bazelon, the *Volpe* case served, inter alia, to demonstrate the vast discretionary powers administrators possess to make final decisions, thus likely determining the fate of major public programs. It is true that administrators possess great discretionary powers in making final decisions, but Davis holds that administrators’ interim decisions may actually carry more weight than their final decisions.^[4] Interim actions may be more subtle, but they can nevertheless make or break programs. In essence, interim decisions determine what steps, if any, should be taken to reach some end. Of course, an administrator’s interim decisions play a significant role in determining the possible ends that will result. Typically, agency administrators use their discretion to make interim decisions in response to such key questions as: Should we investigate the X Company? Should we institute an investigation proceeding? Should we make it a prosecutory proceeding? Or should we make it somewhat one and somewhat the other, allowing its nature to change as it progresses? Should we give X a chance to escape by agreeing to comply? Or should we make an example of X? Should we start a cease-and-desist order proceeding, or should we recommend a criminal prosecution to the Department of Justice?^[4]

Such interim questions appear to be endless. Yet, how each question is answered may make a big difference to, say, Company X, as X may be vitally affected

at every step. Even if it is found guilty of nothing, the investigation may cost half a million dollars. Or its license worth millions of dollars may be jeopardized. Adverse publicity may cripple its business.^[9–11] Discretionary interim decision making may also play a critical role in determining the fate of public policies. In *Implementation*, Jeffrey Pressman and Aaron Wildavsky held that success in implementing public policy programs depends heavily on whether administrators make the vital discretionary interim decisions that permit programs to pass the crucial clearance points necessary for program survival.^[3] In sum, dramatic final actions by public agencies may make frontpage news headlines (“Fed hikes interest rates against Bush’s wishes”), but the daily, routine, nonnewsworthy interim decisions (“Fed decides to discount Labor Department’s 2002 projection on unemployment”) are largely responsible for preventing such agency headline stories from ever occurring.

Thus far, administrative discretion has been defined in the context of substantive decision making, but the scope of informal discretion extends much further. Says Davis: “Discretion is not limited to substantive choices but extends to procedures, methods, forms, timing, degrees of emphasis, and many other subsidiary factors.”^[3] Actually, because administrative law focuses on procedural due process, administrative law scholars and the courts are not so much interested in what substantive choices were made, but in how discretion was used by administrators in making their choices. Fig. 2 illustrates how administrators can ideally exercise their discretion in resolving an administrative problem. This simple model conveys that to resolve an administrative problem, administrators would first have to employ discretion, acknowledge the problem, discover the relevant facts, and then use discretion to search for those applicable laws, rules, and so on, that would settle the matter. But this ideal conceptualization is grossly complicated in reality. In the real world, administrators may use their discretionary powers to deny that a problem even exists. This action in itself must be regarded as a powerful exercise of discretion.

Thus, even if administrators acknowledge the existence of the problem, discretion may be used to define it in a particular way. Naturally, the discretion to define the problem in various ways provides administrators with the flexibility to arrive at different possible solutions. In addition, administrators can employ their discretion to determine what facts are relevant and what laws, rules, and so on, are pertinent. Of course, the discretion to include and exclude facts that are perceived

Administrative problem | Discovery of relevant facts | Application of relevant laws, rules, etc., to facts | Responsible administrative decision

Fig. 2 Simplistic, ideal exercise of discretion.

as relevant or irrelevant allows administrators the opportunity to apply or not to apply various laws, rules, and so forth, because the facts of a situation obviously determine what laws and rules should be applied.

CONCLUSION

Discretionary decisions are also clouded by administrators' past experiences, present environmental demands, politics, and personal values. Regarding police discretion, Joel Samaha commented: "Discretion creates an enormous gap between what law books say and what law officers actually do, because class, race, economics, and politics influence discretion."^[12] Uncertainty also plays an intervening role. A decision as to what is desirable may include not only weighing desirability, but also guessing about unknown facts and making a judgment about doubtful law, and the mind that makes the decision does not necessarily separate facts, law, and discretion. Furthermore, the term discretion may or may not include the judgment that goes into finding facts from conflicting evidence and into interpreting unclear law; the usage is divided.^[13] In sum, discretionary decisions are complicated informal administrative actions that involve factors and processes that are extremely difficult to pinpoint and assess. It is partly for these reasons that abuses of administrative discretion are hard to prove, especially if bad faith must be shown to substantiate the charge that administrators abused their discretionary authority when reaching a decision.^[14]

REFERENCES

1. Davis, K.C.; Pierce, R.J., Jr. *Administrative Law Treatise*, 3rd Ed.; Little, Brown, and Co.: Boston, 1994; Vol. 1, p. 22; Vol. 3, chapt. 17; p. 105.
2. Mashaw, J.L. *Due Process in the Administrative State*; Yale University Press: New Haven, 1985; 12–28.
3. Davis, K.C. *Discretionary Justice: A Preliminary Inquiry*; Greenwood: Westport, CT, 1980; 4–5. This book is a reprint of his classic work first published by University of Illinois Press (Urbana, 1969).
4. Davis, K.C. *Administrative Law: Cases-Text-Problems*; West Publishing Company: St. Paul, MN, 1977; 440–447.
5. Koenig, L.W. *The Chief Executive*, 5th Ed.; Harcourt Brace Jovanovich, Publishers: New York, 1986; especially chap. 1.
6. Thomas, N.C.; Pika, J.A.; Watson, R.A. *The Politics of The Presidency*, 3rd Ed.; Congressional Quarterly Press: Washington, D.C., 1994; pp. 23–37, 434–453.
7. For a discussion of such court rulings see, Warren, K.F. A special deconcentration: a problem greater than school desegregation. *Adm. Law Rev.*, Fall 1977, 29, 577–99.
8. Pierce, R.J., Jr.; Shapiro, S.A.; Verkuil, P.R. *Administrative Law in Process*, 3rd Ed.; Foundation Press: New York, NY, 1999; 132–139.
9. Pressman, J.; Wildavsky, A. *Implementation*; University of California Press: Berkeley, 1973; chaps. 1–3.
10. Mazmanian, D.A.; Sabatier, P.A. *Implementation and Public Policy*; Scott, Foresman: Glenview, IL, 1983; especially chaps. 1–3.
11. Durant, R.F. EPA, TVA and pollution control: implications for a theory of regulatory policy implementation. *Public Adm. Rev.* July/August 1984, 44, 305–313.
12. Samaha, J. *Criminal Law*, 2nd Ed.; West Publishing Company: St. Paul, MN, 1987; 14 pp.
13. See, for example, the court's arguments in defense of the broad use of discretion in *United States v. International Telephone Corp. and Hartford Fire Insurance Co.*, 349 F.Supp. 22 (1972). It is not as necessary to prove bad faith today because good faith cannot be used as a defense when constitutional rights are violated by public officials.
14. Samaha, J. Some reflections on the anglo-saxon heritage of discretionary justice. In *Social Psychology and Discretionary Law*; Abt, L.E., Stuart, I.R., Eds.; Van Nostrand Reinhold: New York, 1979; 4 pp.

Administrative Law Judges and Agency Adjudication

William D. Schreckhise

Department of Political Science, University of Arkansas, Fayetteville, Arkansas, U.S.A.

INTRODUCTION

Administrative law judges (ALJs) and agency adjudications are two things about which most people know very little, but both play an important role in the operations of government in the United States. Adjudications and agency hearings are an important component of regulatory enforcement, entitlement disbursement, and internal agency management. Administrative law judges preside over disputes between two or more parties, much like a judge presides over cases brought before a court. However, an administrative law judge is an employee of the *executive* branch of government and, often, one of the parties in the dispute is the agency for whom they are employed. Adjudications are the equivalent to cases and are the conflicts over which the administrative law judges preside. This entry will present the origins of administrative law judges and discuss the current role they play at the state and federal government adjudications.

ADMINISTRATIVE PROCEDURES LAWS AND ADMINISTRATIVE LAW JUDGES

Currently, 1286 ALJs serve in the federal government, holding positions in 26 different agencies (Table 1). They deal with such widely varying topics as disputes over continuing Social Security Disability Insurance benefits and the application of regulations of the U.S. Securities Exchange Commission.

The position of administrative law judge originated with the passage of the federal Hepburn Act (1906). In this act, Congress granted the Interstate Commerce Commission (ICC) the power to appoint “hearing examiners” to act on the commission’s behalf in giving oaths, taking testimony, examining witnesses, and viewing evidence. The ICC’s success with this new position prompted other agencies to follow suit. Between the years 1913 and 1940, Congress granted 18 other agencies the power to appoint their own hearing examiners. The federal Administrative Procedure Act (APA) of 1946 established the hearing officer as a clearly distinct judicial power in each agency. These positions were created to constitute an independent corps of judicial actors assigned powers to preside over agency hearings, but were to do it in a manner less

formal and more flexible than courtroom proceedings, and who could develop expertise in more technical areas of policy. These actors were to remain within each agency, yet through the APA, Congress erected institutional safeguards to ensure that ALJs would hear cases in an unbiased manner, ideally free from agency pressure to ensure due process for the parties involved.^[1] State APAs created similar positions following a comparable logic of organization and institutional design. In 1972, the U.S. Civil Service Commission changed the title of “hearing examiner” to that of “administrative law judge” to reflect a recognition that, in many important areas of public life, ALJs would be hearing cases *independent* of agency pressure, i.e., carrying out the role of an impartial judge in the standard sense.

The Administrative Procedure Act granted federal ALJs a substantial degree of autonomy from their agencies. The APA gave to the Civil Service Commission (now the Office of Personnel Management) the power to determine the qualifications and compensation of individual ALJs. Under the APA, ALJs can be removed only for cause, and before one can be disciplined, demoted, suspended, or dismissed, they first must receive a hearing before the Merit Systems Protection Board. The Office of Personnel Management sets administrative law judges’ qualifications, and the APA ties ALJ compensation to the Executive Schedule. The Civil Service Reform Act (1978) further protected ALJs by explicitly exempting them from annual performance appraisals by their agency, and today, ALJs are the only members of the federal Senior Executive Service who are exempt from them.^[2] When these protections are considered along with the other protections afforded them under the APA, the federal administrative judiciary is clearly the single-most protected class of federal employees vis-à-vis employing agency influence.

However, ALJ independence is not absolute. In the eyes of the courts, federal ALJs are not “constitutionally protected” as are their regular courtroom colleagues. They are also subject to the agency in matters of interpreting the law and agency policy, and the courts have concluded that agencies can assign cases to specific ALJs as they see fit. All federal agencies using ALJs employ some type of review within the agency,^[3] and the courts have ruled that in reviewing

Table 1 Federal administrative law judges and their agencies

Agency	Number of ALJs
Commodity Futures Trading Commission	2
Department of Agriculture	2
Department of Education	1
Department of Energy	15
Department of Health and Human Services	9
Department of Homeland Security	7
Department of Housing and Urban Development	5
Department of Interior	11
Department of Justice	4
Department of Labor	47
Department of Transportation	2
Department of Treasury	2
Environmental Protection Agency	6
Federal Communications Commission	2
Federal Maritime Commission	3
Federal Mine Safety and Health Commission	9
Federal Trade Commission	2
International Trade Commission	4
Merit Systems Protection Board	5
National Labor Relations Board	54
National Transportation Safety Board	4
Occupational Health Safety Review Commission	11
Securities and Exchange Commission	5
Social Security Administration	1079
Total	1292

Source: Office of Personnel Management, FedScope, <http://www.fedscope.opm.gov/employment.htm> (accessed July 2003).

an ALJ decision under the APA, the agency maintains “all the powers which it would have in making the institutional decision.”^[4]

STATE ADMINISTRATIVE LAW JUDGES

On the heels of the passage of the federal APA, a working group of representatives from the American Bar Foundation and the National Conference of Commissioners on Uniform State Laws created a model state Administrative Procedure Act. The model act was fashioned after the federal APA and included provisions for the creation of state ALJs. The model APA has been adopted at least in part by all 50 states

and by the government of the District of Columbia. Today, every state employs ALJs in a fashion similar to federal ALJs.^[5]

State-level ALJs hear different types of cases from their federal counterparts. The bulk of federal ALJs hears cases involving benefits for Social Security claims, with the remainder largely presiding over regulatory hearings (environmental, transportation, securities, mergers, labor, and tariffs). State ALJs hear cases involving workers compensation, alcohol sale permits, public health regulations, environmental protection, utilities regulations, and employment discrimination. Some states will even relegate to their ALJs the authority to preside over driver’s license revocation proceedings.

The ways in which ALJs perform their duties vary somewhat from state to state. In some state jurisdictions, an ALJ is an employee of the agency for which he or she hears cases. In other states, ALJs are employed in a separate “central panel” agency. Central panel agencies were created at the state level to give state ALJs even greater decisional independence. The creation of central panels began after the Administrative Conference of the United States recommended the creation of such agencies for the federal government in the 1970s, and the idea was vigorously supported by state bar organizations.^[6] By 2003, 26 states employed central panels systems, most of them modeled after the California central panel system that was adopted originally in the 1940s^[7] (Table 2). In states without central panels, it is often the practice for the ALJs to be watched over by a board or commission, whose members may be drawn from the interests in cases the ALJs are to hear, as some states require in statute representation on the boards of particular interests.^[8]

It should be noted that some states recognize the difference between formal and informal adjudicative hearings. Generally speaking, state-level adjudications are even less formal than are federal adjudications, generally require less proof to prove a case, and are less likely to involve attorneys for private parties involved in the hearing.^[8]

AGENCY ADJUDICATIONS

Administrative law judges are the most visible actor in state and federal agency adjudications because they hear the vast majority of them. However, they are not the only ones who can preside over adjudications. Namely, the agency commission, board, or agency head may also preside over them. There are no specific processes for how hearings are held. Although all federal agencies must adhere to portions of the APA, each agency is free to develop its own specific methods.

Table 2 State central panel agencies

State	Agency
Alabama	Alabama Administrative Law Judge Central Panel
Arizona	Office of Administrative Hearings
California	Office of Administrative Hearings
Colorado	Division of Administrative Hearings
Florida	Division of Administrative Hearings
Georgia	Office of State Administrative Hearings
Iowa	Department of Inspections and Appeals
Louisiana	Division of Administrative Law
Maine	Division of Administrative Hearings
Maryland	Office of Administrative Hearings
Massachusetts	Division of Administrative Law Appeals
Michigan	Bureau of Hearings
Minnesota	Office of Administrative Hearings
Missouri	Missouri Administrative Hearing Commission
New Jersey	Office of Administrative Law
North Carolina	Office of Administrative Hearings
North Dakota	Office of Administrative Hearings
Oregon	Office of Administrative Hearings
South Carolina	Administrative Law Judge Division
South Dakota	Office of Hearing Examiners
Tennessee	Administrative Procedures Division
Texas	State Office of Administrative Hearings
Washington	Office of Administrative Hearings
Wisconsin	Division of Hearings and Appeals
Wyoming	Wyoming Office of Administrative Hearings

Source: National Association of Administrative Law Judges (2003).

In many respects, agency adjudications look like regular courtroom trials, but are less formal than a courtroom proceeding. Exactly how formal the proceedings are to be is up to each agency, and state-level adjudications tend to be even less formal than federal ones. At a minimum, administrative law judges (or whoever else is presiding over the adjudication) will manage the hearing (e.g., deciding who gives testimony and when), develop the record, maintain the integrity of the hearing, and render a decision based on the record generated in the hearing.

Who gets a hearing? The answer would seem to be a simple one, but it is not. Generally speaking, under the Fifth and Fourteenth Amendments to the U.S. Constitution, anyone in jeopardy of losing their life, liberty, or property at the hands of the government must be afforded due process. At the very least, this means they have to be given a hearing before an independent and

unbiased decision maker. Because it is possible that an individual could lose his or her property at the hands of an agency (questions of life and liberty are handled exclusively by the courts), it seems anyone in this position should be given a hearing. However, the courts have been reluctant to require agencies to conduct adjudicatory hearings in *all* cases where a person stands to lose some type of property. When Congress has passed a law stating an agency must hold a hearing “on the record” for a particular type of case (such as in licensing radio stations), or when the courts have inferred Congress intended to require the agency to hold such hearings (absent of any specific language on the matter), the adjudication provisions of the APA will apply. If there is no such statute, and the agency fails to grant the hearing, then it is up to the courts to decide whether a hearing is in fact required. However, scholars have been critical of the courts for failing to develop a coherent set of principles governing administrative due process and what constitutes a valid hearing.^[9] Even if an individual can show a court they stand to lose property, the Supreme Court ruled in *Mathews v. Eldridge* (424 U.S. 319, 1976) that the agency can take into consideration the burden a hearing would impose on the agency when deciding to hold one, which means the court might not require the hearing even if a person’s property is at stake. Furthermore, even if it is apparent that the individual has a right to a hearing, the Supreme Court has ruled that the hearing can be held *after* the property has been taken away by the agency.^[10]

Generally speaking, adjudications can be grouped into one of four types, increasing in degrees of formality from alternative dispute resolution (ADR) hearings, paper hearings, oral hearings, to formal hearings.^[11] In ADR proceedings, the parties in the case agree to reach a settlement through compromise and negotiation. So-called “paper hearings” are a bit more formal in the respect that the outcome is the product of the hearing officer (and not negotiation), but there is no actual hearing wherein oral testimony is presented. Instead, arguments are presented exclusively via written briefs. Even more formal are oral hearings where the parties present their arguments orally, but without discovery, prehearing conferences, or cross-examination during testimony, as in the case of formal hearings.

The federal APA does not provide a set of comprehensive procedures governing the holding of hearings. Generally, formal hearings will include opening statements, the presentation of the case by the proponent (usually the agency in regulatory hearings), the presentation of the case by the defendant, rebuttal (as allowed by the presiding officer), and closing arguments. Agencies vary by what can be admitted into evidence and are not bound by the Federal Rules of Civil Procedure’s rules on allowable evidence,

although each agency is required to generate their own rules. Agencies may base their decision on evidence that would not otherwise be admitted in a court case (such as hearsay), but the decision must also be based on enough evidence that would have been admitted or show why such evidence was not available.^[12]

Depending on the type of hearing, the administrative law judge will issue an initial decision or recommended decision. Initial decisions are generally issued when the case deals with a well-established issue and, unless overturned by the agency, becomes the agency's decision. If the case deals with a novel issue, the ALJ will issue a recommended decision that may require further action (such as additional proceedings) by the agency. The burden of proof is on the party bringing the case. Generally speaking, the standard of proof is "preponderance of evidence" if Congress has not set the standard already.^[13] However, for immigration deportation cases, reviewing courts have required higher standards, such as the more stringent "clear and convincing evidence" standard.^[14] Many agencies have internal appellate processes, having provisions that allow individuals to appeal the case to the agency board or an intermediate appellate body. Once all appellate remedies have been exhausted in federal adjudications, parties may appeal the case to a U.S. Court of Appeals, except cases from the Social Security Administration which are heard by U.S. District Courts.

CONCLUSION

Few of us who have not had contact with ALJs and agency adjudications know of their existence, but they warrant a degree of attention. Although the courts have recognized their constitutional legitimacy, their platypus-like characteristics of having both judicial and executive functions make them intrinsically interesting. The sheer volume of ALJ activity should garner our attention as well. In the case of the U.S. Social Security Administration alone, ALJs presided over 377,163 cases in fiscal year 1996; in contrast, in 1998,

all 94 U.S. District Courts presided over only 280, 293 civil cases.^[15] Although the average dollar amounts at stake in District Court cases were probably greater, ALJs have been deemed an integral part in the implementation of portions of the Social Security program, the largest of the federal domestic programs. They play important roles in state-level programs as well.

REFERENCES

1. Rich, M. Central panels of administrative law judges: an introduction. *Judicature* **1981**, 65 (5), 233-234.
2. Lubbers, J. The federal administrative judiciary: establishing an appropriate system of performance evaluation for ALJs. *Adm. Law J. Am. Univ.* **1994**, 7 (3), 598-628.
3. 1992 A.C.U.S. 35.
4. See *The Association of Administrative Law Judges, Inc. v. Heckler*, 594 F. Supp. 1134 (D.D.C., 1984).
5. Schwartz, B. *Administrative Law*, 3rd Ed.; Little Brown: Boston, 1991.
6. Rosenblum, V. The central panel system: enhancing administrative justice. *Judicature* **1981**, 65 (2), 235.
7. Rich, M.; Brucar, W. *The Central Panel System for Administrative Law Judges: A Survey of Seven States*; American Judicature Society: Chicago, 1983.
8. Bonfield, A. State law in the teaching of administrative law: a critical analysis of the status quo. *Tex. Law Rev.* **1982**, 61 (1), 95.
9. Carter, L.; Harrington, C. *Administrative Law and Politics: Cases and Comments*, 3rd Ed.; Longman: New York, 2000.
10. *Cleveland Board of Education v. Loudermill*, 470 U.S. 532 (1985).
11. Cooper, P. *Public Law and Public Administration*; F.E. Peacock: Itasca, IL, 2000.
12. *Perales v. Richardson*, 402 U.S. 389 (1971).
13. *Steadman v. U.S.*, 450 U.S. 91 (1981).
14. *Woodby v. INS*, 385 U.S. 276 (1966).
15. Social Security Administration. *1998 SSI Annual Report*. U.S. Government Printing Office: Washington, DC, 1998.

Administrative Reform in Southeast Asia

M. Shamsul Haque

Department of Political Science, National University of Singapore, Singapore

INTRODUCTION

Since the end of colonial rule, administrative reform has been one of the most common domains of social change pursued by developing nations. In these countries, the process of administrative reform became quite intensive immediately after their decolonization to reduce their dependence on colonial administration, create more indigenous institutional structure, and expand the role of government in national economies.^[1] Beyond this historical reason, there were inherent weaknesses in governing institutions—including bureaucratic inefficiency, administrative incapacity, social instability, market failure, and unemployment problem, which also required administrative reform. In more recent years, however, administrative reform in developing countries has been driven by the increasing local needs and demands as well as the contemporary global forces and pressures, in response to which the state has to restructure its administration and revive its managerial capacity in line with the similar reinvention in governance found in advanced capitalist nations.^[2,3]

Administrative reform has traditionally been understood as a formal, planned, and deliberate change in various dimensions (e.g., structure, process, behavior, norms) of the administrative system to improve its efficiency, quality, coordination, motivation, responsiveness, accountability, and so on.^[1,4] More recently, instead of administrative reform, many scholars and experts use the term “governance reform,” which has much broader connotation, encompassing changes in the civil service, policy process, civil society, and state-market relations.^[5] In terms of scope, administrative reform covers all major levels of government (federal, state, local) and public sectors (agriculture, industry, commerce, finance, transport, education, health).^[3,6] Internally, it includes various structural, functional, procedural, normative, and attitudinal changes in the administrative system.

During the recent decades, there has emerged almost a paradigm shift, especially in terms of greater emphasis on the market-driven objective, role, structure, policy orientation, and norms of government administration. Compared to the earlier state-centered model, this newly emerging market-centered model of reform has become the major framework for undertaking administrative restructuring in most countries of

the world. This current reform emphasizes the goal of efficiency and effectiveness, supportive or facilitating role, disaggregated and flexible managerial structure, result-based performance, promarket policies, customer orientation, and business sector norms.^[7,8] Such an antibureaucratic mode of administrative reform largely represents the basic components of “reinventing government” prescribed by Osborne and Gaebler^[9] as well as the principles of the so-called “new public management” presented by Hood,^[10] and it differs substantively from the past reform initiatives undertaken by various governments in line with the traditional bureaucratic model.^[9–11]

The above market-centered model of administrative reform, which emerged in advanced capitalist nations, was gradually adopted by developing countries often under the structural adjustment program prescribed or imposed by international aid agencies. In line with this global trend, in various degrees, Southeast Asian countries have embraced the major components of such reform agenda—including the downsizing of the public sector, deregulation of service delivery, divestment of state enterprises, liberalization of trade and investment, corporatization of public agencies, use of result-based budget and performance indicators, and so on—in the name of economic efficiency, competitiveness, service quality, value for money, and customer satisfaction.^[12,13] There are some major causes and implications of this reform process in Southeast Asia.

FEATURES AND TRENDS OF REFORMS IN SOUTHEAST ASIA

Countries in Southeast Asia—including Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, and Vietnam—vary significantly with regard to their territorial size, demographic pattern, ethnic and religious composition, colonial background, social structure, political system, and economic development.^[14] In the past, this diversity created certain differences among these countries in terms of the nature and objectives of administrative reform. In recent years, however, most of these countries have adopted reforms in line with the globalized market-led model of public management mentioned above. This section discusses the major features and

trends of current administrative reform, including the institutional, functional, structural, normative, and policy reforms in various Southeast Asian countries.

Institutional Reform

Most governments in Southeast Asia have undertaken major reform initiatives to restructure the public sector based on promarket assumptions. Examples of such initiatives include Public Service for the 21st Century (PS21) in Singapore, Panibagong Sigla 2000 (Renewed Vigor 2000) in the Philippines, Public Sector Management Reform in Thailand, Malaysia Incorporated in Malaysia, Resolution on Public Administration Reform in Vietnam, and National Development Program in Indonesia.^[13,15] In addition, a new set of market-friendly state institutions has been introduced, such as the Public Sector Divestment Committee in Singapore, the Committee on Privatization and the Asset Privatization Trust in the Philippines, the Steering Committee on Reduction in the Size of the Public Service in Malaysia, and the Public and Private Sector Committee in Thailand.

A more critical institutional feature of administrative reform in the region, however, is the streamlining or downsizing of the public sector in line with the current global trend. For example, the Thai government adopted the downsizing strategies such as the recruitment freeze, early retirement scheme, abolition of postretirement vacancies, and replacement of underutilized employees.^[15,16] The Philippine government also decided to downsize the number of public sector employees by 5–10% and Singapore government by 10%.^[17,18] Similarly, the governments in Indonesia and Malaysia have decided to streamline state bureaucracy by cutting public expenditure and reducing the number of civil servants.^[19,20] Even in communist Vietnam, the government has introduced unprecedented reform measures and reduced the number of ministries and agencies from 76 in 1986 to 48 in 2001 and the percentage of public sector employment from 10% in 1991 to 8.7% in 1994.^[19,21]

Functional Reform

In Southeast Asian countries, the state bureaucracy played a crucial role in socioeconomic development, whereas the private sector took part in this development process under state regulation, control, and coordination. However, the recent administrative reform has moved away from such a dominant role of the monopolistic public sector and emphasized the role of market competition and private enterprise. This new mode of reform defines the function of public administration as that of a catalyst or facilitator rather

than main actor or leader—the purpose is to restructure the role or function of the administrative system in such a manner that it can enable (rather than control) the role played by the business sector.^[22] In both Malaysia and Singapore, the role of the public sector has been redesigned to reduce its functional scope and to transform it into an enabling agent for the private sector, which now is supposed to play the leading role in economy and society.^[11,20] Such a facilitating role of the public sector has also been prescribed in the Philippines, Thailand, and Vietnam in their recent initiatives of administrative reform.^[23,24]

The functional dimension of the current administrative reform in the region has also greater emphasis on the satisfaction of public sector “customers,” which represents a shift from the earlier focus on the *entitlement* of “citizens” to basic services provided by the state. Similar to the businesslike customer-oriented approach adopted recently in the public sector by most developed nations, the concern for customers or clients has gained increasing significance in public management in Singapore, Malaysia, Thailand, and the Philippines. Top public officials are now encouraged to develop customer-oriented outlook and customer-friendly attitude in these countries.^[13]

Policy Reform

In most Southeast Asian countries, compared with their earlier state-centric public policies such as nationalization, regulation, and protectionism, the recent reform initiatives have moved toward more promarket policies such as privatization, deregulation, outsourcing, and liberalization. In the region, the privatization policy encompasses all major sectors, including telecommunications, electricity, airlines, railway, banking, finance, petroleum, transport, mining, construction, tourism, and so on. For instance, Malaysia introduced massive privatization under its Privatization Masterplan, Singapore adopted privatization policy planned by its Public Sector Divestment Committee, and the Philippines launched the privatization program through its Committee on Privatization. Similar privatization exercises were carried out in Thailand, Indonesia, Cambodia, and Vietnam.^[19]

Southeast Asian countries have also pursued the deregulation and liberalization of trade, investment, and foreign ownership. For foreign investment, Indonesia has liberalized such sectors as electricity, railways, telecommunications, and airlines, which used to be under state control. Malaysia, Thailand, and the Philippines have withdrawn restrictions on foreign ownership and adopted incentives for foreign investors such as tax exemption, duty-free imports, and so on. The liberalization of trade and investment has also

taken place in Cambodia, Laos, and Vietnam.^[25] These recent changes certainly represent a significant policy reform in Southeast Asia.

Structural Reforms

Despite the diverse historical origins of the administrative systems in Southeast Asian countries, in general, they became increasingly based on the principles of a bureaucratic model that prescribes central control over financial, procedural, and personnel matters. However, in line with the abovementioned “new public management” model suggesting disaggregated and decentralized management practices, most countries in the region have moved toward reforming their administrative systems based on greater managerial and financial autonomy and flexibility. In Singapore, for instance, the government has restructured various agencies or departments into the so-called “autonomous agencies” assigned with considerable managerial autonomy in personnel and financial matters.^[11]

In the case of Thailand, the government has also created various “autonomous public organizations” that have been provided with autonomy in their respective financial and personnel policies. In Malaysia, the traditional financial controls over government departments have been transformed into more decentralized financial management with greater autonomy in the budget matters. Similar trends of administrative reform in favor of operational autonomy in management can be found in Indonesia and the Philippines.^[15]

Normative-Attitudinal Reform

In the past, in line with the bureaucratic model that emerged largely in Western liberal democracies, Southeast Asian countries adopted various reform measures to inculcate administrative norms and attitudes based on neutrality, equity, representation, and accountability, although such standards were not often observed in practice. However, under the recent reform in these countries, greater emphasis has been placed on values such as efficiency, economy, competition, value for money, service quality, and customer satisfaction. In Malaysia and the Philippines, there is greater concern for administrative efficiency, effectiveness, quality, partnership, and customer-orientedness.

Similarly, in recent development plans and programs, Indonesia and Thailand have emphasized these market-based normative and attitudinal standards. In the case of Singapore, the current reform initiatives for the public service, especially the PS21, also stress the realization of values such as competition, efficiency, quality, and customer orientation.^[13] All these countries have also adopted various

training programs to attitudinally reorient their public employees to make them more quality-conscious and customer-friendly.

CAUSES AND IMPLICATIONS OF REFORMS IN THE REGION

In general, the main factor or force behind the current market-driven reform in the public sector has often been attributed to the globalization of trade and investment, decline in state capacity, and erosion of state autonomy, which allegedly led to the crisis of the state and thus required substantive reforms in state policy and administration. According to Pereira,^[3] this pressure of globalization on the state was reinforced by the international business interests and neoliberal policy elites demanding or advocating privatization, deregulation, liberalization, and other market-friendly institutional reforms. Because Southeast Asian countries were already integrated with the world capitalist system, they had to respond to the new globalized political economy and its forces by undertaking such reform initiatives often prescribed by international aid agencies.

For most developing countries, including those in Southeast Asia, the World Bank identified some major problems with the public sector, suggested reform measures such as downsizing, retrenchment, divestment, and so on, and extended loans mainly to those countries which could demonstrate commitment to these prescribed reforms.^[5] To a certain extent, countries such as Thailand, Indonesia, the Philippines, and Vietnam introduced divestment, deregulation, and liberalization under the influence of the International Monetary Fund and the World Bank.^[19] It is often the external debt and dependence of these countries that created an opportune context for these international financial institutions to exert such policy influence or pressure. However, it should be added that there are also internal factors—including domestic fiscal crisis, public sector inefficiency, and local business interests—which required such market-led reforms in these countries.

What are the implications of these contemporary administrative reforms? In terms of positive outcomes, the current promarket administrative reform represents a serious challenge to the traditional, centralized, monopolistic, and elitist state bureaucracy that existed in most Southeast Asian countries. This recent reform is likely to make the public sector more competitive, cost conscious, and performance-oriented and contribute to a greater degree of transparency, decentralization, service quality, and responsiveness to its clients or customers. However, there are critics who often point out various negative consequences of this

contemporary reform. More specifically, for some scholars, the new model of administrative reform based on neoliberal assumptions and market-driven principles and policies has been largely imposed on developing nations, and it represents a form of ideological hegemony.^[26]

It is stressed that under the current public sector reform, the privatization and restructuring may have diminished state sovereignty and increased foreign ownership in certain Southeast Asian countries (e.g., Thailand, Indonesia, Malaysia, and the Philippines) that suffer from heavy foreign debt and dependence and thus are vulnerable to external policy imposition.^[27] Internally, on the other hand, policy reforms such as deregulation, divestment, welfare cut, and customer focus may not equally benefit all social sections or income groups—these reform measures are likely to benefit the business and political elites, overlook the concerns of ordinary citizens, and expand the gap between public officials and common citizens.^[19,28]

In terms of the impacts of such market-led reform on the administrative system itself, the increasing similarities between public administration and business management may pose a challenge to the “public” nature of public service, perpetuate its identity crisis, and adversely affect the pride and morale of public employees.^[13] In addition, under the current reform, the expansion of operational and financial autonomy of public managers and their growing interaction and partnership with business executives may generate new avenues for administrative corruption and thus compromise public service integrity. In other words, in the current context of managerial autonomy and partnership with the private sector, it may be necessary to introduce additional safeguards to prevent any kind of wrongdoing resulting from these recent changes.

CONCLUSION

It is evident from the above discussion that during the recent decades, Southeast Asian countries have introduced considerable administrative reforms based on market-based assumptions and principles, which include changes in the public administration system in terms of its institutional nature and scope, role and function, public orientation, structural pattern, and normative priority. Although there are some potential benefits from such administrative reforms, there are also some major adverse consequences. In this regard, there is a need to consider some major critical concerns by top policymakers in charge of initiating and implementing administrative reform in each of these countries.

First, in pursuing administrative reform, most developing countries, including those in Southeast

Asia, have often followed models that emerged in advanced capitalist nations. In particular, the Weberian bureaucratic model that evolved in Western nations was imitated by Southeast Asian countries in modernizing their administrative systems without much attention paid to their unique local contexts. On the other hand, the current process of administrative reform in the region has largely been based on the neomanagerial model (“new public management”) that originated in countries such as America, Australia, Britain, Canada, New Zealand, and so on. This new model of reform, which is globally touted as “good governance” or “best practices” by international agencies, is also being reproduced in developing nations without much concern for the “peculiarities of public service problems and their settings” in these countries.^[26] On this ground, countries in Southeast Asia should try to replace the imitative models of administrative reform borrowed from outside or imposed on them by external actors or agencies.

Second, to pursue a need-based, indigenous model of administrative reform, the policymakers in Southeast Asian countries should examine and seriously consider their respective contextual factors. It is stressed by some scholars that in general, administrative reform should not be based on the assumption of “ideological or cultural supremacy” of any particular society (especially the West); it should rather be determined by each nation’s contextual factors such as political history, cultural tradition, nature of government, and constitutional features.^[26,29] Each country in Southeast Asia has its own unique contextual factors—including the colonial legacy, state formation, political culture, social relations, cultural and religious beliefs, economic resources, and citizens’ needs and expectations—which should be seriously taken into account in formulating administrative reform policies and their eventual consequences. As Heeks and Bhatnagar^[30] mention, in general, “Public managers would do better to open their eyes and ears to their immediate surroundings rather than burying their noses in MBA textbooks in seeking guidance on reform.”

Finally, it is essential to understand that administrative reform should be based on a comprehensive framework, which emphasizes reforms in other social domains (political, economic, and cultural) that affect and are affected by changes made in the administrative system. It is often emphasized that administrative reform can be successful only when it is complemented by corresponding reforms in other realms of society within which the administrative system exists and functions.^[31] In Southeast Asian countries, most initiatives for administrative reform have been undertaken without much innovations made in politics, economy, society, and culture—this parochial or unbalanced nature of administrative reform represents a

major obstacle to its successful realization or implementation. In this regard, the policymakers in each of these countries must reexamine the shortcomings of such a parochial approach to reform that focuses mainly on the administrative system and adopt a more comprehensive and multidimensional reform outlook that covers other relevant and important domains of society.

REFERENCES

- Caiden, G.E. *Administrative Reform*; The Penguin Press: London, 1969.
- Rosenbloom, D.H. Administrative reformers in a global world: diagnosis, prescription, and the limits of transferability. In *Rethinking Administrative Theory: The Challenge of the New Century*; Jun, J.S., Ed.; Greenwood: New York, 2001.
- Pereira, L.C.B. *State Reform in the 1990s: Logic and Control Mechanisms*; Seminar on the Changing Role of the State (sponsored by The World Bank): Hong Kong, September 23, 1997.
- World Bank. *The East Asian Miracle: Economic Growth and Public Policy*. The World Bank: Washington, DC, 1993.
- Management Development and Governance Division (MDGD). *Civil Service Reform Paper*. MDGD, United Nations: New York, 2001; 30–36.
- König, K. Administrative sciences and administrative reforms. *Strategies for Administrative Reform*. Caiden, G.E., Siedentopf, H., Eds.; Lexington Books: Lexington, MA, 1982.
- Reyes, D.R. Public sector reengineering: practice, problems and prospects. *Philipp. J. Public Adm.* **1998**, *42* (3–4).
- Liou, K.T. Linking administrative reform to economic development: issues and the national experience. In *Administrative Reform and National Economic Development*; Liou, K.T., Ed.; Ashgate: Aldershot, 2000.
- Osborne, D.; Gaebler, T. *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*; Plume Books: New York, 1993.
- Hood, C. A public management for all seasons? *Public Adm.* **1991**, *69* (1), 3–19.
- Haque, M.S. Structures of new public management in Malaysia and Singapore: Alternative views. *J. Comp. Asian Dev.* **2002**, *1* (1), 71–86.
- World Bank. *World Bank Annual Report 1996*. The World Bank: Washington, DC, 1996.
- Haque, M.S. New directions in bureaucratic change in Southeast Asia: selected experiences. *J. Polit. Mil. Soc.* **1998**, *26* (1), 96–114.
- Aldaba, F.; Petilla, M.J. *Poverty Situation in Southeast Asia and NGO Responses*; Research Paper No. 1, March 2002, Catholic Institute for International Relations, South East Asia Research and Advocacy Unit, 2002.
- Research Institute for Asia and the Pacific (RIAP). *Public Sector Challenges and Government Reforms in South East Asia: Report 2001*. RIAP, University of Sydney: Sydney, 2001.
- Asian Development Bank. *Governance in Thailand: Challenges, Issues and Prospects*; Asian Development Bank: Manila, 1999.
- Halligan, J.; Turner, M. *Profiles of Government Administration in Asia*; Australian Government Publishing Service: Canberra, 1995.
- World Bank. *Philippines: The Challenge of Economic Recovery*. The World Bank: Washington, DC, 1999.
- Haque, M.S. Reinventing governance in Southeast Asia: what are its impacts on economic sovereignty and self-reliance? The Second International Conference of Japan Economic Policy Association, Nagoya, Japan, Nov 29–30, 2003.
- Kristiadi, J.B. *Administrative reform in Indonesia. The Conference on Public Sector Challenges and Government Reforms in South East Asia*; Research Institute for Asia and the Pacific, University of Sydney: Australia, March 12, 2001.
- McCarty, A. Governance institutions and incentive structures in Vietnam. In *The Conference on Public Sector Challenges and Government Reforms in South East Asia*; Research Institute for Asia and the Pacific, University of Sydney: Australia, March 12, 2001.
- World Bank. *East Asia: Recovery and Beyond*. The World Bank: Washington, DC, 2000.
- United Nations. *Administrative Reforms: Country Profiles of Five Asian Countries*. United Nations Department of Economic and Social Affairs: New York, 1997.
- Tiep, N.T. Government reform for socio-economic development in vietnam. *Asian Rev. Public Adm.* **1998**, *10* (1–2), 172–185.
- Montes, M.F. The economic miracle in a haze. In *Growing Pains: ASEAN's Economic and Political Challenges*; Montes, M.F., Quigley, K.F.F., Weatherbee, D.E., Eds.; Asia Society: New York, 1997.
- Blunt, P. Public administrative reform and management innovation for developing countries. Fourth Global Forum on Reinventing Government—Citizens, Businesses, and Governments: Partnerships for Development and Democracy, Marrakech, Morocco, Dec 11–13, 2002.
- Higgot, R. The Asian economic crisis: a study in the politics of resentment. *New Polit. Econ.* **1998**, *3* (3), 333–356.
- Cook, P.; Kirkpatrick, C. The distributional impact of privatization in developing countries: who gets what, and why. In *Privatization and Equity*; Ramanadham, V.V., Ed.; Routledge: London, 1995; 35–48.
- Roness, P.G. Structural features, institutional characteristics and administrative reforms: theoretical reasoning and empirical measures. In *The ECPR Workshop: Institutional Theory; Issues of Measurement and Change*; Edinburgh, UK, March 28–April 2, 2003.
- Heeks, R.; Bhatnagar, S. Understanding success and failure in information age reform. In *Reinventing Government in the Information Age*; Heeks, R., Ed.; Routledge: London, 1999; 71 pp.
- Haque, M.S. The contextless nature of public administration in third world countries. *Int. Rev. Adm. Sci.* **1996**, *62* (3), 315–329.

Agenda Setting

David A. Rochefort

Department of Political Science, Northeastern University, Boston, Massachusetts, U.S.A.

INTRODUCTION

Agenda setting is the process by which issues are selected for attention by government. Agenda setting is generally defined as the first phase of policy making, before formal consideration of policy proposals. As a scholarly perspective, agenda-setting research focuses on the linkage between the social environment and the polity, highlighting the role of ideas, current events, media coverage, and conflict expansion in the origin and prioritization of issues. Institutional gatekeeping also plays an important role in controlling access to the political agenda.

BACKGROUND AND OVERVIEW

Prior to the emergence of the agenda-setting framework, the study of public policy was limited in several respects. Of primary concern had been the question of participation in political decision-making. Elitists maintained that power and control lie in the hands of a small ruling group,^[1] while pluralists argued that the influence of organized interests varied according to issue.^[2] These approaches gave much attention neither to the predecisional phase of policymaking nor to the means by which private concerns become identified as public issues. Easton's^[3] system model of government did specify societal demands as a key institutional "input," yet without analyzing the pathways by which some demands succeed in entering the political mainstream while others languish without official interest.

The agenda-setting model is concerned with just this transition from social environment to political arena. It arrived on the scene when scholars were increasingly prone to view the policy process as having a natural history that exhibits the properties of a cycle with well-defined beginning, middle, and ending stages.^[4,5] The agenda-setting perspective supplied a new vocabulary and set of analytical tools for examining the early stages of this process in which an issue first acquires its political identification. It also addressed the interplay of political advocacy, media coverage, and public opinion formation as crucial in determining access to the polity. If we recognize society as the breeding ground for an almost unlimited number of conflicts, losses,

and grievances at any point in time, then agenda setting examines the funneling process by which only certain claims among these attain official standing and elicit a response from those in power.

Cobb and Elder's^[6,7] *Participation in American Politics* was pivotal in laying the foundation for this agenda-setting framework. Rather than elite or pluralist theory, Cobb and Elder based their work on Schattschneider's^[8] model of social conflict. Schattschneider acknowledged the bias of the political system, which favors those with power, money, and social status, while noting the tendency of political officials to avoid conflict. However, Schattschneider portrayed conflict expansion as a means by which disadvantaged groups compel official attention to their concerns, to the extent that the group that builds scope, intensity, and visibility in a conflict increases the possibility of agenda access.

For Cobb and Elder, the issues under debate are just as important as the groups involved in political action. Building on Schattschneider's idea of "issue redefinition," i.e., broadening the meaning and value implications of an issue, they proposed that certain issue characteristics—specificity, social significance, temporal relevance, complexity, and categorical precedence—are related to issue expansion. "Triggering mechanisms" are unforeseen events, such as natural catastrophes, changes in technology, social conditions, and scientific knowledge, or a sudden act of violence, which launch an issue by moving it into the spotlight of media coverage and public awareness. Issue advocates promote their cause by taking advantage of these circumstances.

Language animates and directs political conflict. By combining facts and emotion-laden symbols, groups use rhetoric to mobilize and maintain their followers, arouse latent public support, provoke media coverage, and dissuade the opposition.^[6,7] "Initiators," or policy entrepreneurs, are those individuals who lead the way in issue expansion by gathering information, gaining publicity, contributing to organizing efforts, and engaging in symbolic manipulation.

According to Cobb and Elder's formulation, two kinds of agendas exist. The systemic agenda refers to the "general set of political controversies that will be viewed at any point in time as falling within the range of legitimate concerns meriting the attention of the

polity.”^[7] The institutional, or formal, agenda is limited to that subset of issues “scheduled for active and serious consideration by a particular institutional decision-making body.”^[7] When the gap between these two types of agendas becomes extreme in a society, it is a formula for political delegitimation and instability.

AGENDA-SETTING RESEARCH

Nelson^[9] was one of the first scholars to apply the agenda-setting framework in a detailed policy case study. Although the mistreatment and neglect of children is not a new phenomenon, only in the early 1970s did child abuse gain a place on the national legislative agenda. Agenda entrance resulted from the U.S. Children’s Bureau’s interest in child abuse during the 1950s, steady growth of mass media coverage in the 1960s and 1970s, and the diffusion of state child abuse reporting laws in the 1960s. When the Congress adopted the Child Abuse Prevention and Treatment Act of 1974, it not only reflected the special advocacy of a powerful Senator, Walter Mondale, but also increased public awareness of the problem. The issue’s success owed much to its symbolic dimension, as the image of children victimized by adult violence invoked sympathy, stimulated moral indignation, and promoted political consensus.

In the aftermath of Nelson’s work, there has been a steady outpouring of public policy research informed by agenda-setting concepts. The topics of these studies are diverse—aging policy,^[10] social welfare,^[11] pesticides,^[12] drug policy,^[13] children’s health care,^[14] transportation policy,^[15] and more. Today, it would be hard to imagine a public policy textbook that omitted discussion of the politics of agenda setting. Meanwhile, the agenda-setting approach has not remained static. Researchers have both adopted and adapted it, fleshing out the framework’s details while extending it in new directions.

Kingdon^[16] examined the “long smoldering” process of policy initiation. He identified three streams that ebb and flow constantly through the system of government decision-making. A “problem stream” consists of the social problems perceived by government officials to be of interest and importance. A “policy stream” is the “primeval soup” of ideas for government action. Specialized policy communities spanning public and private sectors generate these alternatives and proposals, while policy entrepreneurs work to gain prominence for certain ideas over others. A “political stream” is the combination of several factors—national mood, the alignment of organized political forces, the balance of power and administrative capacity within the government, and the propensity

for bargaining—that define the political system’s readiness for action. Significant changes in the political or problem streams can “open the window” of opportunity in the policy process. At such a time, the usual obstacles are reduced and a compatible policy proposal enjoys heightened chance of adoption. But these conditions are fleeting, and policy entrepreneurs must seize the moment by attaching their preferred proposals to the problem needing solution. For Kingdon, the best way to understand movement on the government agenda is to pay attention to those instances when two or, most powerfully, three decision-making streams become coupled and carry an issue forward on an almost irresistible current of support.

“Punctuated equilibria” is the term used by Baumgartner and Jones^[17] to describe the rise and fall of policy subsystems. A subsystem is an organized institutional structure involving political decision-makers, stakeholder groups, and technical experts who share a central idea about the purpose of a policy area and produce a stable equilibrium in policy making over time. Eventually, however, equilibrium conditions can be overturned, or punctuated, by an alternative “policy image” that captures media interest and broadens policy debate. Often, this process involves new venues of decision-making, such as a change in legislative committee jurisdiction, a shift between branches of government, or movement to another level of government. Issue advocates strategically engage in “venue shopping” by looking for the most advantageous arena in which their issue could be advanced. Sabatier and Jenkins-Smith^[18] in their work on “advocacy coalitions” clarify how issue reconceptualizations of this kind fragment existing policy subsystems while attracting participants from other policy domains.

Birkland’s^[19] consideration of the role of events as triggers in agenda setting encourages attention to all “potential focusing events” to understand those qualities and circumstances that lead to a political response. In general, a focusing event is a rare, sudden, and harmful occurrence which affects a particular locale or social group and receives widespread publicity. But the impact of focusing events is not standard for all issues. Rather, the attributes of the event, as well as the policy domain affected, matter. Whether or not a well-organized policy community exists, the degree of polarization inside that community, and the role of experts all modulate the effect of focusing events in terms of an issue’s placement on the political agenda, the scope of policy change, and the distribution of policy benefits.

The original insight that issue characteristics affect conflict expansion has led to a more thorough identification of issue attributes and their political relevance.^[20] Problems can be defined along many

dimensions, including causality, severity, incidence, novelty, and proximity. The characterization of a problem as a “crisis” is a common ploy by issue advocates seeking to gain rapid agenda access. In social policy making, the description of problem populations—as deserving or undeserving, socially deviant or mainstream, and culpable or innocent—helps predict how much attention is devoted to the problem and which policy instruments are selected. In addition to problems, solutions are defined with respect to their availability, acceptability, and affordability. Numbers pervade the process of problem definition in their use, by experts and advocates, to document the existence of social problems. Numbers can also stimulate fierce political disagreements over measurement techniques and normative standards.^[21] Through force of repetition, numbers become potent symbols in their own right, evoking shared meanings and strong emotions about an issue.

Agenda setting is a two-way process in which opponents seek to contain conflicts that threaten their interests. Cobb and Ross^[22] revisited the idea of conflict containment, subsuming it under the more global dynamic of agenda denial. Their concern is the “non-consideration” of issues, despite social conditions that meet objective standards of seriousness. A narrow resource-based calculation often does little to explain why issue opponents are victorious in these situations. What is needed, instead, is an understanding of how the manipulation of cultural and political symbols can limit the appeal of an issue or cast it as unsuitable for public intervention. Cobb and Ross present a catalog of low-cost, medium-cost, and high-cost strategies used for keeping issues off the formal agenda in this way.

An analysis of the failed health-reform effort of the first Clinton administration shows how agenda setting on the elite and public levels need not be in sync. Hacker^[23] traces how media coverage, shifting public opinion, and responsive public officials together expanded the health care issue and propelled it onto the national agenda. Yet, an intriguing question is why Clinton and his advisers put forward a reform proposal built on the relatively unknown model of “managed competition,” which is a confusing hybrid of regulatory and free-market principles. Hacker finds that managed competition was a compromise meant to find favor with liberals and conservatives while achieving the benefits of a strong public program without “big government.” The notion owed its prominence to an influential network of actors in the health policy, health industry, and media sectors. In the end, however, the value of managed competition as a technical and rhetorical solution proved illusory. Once opponents redefined managed competition by emphasizing its uncertainties, public support quickly

evaporated. This episode demonstrates the danger in agenda setting that policy entrepreneurs will succeed in advancing policy designs that conflict with broad political, institutional, and cultural requirements.

Considering all that has been said, the media’s role in agenda setting is no less complex than it is essential. Media coverage gives an issue visibility, a key ingredient for standing on the public agenda. The media also provide a battleground for groups seeking to increase or limit conflict expansion. Media coverage is, in part, reactive to the established policy agenda, but it is not identical with this group of issues or the concerns of public opinion.^[16,17,24] At times, reporters and editors act as policy entrepreneurs themselves when elevating issues through investigative work or by advancing favored policy positions. Finally, the media contribute to problem definition through the causal frames adopted for story presentation.^[25] With the spread of new media forms such as the world wide web and 24-hour cable news, the possibilities for media influence within agenda setting are only multiplied.

NEGLECTED AREAS OF INQUIRY

Early on, the agenda-setting framework was suggested as a tool for studying comparative politics and government, as well as U.S. policy making.^[26] A focus on agenda setting can highlight how politics differ with regard to the regulation of social conflict, problem definition strategies, the groups involved in policy formation, and the relationship between cultural values and policy choice. Institutional processing of new issues is also an important regime-level variable across nations. Despite a number of studies of agenda setting in contexts other than the U.S.A., truly comparative research applying the model to two or more polities, either in terms of political process or a single policy domain, has been rare.^[27]

Similarly, the analysis of agenda setting in international politics and policy making is a promising subject, although it is difficult because of the multiplicity of decision-making venues for many of these issues.^[28] A conceptually sophisticated investigation of agenda setting in the European Parliament has been produced by Tsebelis,^[29] who modeled the rational preferences of participants under the rules that make it easier to accept than to modify proposals. Tsebelis’s concept of “conditional” agenda setting describes how an unusual set of institutional procedures gives the Parliament an unexpected power over member countries for policy proposals that reach a certain threshold of support, but little influence outside these limits.

To date, the agenda-setting framework has also been used primarily for political situations in which an outside group seeks to gain access to the formal agenda by activating uninvolved groups and the general public. However, “outside initiative” is only one of a set of alternative models featuring different patterns of conflict expansion. Under the “inside access” model, an issue that originates with government officials or a strategically positioned interest group attains the formal agenda directly, without expansion to the public. The “mobilization” model refers to a process in which political leaders initiate an issue or policy and then attempt to gain the public’s support for policy adoption or program implementation. Contemporary policy issues in the U.S.A., such as the war in Iraq and the proposal for social security privatization, suggest the potential value of these latter descriptive schemas.

With respect to methodology, most agenda-setting research has taken the form of qualitative case studies. With the exception of works done by Baumgartner and Jones^[17] and a few others, the framework has not been associated with quantitative empirical analysis. Although the variables involved are many and not always easy to measure, there are aspects of the model pertaining to media coverage, public opinion, and institutional agendas that invite further efforts at quantification.

CONCLUSIONS

At the time it was introduced, the agenda-setting framework supported an understanding of the policy process as a sequence of well-delineated stages, and its applications were generally confined to the beginning of that sequence. Over time, scholars have recognized the relevance of agenda-setting concepts across the entire policy cycle, whenever issues and policies are being prioritized for action or review. Also, it is now evident that policy information flows in more than one direction, creating feedback loops between stages that contribute to problem definition and political mobilization. The agenda-setting framework has shaped a generation of public policy research increasingly sensitive to this fluidity.^[19] Its versatility is confirmed by the variety of policy areas it has illuminated and the rich body of scholarly work it continues to inspire.

ACKNOWLEDGMENTS

Thanks to Roger W. Cobb, Brown University, for comments on this essay.

REFERENCES

1. Mills, C.W. *The Power Elite*; Oxford University Press: New York, 1956.
2. Dahl, R.A. *Who Governs? Democracy and Power in an American City*; Yale University Press: New Haven, CT, 1961.
3. Easton, D. *A Framework for Political Analysis*; Prentice-Hall: Englewood Cliffs, NJ, 1965.
4. Lasswell, H.D. *The Decision Process*; University of Maryland: College Park, MD, 1956.
5. Jones, C.O. *An Introduction to the Study of Public Policy*, 2nd Ed.; Wadsworth: Belmont, CA, 1977.
6. Cobb, R.W.; Elder, C.D. *Participation in American Politics: The Dynamics of Agenda-Building*; Allyn and Bacon: Boston, MA, 1972.
7. Cobb, R.W.; Elder, C.D. *Participation in American Politics: The Dynamics of Agenda-Building*; Revised Ed.; The Johns Hopkins University Press: Baltimore, MD, 1983; 14 pp.
8. Schattschneider, E.E. *The Semi-Sovereign People*; Holt, Rinehart and Winston: New York, 1960.
9. Nelson, B.J. *Making an Issue of Child Abuse*; University of Chicago Press: Chicago, IL, 1984.
10. Cutler, N. Aging and public policy: the politics of agenda-setting. A symposium. *Policy Stud. J.* **1984**, *13*, 111–203.
11. Rochefort, D.A. *American Social Welfare Policy: Dynamics of Formulation and Change*; Westview Press: Boulder, CO, 1986.
12. Bosso, C.J. *Pesticides and Politics: The Life Cycle of a Public Issue*; University of Pittsburgh Press: Pittsburgh, PA, 1987.
13. Sharp, E.B. *The Dilemma of Drug Policy in the United States*; Harper Collins: New York, 1994.
14. McDonough, J.E. *Experiencing Politics: A Legislator's Stories of Government and Health Care*; University of California Press: Berkeley, CA, 2000.
15. Cobb, R.W.; Primo, D.M. *The Plane Truth: Airline Crashes, the Media, and Transportation Policy*; The Brookings Institution: Washington, DC, 2003.
16. Kingdon, J.W. *Agendas, Alternatives, and Public Policies*; Little Brown: Boston, MA, 1984.
17. Baumgartner, F.R.; Jones, B.D. *Agendas and Instability in American Politics*; University of Chicago Press: Chicago, IL, 1993.
18. Sabatier, P.A.; Jenkins-Smith, H.C. The advocacy coalition framework: an assessment. In *Theories of the Policy Process*; Sabatier, P.A., Ed.; Westview Press: Boulder, CO, 1999; 117–166.
19. Birkland, T.A. *After Disaster: Agenda Setting, Public Policy, and Focusing Events*; Georgetown University Press: Washington, D.C., 1997.
20. Rochefort, D.A., Cobb, R.W., Eds.; *The Politics of Problem Definition*; University Press of Kansas: Lawrence, KS, 1994.
21. Stone, D. *Policy Paradox: The Art of Political Decision Making*; Norton: New York, 1988.

22. Cobb, R.W., Ross, M.H., Eds.; *Cultural Strategies of Agenda Denial: Avoidance, Attack, and Redefinition*; University Press of Kansas: Lawrence, KS, 1997.
23. Hacker, J.S. *The Road To Nowhere: The Genesis of President Clinton's Plan for Health Security*; Princeton University Press: Princeton, NJ, 1997.
24. Dearing, J.W.; Rogers, M. *Agenda-Setting*; Sage Publications: Thousand Oaks, CA, 1996.
25. Iyengar, S. *Is Anyone Responsible? How Television Frames Political Issues*; University of Chicago Press: Chicago, IL, 1991.
26. Cobb, R.W.; Ross, J.-K.; Ross, M.H. Agenda building as a comparative political process. *Am. Pol. Sci. Rev.* **1976**, 70 (1), 126–138.
27. Boneparth, E., Stoper, E., Eds.; *Women, Power and Policy: Towards the Year 2000*, 2nd Ed.; Pergamon Press: New York, 1988.
28. Livingston, S.G. The politics of international agenda-setting: reagan and north–south relations. *Int. Stud. Q* **1992**, 36 (3), 313–329.
29. Tsebelis, G. The power of the European Parliament as a conditional agenda setter. *Am. Pol. Sci. Rev.* **1994**, 88 (1), 128–142.

Agricultural Policy

William P. Browne

*Department of Political Science, Central Michigan University,
Mount Pleasant, Michigan, U.S.A.*

INTRODUCTION

Agricultural policy was once synonymous with farm (and ranch) public policy, the actions of U.S. government to modernize and industrialize producers of food and fiber commodities away from subsistence agrarianism. Thus, agricultural policy has been categorized as economic policy, or sector support. Today this is a misleading description. The evolution of modernizing agricultural policy touches myriad domestic social and international affairs matters through a formidable Agricultural Establishment. Issues of hunger, food safety and nutrition, trade, world food needs, environmental degradation, soil and water conservation, farm worker status and protection, rural poverty, and rural development are each institutionally central to contemporary agricultural policy.

POLICY INNOVATION: THE CREATION OF AN AMERICAN AGRICULTURAL ESTABLISHMENT

Public policies to benefit from American food and fiber producers understandably predate the formation of the Republic.^[1] Farmers and what they produced were assets deserving of attention. That should hardly be surprising since both British sovereigns and constitutional lawmakers governed lands dominated and hopefully stabilized by farmers and ranchers once the wilderness trail blazers and adventurers passed through. The early Americas were truly low-density rural societies, characterized by small, family-controlled farms, except in the South where large plantation owners shared space with generally small-scale growers, often tenants.^[2]

Policy debates over improving the status of farmers flourished for 20 years prior to the Civil War; farm prosperity was of obvious value to national growth. Resolution of these debates proved impossible due to national geographic divisions. With the outbreak of war and with a shaken northern electorate, otherwise divided Republicans quickly coalesced to pass popular legislation intended to keep their party in congressional control.^[3] Southern congressional members who had withdrawn from Washington were no longer an obstacle to public policy that assisted small

landowners. In 1862, four significant acts were passed: the first Homestead Act, to provide tracts of federal land to new small-scale settlers; the Morrill Act, which deeded federal lands to the states for use in capitalizing state "land grant" colleges of agriculture and mechanics; the establishment of the U.S. Department of Agriculture (USDA), as a small agency to identify and promote proper seeds and plants to growers who lacked both funds and horticultural knowledge; and the Railroad Act, which subsidized railroad expansion and, incidentally, brought means of transportation for farm supplies to the West and farm products to the East.

Although these distinctive policy innovations were administratively unplanned by politically troubled lawmakers and rarely seen by any contemporaries as an integrated approach to agrarian modernization, they proved to be just that.^[4] New federal and state institutions were being created to serve an ever-expanding cadre of American family farmers. Science and discovery were being put in motion through the creation of public sector experts who could bring applied knowledge to farmers.

Expertise and administrative outreach were facilitated throughout the remainder of the nineteenth century, in no small part because so many farmers were not commercially viable. The states were building their land grant colleges and creating state departments of agriculture to promote farming and ranching. Because so few farmers availed themselves of educational services, many state institutions employed extension administrators to work personally with farmers, usually housing these case workers within county governments.^[5]

The federal government responded similarly. Legislation created federal agricultural experiment stations in 1893, locating them within the states and usually adjacent to land grant colleges. These non-integrated institutions were to engage applied science even further, conducting research specifically on local agricultural problems and local farm failures. The Black land grants were created in 1890, to bring science, education, and outreach to large numbers of unserved southern farmers. The USDA was given full cabinet status in 1889, after intensive farm lobbying. Thus, an ever-growing department could represent its increasing number of programs on an equal footing with those of military, foreign affairs, public lands, and economic

policy makers. Finally, in 1914, the federal government passed an extension act to ensure a national farm outreach policy and join federal agricultural institutions to those of the states and counties in a highly decentralized process.

By World War I, two lasting innovations of general policy importance to the development of modern public administration had occurred. First, an unprecedented network of federal institutions had been created in a *laissez-faire* American government. Big government, fostering expertise and promoting applied science was in place; its institutions provided a *mélange* of policy for federal lawmakers to oversee, fund, reconsider, and recast into the twenty-first century. Second, what public administration scholars refer to as “picket-fence federalism” was in practice on a large scale. The plight of American farmers, chronically plagued by low prices and extensive supplies of their products, became a joint undertaking of all three levels of American government. No level was truly “in charge.” Agricultural specialists emerged at each level, maintained an ongoing dialog with one another, relied on often shared field administrators to demonstrate institutional worth, and provided mutual support both administratively and politically. It was little wonder that, in the 1950s, economist Theodore Shultz looked back at this maze of public institutions and pro-farm sector policies and almost alarmingly labeled it the “Agricultural Establishment.”^[6]

POLICY EFFECTS, POLICY EVOLUTION

The concept of Establishment unwittingly communicates the impression of a favored elite; and, surely, agricultural institutions brought elitism to American farming. Farmers who were better prepared, better funded, and more likely to join in local organizations were also more likely to gain educations for themselves and family members, pay attention to agricultural field administrators, and employ applied science and its technology.^[4] As a consequence, larger-scale farmers gained the most from publicly assisted technological adaptation as early innovators. They then produced greater amounts of commodities, did so at lower costs, and at least for the short term gained profits even with low prices. Smaller-scale, less-positioned farmers did not, bringing poverty, economic failure, and farm exits. Larger-scale neighbors having excess capital assets in turn purchased their lands. The ongoing demise of the subsistence, peasant farmers of America led to the emergence of commercial farm elites who made use of Establishment resources.

Agricultural policy prior to and just after 1900 provided more than field administration services to producers. A regulatory era ensued. Substantive legislation

provided opportunities to those who could and would take them, which again disproportionately meant farm elites. Policies to federally fund irrigation and drainage were established, as were programs for land improvement to facilitate production. Both farmer associations and marketing cooperatives were subsidized. Local farm bureaus were encouraged to form with institutional assistance. Loans and credit were extended to cooperatives. The cooperatives were freed from antitrust laws and assisted in marketing safe products. The net effect was to give agricultural producers special marketing and political status and, more important, lower their cost of production due to federal contributions.

The legacy of such opportunistic policy lingers on. The USDA still supports and regulates cooperatives. Food handlers are inspected. Water projects, loan programs, and farm credit remain consequential components of agricultural policy, in part because of strong state government advocacy. The USDA’s Forest Service provides roads through national forests to subsidize lumber companies. Public policy makers, accordingly, struggle to supervise and reshape the allocation of policy set in motion over 100 years ago. They engage these evaluative struggles with the same institutional supporters who promoted farm assistance in its beginning, agricultural field administrators with their local farm connections and technological interests.^[7] Thus, administrative components of agricultural public policy serve as advocates of both continued and reformulated service and regulatory policy, much as did the many patronage appointees in agriculture in the nineteenth century. While that early generation of agricultural experts capitalized on their political contacts at home and in Washington, modern field administrators more fully embrace the expert guise of home state economists, horticulturalists, agronomists, and nutritionists. The single constant is that a locally and professionally fragmented array of field administrators sets much of the agenda and dominates many of the debates over future agricultural policy.

THE CREATION OF MARKET POLICY: THE MAJOR CONCERN

Agriculture enjoyed several years of good production and high prices in the early twentieth century, through World War I. The transition to commercial agriculture flourished as a result, solidifying cooperation between producers and field administrators. However, commercial producers, often indebted by expansion, were ill prepared for a postwar fall off in demand and the unavoidable collapse of prices. Agricultural depression began in 1920, predating general economic collapse by 10 years. Both agricultural institutions and newly

strong farmer organizations became advocates of direct government intervention in the farm economy, displacing the emphasis on more and better farm services and regulation.

It was not until 1933 that policy makers responded to grassroots demands and moved to directly subsidize farmers by intervening in the marketplace of production and sales. That action, which was intended to raise prices by limiting production, set in motion a complex series of policies that would essentially drive agricultural policy throughout the century. The policy vehicles were the Agricultural Adjustment Acts of 1933 and 1938, along with the Soil Conservation and Domestic Allotment Act of 1936.^[8]

The featured program types that evolved from this legislation included marketing orders, which licensed agreements by commodity cooperatives to hold exclusive sales rights to their sorts of products by negotiating pricing and trade practices; production controls, either mandatory or voluntary, which limited grower planting in return for financial payments; direct supports, payments made by the federal government to landowners, for cutting back production or some other contribution to policy, such as constructing soil conservation projects; and loan guarantees, which provided preproduction loans to farmers at guaranteed rates and allowed producers to voluntarily forfeit crops to government if commodity prices fell below loan rate levels.

Four other consequential and lasting features were introduced into agricultural policy by these programs. First, it was increasingly logical to wrap the many service, regulatory, and market policies of agriculture into one extensive renewable bill, called an omnibus act. These farm bills allowed policy makers to oversee and recast policy in cycles that concentrated attention to agriculture's many details. Second, local (mostly county) and state farmer-operated committees were established to rule on producer eligibility for the previous programs. Third, policies were set by individual commodity, with no general allocation scheme and the omission of numerous products. Fourth, the largest payments and benefits accrued to the largest producers, the most commercially able farmers. Only by rewarding the largest producers would these growers cut significant sums of production and effect prices. Small-scale farmers and their small cuts in production would negligibly influence prices. Of course, this policy rewarded those who had found favor with institutional administrators. Moreover, programs for the rural poor, for small farms, and for tenants were strongly resisted by administrators and given diminished attention in the New Deal and beyond.

Agricultural policy, through 2002, saw a confusing mix of these programs recurring, coming and going, and being constantly readjusted from one omnibus

farm bill to another. Variances in crop prices, partisan politics, congressional control, and interest group strength heavily affected changes. For example, after the House and Senate agricultural committees lost control, Congress passed the Federal Agriculture Improvement and Reform Act (FAIR) as the 1996 farm bill. This act was directed by congressional leaders who had little contact with agricultural professionals.^[3] Their intent was to phase out intervention and all commodity support programs by 2002, letting the marketplace then set prices.^[9] That act provided large direct payments as phase-outs to farmers over a 7-year period; it was later supplemented with increased emergency direct payments from 1998 through 2001. In 2002, with farm prices again depressed, Congress, with the committees once more in control, reversed FAIR and reinstituted commodity programs. A review of price policy effectiveness was scheduled for 2008. Market intervention in farm incomes and pricing thus have been the central but far from only concern of federal policy makers for over three-quarters of a century.

AN EXPANDED AGRICULTURAL POLICY AGENDA

Even if market intervention, or price policy, remains the central concern of federal officials, that focus is difficult to maintain. Trade-offs are necessary. As farm voters diminish in number, as the comparative economic contribution of agriculture shrinks, and as non-farm citizens expect greater federal attention to other issues of food and fiber production, agricultural officials expand their policy agenda. Omnibus farm bills and the already extensive base of agricultural institutions make that possible, through simple accommodation.^[10]

The regulatory era of agricultural policy instituted programs to inspect meat-packing plants, set standards for food processing, and protect against manipulative pricing and food contamination by middle market businesses that brought products to consumers. Current policies extend these efforts and bring federal agencies into a review of emerging citizen worries, such as pesticide contamination and genetically engineered crops. Thus, much of the ongoing work of agricultural policy agencies is evolutionary in detail and politically responsive to shifting public opinion.

As are all public agencies, however, agricultural officials are often charged with having less than total loyalty to new agenda items and non-traditional clientele. Often, new legislation must be initiated to legitimize agency actions and overcome resistance. Omnibus farm bills have been expansive; statutes grew in length from a few dozen pages in the 1960s to over 700 in 1990. These have added important dimensions to what was once economic policy.^[11]

In 1964, the provision of food stamps to low-income Americans for use in making retail purchases was made a “permanent” agricultural agency responsibility. Supplemental assistance to low-income children and mothers followed in later years, as did a renewed USDA emphasis on nutrition. The 1973 Agriculture and Consumer Protection Act responded to twin political realities: inflation and consumers. Basic price program features were remixed to provide more direct payments to farmers, keep food prices low, and expand production with overseas markets in mind. After 1973, both consumer matters and international trade became ever more important components of agricultural policy, particularly in efforts to determine which user preferences increased product sales. Environmental advocates turned considerable attention to the farm bill in 1985, notably in preserving prairies and wetlands. As that bill passed, it was clear that environmental interests had attached themselves to existing soil and water programs and became long-term partners in policy making. Irrigation use, the release of genetically engineered products, safe food, erosion, and groundwater contamination all became ongoing issues. In sum, a variety of social causes, all of which intermingle, are now enmeshed in the economics of commercial farming and food production.

CONCLUSION

Today’s U.S. agricultural policy should be understood as general social policy, taking place in a global economy. In that economy, America is a huge food exporter and gigantic food importer. Agricultural policy in that context struggles with three problems: continued fostering of U.S. commercial production by large-scale farms, developing commercially viable producers of organic and sustainable farms that satisfy non-economic public food wants, and integrating often poorly informed social values about food safety and health into economic policies of production and trade. Integration is the critical challenge: integrating sound trade practices that ensure worldwide environmental

standards in producing safe foods; integrating USDA and other agricultural institutions with the interrelated policies of the Food and Drug Administration and the Environmental Protection Agency; and integrating organic and sustainable producers, as favored by the public, into a dominant system of large-scale producers and into the commercially directed Agricultural Establishment.

REFERENCES

1. Benedict, M.R. *Farm Policies of the United States, 1750–1950*; Twentieth Century Fund, 1953.
2. Cochrane, W. *The Development of American Agriculture: A Historical Analysis*; University of Minnesota Press, 1979.
3. Scheingate, A.D. *The Rise of the Agricultural Welfare State: Institutions and Interest Group Power in the United States, France, and Japan*; Princeton University Press, 2001.
4. Browne, W.P. *The Failure of National Rural Policy: Institutions and Interests*; Georgetown University Press, 2001.
5. Baker, G.L. *The County Agent*; University of Chicago Press, 1939.
6. Schultz, T.W. *The Economic Organization of Agriculture*; McGraw-Hill, 1953.
7. Hallberg, M.C. *Policy for American Agriculture: Choices and Consequences*; Iowa State University Press, 1992.
8. Bowers, D.E.; Rasmussen, W.D.; Baker, G.L. *History of Agricultural Price Support and Adjustment Programs, 1933–84*; Agricultural Information Bulletin; Economic Research Service, U.S. Department of Agriculture: Washington, DC, 1984; Vol. 485.
9. Orden, D.; Paarlberg, R.; Roe, T. *Policy Reform in American Agriculture: Analysis and Prognosis*; University of Chicago Press, 1999.
10. Browne, W.P. *Private Interests, Public Policy, and American Agriculture*; University Press of Kansas, 1988.
11. Browne, W.P. *Cultivating Congress: Constituents, Issues, and Agricultural Policymaking*; University Press of Kansas, 1995.

Alternative Dispute Resolution Processes

Tina Nabatchi

Lisa B. Bingham

*School of Public and Environmental Affairs, Indiana University,
Bloomington, Indiana, U.S.A.*

INTRODUCTION

This chapter explores the basics of alternative dispute resolution (ADR). The chapter begins by defining ADR and examining its use in the public sector. It then discusses several processes within each of the three categories along the ADR continuum: unassisted negotiation, consensus-building, and quasi-adjudication.

WHAT IS ADR AND WHY IS IT USED IN THE PUBLIC SECTOR?

Alternative dispute resolution (ADR), also called appropriate dispute resolution, is an umbrella term for a wide variety of conflict management techniques and processes used in lieu of traditional judicial and administrative adjudication. In general, ADR processes are voluntary, and most use a third party neutral, such as a facilitator, mediator, or arbitrator. ADR processes are designed to resolve disputes in a faster, less expensive, and more amicable manner, and because the processes are less adversarial and formal than traditional litigation, ADR is often able to preserve, and sometimes strengthen, the relationships of the disputing parties. Today, ADR processes are used at all levels of government to resolve a wide variety of public sector disputes (see [Table 1](#) for a list of government ADR web resources).

In the 1990s, several legislative acts incorporated ADR into all three branches of the federal government. The Civil Justice Reform Act of 1990 [28 U.S.C. §§ 471–482 (1994)] and the Alternative Dispute Resolution Act of 1998 [28 U.S.C. §§ 651–658 (Suppl. IV 1998)] brought ADR into the federal judicial branch. Legislative agencies were instructed to use ADR for employment disputes by the Congressional Accountability Act of 1995 [Public Law 104-1]. Executive branch agencies were encouraged to use ADR when Congress passed the Negotiated Rulemaking Act (NRA) of 1990 [5 U.S.C. §§ 561–570] and the Administrative Dispute Resolution Act (ADRA) of 1990 [Public Law 101-552 (codified in scattered

sections of 5 U. S. C. and 9 U.S.C.)], as amendments to the federal Administrative Procedure Act (APA) [5 U.S.C. § 553]. The ADRA and the NRA were amended in 1996, making ADR and negotiated rule-making permanent fixtures in federal agencies.^[1,2]

State agencies derive their authority to use ADR from three sources. First, agencies may infer authority from the state Administrative Procedures Act (APA), which often provides for informal proceedings and the resolution of complaints by settlement.^[3] Second, agencies may imply authority from a general enabling statute that gives them the power to enter into contracts.^[3] Third, some state legislatures have passed statutes expressly authorizing state, and sometimes local, governments to use ADR and/or negotiated rulemaking.

Local governments and municipalities have also exercised their inherent police powers and budgetary and legislative authority to use ADR. Often, they collaborate with local community mediation programs, non-profit organizations with volunteers who are available to mediate typical neighborhood disputes.

Given the proliferation ADR legislation, such processes are now being used to address a wide variety of public sector disputes. ADR processes are used to resolve intra-agency disputes, interagency disputes among government agencies at the same or different levels, and disputes among agencies and private parties. In these contexts, ADR has been used in labor-management disputes, contracting and procurement disputes, regulatory and enforcement disputes, and for claims against the government. The following section of the chapter discusses the broad range of ADR processes that might be used in these different settings.

ADR PROCESSES

ADR processes can be arranged along a continuum ranging from informal, unassisted techniques to more formal adjudicatory arrangements. At one end of the continuum are certain approaches to negotiation; consensual ADR processes involving a third-party neutral fall in the middle of the continuum; and, at the other

Table 1 ADR resources

For resources about ADR use in the federal government, see the Federal ADR Interagency Working Group web site at <http://www.adr.gov>

For resources about ADR use in state government, see the Policy Consensus Initiative (PCI) web site at <http://www.policyconsensus.org>

For resources about ADR use in local government, see the National Association for Community Mediation web site at <http://www.nafcm.org>

For a comprehensive gateway to ADR and conflict resolution resources, see the Conflict Resolution Information Source (CRInfo) web site at <http://www.crinfo.org>

end of the continuum are quasi-adjudicatory processes involving a third-party neutral (see Table 2). Several ADR processes within each of these categories are discussed below.

Unassisted Negotiation

Although they do not require the use of a third-party neutral, certain approaches to negotiation, specifically principled or interest-based negotiation and partnering, fall on the ADR continuum.

Principled or interest-based negotiation

The terms principled negotiation and interest-based bargaining were developed to represent negotiation approaches that stand in contrast to traditional positional bargaining.^[4] Positional bargaining, also called win-lose, competitive, or adversarial negotiation, is relatively confrontational; the disputing parties take sides (positions) and argue in such a manner as to prevail over the other. Conversely, principled or interest-based negotiation is relatively cooperative; the disputing parties seek a solution that is mutually beneficial. There are four steps in principled negotiation:

1) separate people from the problem; 2) focus on interests, not positions; 3) invent options for mutual gain; and 4) use objective criteria to assess possible solutions.^[4] The goal of principled negotiation is to find a solution that will meet the needs and interests of all parties in the dispute.

Partnering

Partnering is a relatively new ADR process developed by agencies for use in the area of procurement and government contracting. Partnering is intended to help avoid the occurrence of disputes by building strong, collaborative working relationships among the contracting parties before disputes arise. The goal is to establish channels of communication that are immediately used at the first sign of a dispute in order to catch and resolve potential problems before they become real disputes and to reduce the need for litigation in the future.^[3]

Consensual Processes Involving a Third-Party Neutral

Consensus-building or consensus-based processes describe a number of collaborative decision-making techniques in which a third-party neutral assists diverse or competing interest groups in reaching an agreement about the issues in conflict.^[5] These processes are typically used to foster dialogue, clarify areas of agreement, improve the information on which decisions are based, and resolve controversial issues in ways that all parties find acceptable. The most common consensus-based processes are conflict assessment or convening, facilitation, mediation, conciliation, negotiated rulemaking, and policy dialogues. In addition, there are processes intended to result in consensus that have adjudicatory elements, such as early neutral evaluation, minitrial, and summary jury trial. These processes supply the disputing parties with an expert

Table 2 The continuum of ADR process

Unassisted negotiation	Consensual process	Quasi-adjudicatory process
Principled or interest-based negotiation	Conflict assessment (convening)	Fact-finding
Partnering	Facilitation	Settlement judges
	Mediation	Private judges
	Conciliation	Arbitration
	Negotiated rulemaking	Med-arb
	Policy dialogues	
	Early neutral evaluation	
	Minitrial	
	Summary jury trial	

opinion about the merits of their case and furnish more information about their best alternative to a negotiated agreement (BATNA). This, in turn, can provide the disputants with a loop-back to negotiation. In all of the consensus building processes, the third party has no power to impose a settlement on the disputing parties, but rather simply aids the parties in reaching an agreement. The differences among these consensus-based processes stem from the third party's degree of activism in the dispute.

Conflict assessment (convening)

Conflict assessment, also known as convening, can be a valuable first step in many ADR processes. The assessment process begins with a discussion among potential stakeholders to evaluate the causes of the conflict and identify the entities and individuals who would be substantively affected by the outcome of the conflict. Next, the interests and needs of the stakeholders are assessed, a preliminary set of relevant issues is identified for discussion, and the feasibility of using various collaborative ADR processes to address the dispute is evaluated. In the final step, often called process design, the neutral recommends and assists in developing an appropriate ADR technique for addressing the dispute.

Facilitation

Facilitation is a collaborative process in which a neutral assists a group of stakeholders in constructively discussing the issues in controversy. The facilitator typically works with participants before and during discussions to assure that appropriate persons are at the table. The facilitator also helps the parties set and enforce the ground rules and agendas, assists the parties in effectively communicating, and helps keep them on track and working toward their goals. Facilitation is commonly used in negotiated rulemaking and to help resolve complex environmental or public disputes; however, the process may work in any number of situations where parties of diverse interests or experience are in discussion.

Mediation

Mediation, one of the oldest forms of conflict resolution and the most common ADR process used in the federal government, is a style of facilitated negotiation, where a skilled, impartial third party assists disputants in reaching a voluntary, mutually agreeable resolution to all or some of the disputed issues.^[5] The mediator works with disputants to help them analyze the conflict, improve communication, identify interests, and explore possibilities for a mutually agreeable resolution. The mediator lacks power to impose any solution;

instead he/she assists the disputants in designing their own solution. Typically, this involves supervising the bargaining, helping the disputants find areas of common ground and understand their alternatives, offering possible solutions, and helping parties draft a final settlement agreement. Mediation usually occurs in the context of a specific dispute involving a limited number of parties; however, mediation procedures are also employed to develop broad policies or regulatory mandates and may involve dozens of participants who represent a variety of interests. Mediation is most often a voluntary process, but court orders or statutes mandate its use in some jurisdictions.

Conciliation

Conciliation involves efforts by a third party to improve the relationship between two or more disputants. Generally, the third party will work with the disputants to correct misunderstandings, reduce fear and distrust, and improve communication. The term conciliation was used in the early 20th century to refer to labor-management mediation, and was later used in Title VII of the Civil Rights Act of 1964 to refer to settlement efforts conducted by an investigator in the Equal Employment Opportunity Commission (EEOC) after reasonable cause to believe that discrimination under the Act has occurred.^[3] Today, conciliation is often used to prepare disputants for a future ADR process, and can also be used as a synonym for mediation.

Negotiated rulemaking

Negotiated rulemaking, also known as regulatory negotiation or reg-neg, involves efforts by regulatory agencies to design regulations by negotiating with interested stakeholders.^[6] In this multiparty process, a negotiating committee consisting of the rulemaking agency and interested stakeholders seeks to reach agreement on the substance of a proposed rule, policy, or standard. The purpose and intent of negotiated rulemaking is to avoid the litigation that may arise to challenge the new rule by generating agreement among the affected interests so that they abide by the decision and its implementation. Federal law requires a thorough conflict assessment before the use of reg-neg and the involvement and assistance of a skilled, neutral mediator or facilitator during the process.

Policy dialog

Policy dialogues are a relatively new form of ADR that are generally used to address complex environmental conflicts or public-policy disputes. In this process, representatives of groups with divergent views or interests are assembled to explore and discuss the

issues in controversy. Unlike many other consensus-based ADR processes, policy dialogues usually do not seek to achieve a full, specific agreement. Rather, participants seek to assess the potential for developing a full consensus resolution at some later time or may put forward general, non-binding recommendations or broad policy preferences for an agency (or other governmental entity) to consider in its subsequent decision making.

Early neutral evaluation

Early neutral evaluation (ENE) is a service often initially performed informally by mediators. In this process, a third-party neutral, usually someone with specifically relevant legal, substantive, or technical expertise, hears informal evidence and arguments from all the parties involved in the dispute and issues a non-binding report advising parties about the strengths and weaknesses of their cases.^[3] The report may also evaluate the likely reaction of a judge or jury if settlement is not reached, provide guidance about appropriate range of outcomes, and assist the parties with narrowing the areas of disagreement or identifying information that may enhance the chances of settlement.

Minitrials and summary jury trials

Minitrials and summary jury trials are commonly used to resolve litigation over complex environmental or public issues. In a minitrial, parties are generally represented by a counsel and an agent with the authority to agree to a settlement or decision, e.g., a CEO or agency official. Abbreviated versions of the evidence and arguments are presented, after which the decision-making representatives attempt to negotiate a settlement.^[3] In a summary jury trial, the disputing parties impanel a jury and present short versions of the evidence and arguments. The jury deliberates and makes findings of fact and liability when appropriate, which are then released by the judge.^[3] The parties are not bound by the jury's findings, but rather use the information to assist with settlement negotiations. Minitrials and summary jury trials are alike in that they both serve as a loop-back to future negotiations.

Quasi-Adjudicatory Process Involving a Third-Party Neutral

There are a broad range of ADR processes that resemble administrative agency adjudication. These processes can be non-binding, with a decision that is advisory only and may serve as a loop-back to negotiation, or binding, with a decision that is final and enforceable by the courts. Some of the most common

quasi-adjudicatory processes are fact-finding, settlement judges, private judges, arbitration, and med-arb.

Fact-finding

In this process, the neutral, called a fact finder, receives information and listens to arguments presented by the disputants. The fact finder, who may conduct additional research to investigate the issues in dispute, evaluates the evidence and submits a report that contains findings of fact and sometimes recommendations based on those findings.^[3] Typically, this informal, non-binding process is used in cases where the disputed facts involve highly technical scientific or engineering issues, thus requiring the fact finder to have subject-matter expertise.

Settlement judges

Settlement judges are used for litigation that has already reached administrative adjudication. In this process, a judge, who is different from the presiding judge in the case, acts as a mediator or neutral evaluator and meets both separately and jointly with the parties to find a mutually agreeable solution.^[3] If the efforts of the settlement judge do not produce full agreement, the case returns to the presiding judge. A settlement judge often plays a more authoritative role than a private mediator, by sometimes providing parties with specific, non-binding legal or substantive information and recommendations.

Private judges

Private judging is similar to both settlement judging and arbitration, and is used for cases that have already reached adjudication. In a private-judging process, sometimes called rent-a-judge, the disputing parties or the courts empower a private person, usually a retired judge or magistrate with special expertise, to hear and decide their case after private proceedings.^[7] The private judge acts as an adjudicator and issues a binding decision based on both fact and law.

Arbitration

Arbitration is a quasi-adjudicatory process where the disputants present their case to an impartial third party, who then issues an opinion. Arbitration may be of rights or interests. Rights arbitration is retrospective; the issues involve an existing contract that one party claims the other has breached. Interest arbitration is prospective; the issues involve the determination of entitlements under some future contract. Arbitration has been used in both unionized and non-unionized labor settings, where it is referred to

as rights arbitration, grievance arbitration, interest arbitration, employment arbitration, and final-offer or baseball arbitration.^[8] The success of arbitration in labor relations has led to its use in commercial settings. Commercial arbitration is used for disputes involving architectural, construction, consumer, and sales contracts, as well as for divorce, environmental, and other disputes.

In all forms arbitration, neither the decision-maker nor the parties are bound by the rules of evidence used in a court of law. However, the hearings themselves generally follow the steps of adjudication: the parties make opening statements; the party with the burden of proof presents its case, then the other party presents a rebuttal; witnesses may be cross-examined and the parties may make closing statements or present briefs arguing their cases.

Med-arb

Med-arb is a relatively new procedure in which a neutral first mediates the case, and if that fails, then goes on to arbitrate the dispute. Sometimes the same neutral mediates and arbitrates the case; at other times different neutrals act as the mediator and arbitrator. In this latter variation, the mediator and arbitrator may or may not be in contact, and the arbitrator may or may not consider the mediator's comments and recommendations.

CONCLUSION

The emergence and proliferation of ADR is one of the most significant movements in U.S. law in the latter half of the 20th century and has had profound effects

on the way the government and public administrators handle conflicts. The institutionalization of ADR in administrative agencies demonstrates both innovation and a concern toward amicable resolution of public disputes. The use of ADR in the public sector will continue to grow as governments recognize the potential of these processes to resolve disputes in a faster, less expensive, and more amicable manner, and as public managers develop expertise in using these processes in decision- and policy-making activities.

REFERENCES

1. Senger, J.M. *Federal Dispute Resolution: Using ADR with the United States Government*; Jossey-Bass: San Francisco, 2003.
2. Breger, M.J.; Lakfer, D.S.; Schatz, G.S. *Federal Administrative Dispute Resolution Deskbook*; American Bar Association: Washington, DC, 2001.
3. Bingham, L.B. Alternative dispute resolution in public administration. In *The Handbook of Public Law and Administration*; Cooper, P.J., Newland, C.A., Eds.; Jossey-Bass: San Francisco, 1997; 546–566.
4. Fisher, R.; Ury, W.; Patton, B. *Getting to Yes*; Penguin Books: New York, 1991.
5. Moore, C. *The Mediation Process*; Jossey-Bass: San Francisco, 1996.
6. Kerwin, C.M. Negotiated rulemaking. In *The Handbook of Public Law and Administration*; Cooper, P.J., Newland, C.A., Eds.; Jossey-Bass: San Francisco, 1997; 225–236.
7. Ponte, L.; Cavenagh, T. *Alternative Dispute Resolution in Business*; West Educational Publishing: Cincinnati, 1999.
8. Elkouri, F.; Elkouri, E.A. *How Arbitration Works*; Bureau of National Affairs, Inc.: Washington, DC, 1985.

Annexation

Jered B. Carr

Department of Political Science, Wayne State University, Detroit, Michigan, U.S.A.

Richard C. Feiock

Askew School of Public Administration and Policy, Florida State University, Tallahassee, Florida, U.S.A.

INTRODUCTION

Annexation is the legal process whereby a municipality adds land, and usually population, to its jurisdiction. In some instances, these actions enable the provision of entirely new services to the territories annexed; in others, they enable higher levels of already-existing services; and yet in others, the level and mix of public services do not change, but the responsibility for providing these services is transferred from another jurisdiction such as a township, a county, or a special district government to a municipal government.

Although annexation involves the extension of the entire institution of municipal government, it is the expansion of city services that typically receives the most attention. The key role service provision plays in annexation cannot be overstated. Annexation is a process that, in many instances, proceeds literally property by property, as services are exchanged for city residence. As the most tangible benefit offered to suburban residents, city services are an inducement used by the city to encourage suburban property owners to join the city. Consequently, annexation issues revolve around service provision and the revenues necessary to pay for them. In some instances, municipal boundary expansion may result in one jurisdiction seizing another's revenue base. When officials from other local governments oppose annexation proposals, it is usually because of the effect of these annexations on their jurisdiction's ability, real or perceived, to fund current and future service levels.

Given the central role of service provision in annexation, the fiscal impacts of these actions are critical. However, current research has shown that the fiscal impacts of annexation are not well understood.^[1] Few local governments have the time and resources to undertake comprehensive fiscal analyses of annexation activities. Moreover, studies show that annexation is not necessarily a "winner-take-all" process.^[1] Annexation can be fiscally advantageous for both, one, or neither community. Conventional wisdom holds that annexation is fiscally beneficial to the

annexing municipality and fiscally disastrous for the government losing the territory. However, in practice, the fiscal impact of annexations is far more complex.

KEY STAKEHOLDERS IN ANNEXATION

Despite the key roles that service provision and financing play in municipal annexation, the racial and income disparities present in communities also influence annexation patterns. Numerous accounts have chronicled the ways in which annexation has been used to intentionally alter the racial composition of cities, and, conversely, how cities have refrained from annexing areas comprised largely of racial minorities or low-income populations.^[2] Importantly, the full effects of municipal annexation on the racial composition of a community may become obvious only when annexations are viewed on a cumulative basis.^[3,4]

Three stakeholders are most often at the center of municipal annexation activity: municipal officials, suburban residents, and the officials of other local governments. In some instances, city residents and state legislators may also become important players in these actions. Whether these stakeholders choose to become involved in a proposed annexation depends on the incentives created by the characteristics (e.g., land area, population demographics, and land use) of the specific parcel considered for annexation, by how revenue sources and functional responsibilities are distributed among the local governments in the area, and by how difficult state annexation laws make their participation in the action. The role played by state laws in affecting the participation of the different stakeholders is discussed in detail in the following section.

Municipal Officials

Municipal officials in cities experiencing population growth and urbanization on their periphery will often view boundary expansion through annexation as

desirable. The thinning and sorting of population that goes on in most urban areas works to diminish the revenue capacity of central cities, while also increasing the demand and need for public services in these cities. The availability of cheaper land, new housing, lower densities, and suburban amenities encourage population dispersion and distribution throughout the metropolitan area. Deterioration of neighborhoods, decreased personal safety, and the widespread apprehension regarding low-income and minority group members all help to push affluent citizens to new residences in areas lying outside the city. Annexation is often seen as an important instrument by which municipal officials can maintain the city's economic base.^[5]

Despite the attractiveness of annexation to municipal officials, whether it makes fiscal sense for a city is often unclear. Annexation may appear to be in the long-term fiscal interests of cities, by expanding the property tax base and increasing revenues from user fees, utility charges, and sales taxes. Yet, in the short term, annexation can result in tremendous capital costs for the extension of services to new residents. Thus, while the annexed area may provide a fiscal surplus, it may also constrain policy options in the short run and affect the ability of a city to undertake other initiatives for years to come.^[6] Even in the longer term, annexation may not prove to be in the fiscal interests of the annexing municipality. For instance, annexation may capture revenues from increased commercial development of the periphery of the municipality, but along with these gains may come increased costs from traffic congestion and infrastructure provision and maintenance.^[1]

State revenue-sharing programs are often an important motivating factor behind municipal expansion, especially where revenue distribution is linked to population or the location of revenue generation. Consequently, city officials may seek to annex territory less for its value in terms of property tax base than for the intergovernmental revenues attached to the area's population and the sales or income tax potential of its businesses.

The extension of municipal boundaries may reduce the costs of service provision borne by current city residents in two different ways. First, the additional customers provided through annexation may lead to lower unit costs produced by economies of scale. Second, annexation expands the base of taxable property that can be tapped to support municipal services. City residents may also see annexation as a way to achieve fiscal equivalence in the wider community. The city's infrastructure and services often directly or indirectly affect residents of bordering unincorporated areas. Thus, annexation can be a means to bring property owners into the municipality

who were previously benefiting from these services at no cost.

However, annexed residents also have the right to participate in municipal elections. Adding a pool of new voters to the city can be controversial, and some city residents may oppose a proposed annexation because of its potential effect on municipal elections. In contrast, race has also been cited as the motivation behind the annexing of White suburban residents. In these instances, city officials have been accused of using annexation to add White population to a city that is nearing a majority-minority status.^[2] Empirical analyses of state annexation patterns have linked reductions in the number of Black elected office-holders to aggressive annexation by communities.^[4]

Suburban Residents

The incentives of suburban residents are less clear than those of city officials. Annexation can be a way to satisfy unmet service demands in these areas and, thus, may be viewed as a means to expand the scope and quality of public services and to increase property values. However, while suburban residents may benefit from the extension of most services and infrastructure, they may instead prefer lower tax and service levels.^[6] Also, suburban residents may resist annexation because they want to maintain the status quo of benefiting from city services and infrastructure at no cost. For example, suburbanites enjoy the benefits of the city's cultural amenities and physical infrastructure at no cost. In addition, some city services have regional effects, such as economic development and urban planning. Annexation can prevent suburban residents from exploiting the central city by imposing marginal costs without full repayment in terms of taxes, fees, or charges.

Even where suburban residents do not seek to "free ride" on city taxpayers, they may still resist annexation because they oppose having to share in the costs of redistributive services that they see as primarily benefiting inner-city residents. Ostensibly, city officials are motivated to add properties that cost less, to support in services, than they generate in tax revenue. This logic is usually apparent to suburban residents as well. Thus, the city must be able to convince them that services through the city will cost less than their other options.

Finally, for a variety of reasons, people within suburban areas and cities sort themselves in many ways. Members of different income groups tend to congregate to form relatively homogeneous communities, both inside and outside of cities. Suburban residents may resist annexation to preserve the "character" of

their community, which they may define in various ways. Numerous anecdotal accounts exist of suburban residents resisting annexation to a city with substantial minority or poor populations. In other instances, opposition may be based on preserving area qualities, such as the rural nature of the neighborhood or low-density development.

Officials from Other Local Governments

The way that officials from local governments outside the annexing city view annexation depends greatly on how functional responsibilities have been divided among the local governments in the community. Conventional wisdom suggests that annexation is a zero-sum game and that the other local governments in the region will always see expansion by neighboring jurisdictions as a threat to their own long-term viability. This is not always the case. In some instances, the annexation of the property displaces a service provider, such as a township, a county, or a special district government. The other local government may strongly oppose annexation to the city because the other government may suffer a loss of revenue from property and/or sales taxes. For example, opposition by county officials is more likely, where county governments have made a significant commitment to providing municipal-level services to the unincorporated areas of the county.^[7] However, in other instances, annexations may involve the acquisition of territory not previously receiving the services in question. The municipal expansion simply adds another layer of government with fundamentally different responsibilities to the property, and no government is displaced. In this situation, other local governments may even support annexation to the city.

STATE STATUTORY REGIMES

State restrictions on local government powers are embodied in Dillon's rule, which holds that local governments are "creatures of the state." As such, local government authority has to be granted by the state constitution or by legislative acts.

For states, the fundamental question is whether annexation is a technical exercise of matching political jurisdictions to service-delivery boundaries, an exercise in political association, or both. State legislatures have generally decided that annexation has elements of both, although they have balanced these two roles in different ways. The different ways that states have resolved this balance has been effectively captured by the Sengstock's classification of annexation laws.^[8] This five-part typology groups state

annexation laws in terms of where the key decision-making authority is placed. It effectively answers the question: Of the three stakeholder groups described in the previous section, which one holds the balance of power to authorize annexations? Generally, each stakeholder has a role in the proceedings, but state law puts one in the favored position where the stakeholder's consent is required for annexation to occur. Given that this diverse set of actors may have different preferences for jurisdictional lines, the inclusion or exclusion of particular actors is expected to have important consequences for annexation activity.^[9]

Regardless of whether it is explicitly stated, all state statutory regimes begin with the presumption that municipalities will generally seek to expand their boundaries. Some states provide municipalities with the authority to expand their boundaries essentially at will, trusting that city officials will not abuse this discretion and expand only where necessary, to ensure efficient and equitable service delivery for the population in the area. Other state legislatures do not display the same level of confidence in municipal governments and require cities to obtain the permission of property owners before expanding. Yet others require that a third party make the decision, apparently trusting neither of the directly affected parties to rise above their own narrow interests on this issue.

Municipal Determination

In these states, municipal governments are authorized to extend their boundaries by the unilateral action of their governing bodies, typically by passing a municipal ordinance annexing the territory. States typically permit unilateral action only if the territory to be annexed meets specific requirements for urban character, such as contiguity, density, or extent of subdivision.^[10] Some states also require that annexation be addressed in a comprehensive plan or overall strategy for annexation. Finally, states that allow unilateral annexation often stipulate that municipalities must produce service plans setting out the timetable for extension of services to the annexed territory.

Popular Determination

Under this approach, the affected electorate or property owners have the right to vote to determine if a municipal boundary change will take place. This approach is rooted in the philosophy that property owners should be able to choose the jurisdiction in which they live, not the other way around. Electors and property owners can grant approval by initiating annexation with a petition signed by a majority of

electors and/or property owners in the territory to be annexed, by referendum or by written permission filed subsequent to a municipality's decision to pass an annexation ordinance or resolution.^[10] In a few cases, a majority of electors and/or property owners have the power to veto an annexation action by petition or by filing written objections.^[10]

Among popular determination states, there are differences with regard to who may vote in annexation referenda. Generally, only the electorate and/or property owners within an area proposed for annexation are allowed to vote on a proposed action. However, in some states, the electors of the municipality may also vote. In a few states, the electors of the diminished territory, typically a county or township government, are also empowered to vote on the action. Referenda commonly require only a simple majority for approval. In cases where the electors of the city are empowered to vote, the two elections and results are usually considered separately, and double majorities are required to approve annexation. Importantly, many states also have provisions that permit annexation to occur without a referendum of property owners. In those instances, where a super-majority of electors or property owners initiate annexation, all referenda requirements are generally waived. Depending on the state, a super-majority of 75% or 100% may be required to avoid the referendum provision.^[10]

Judicial Determination

The final three categories represent different approaches to empowering third parties to approve annexations. The first of these third parties is the judiciary. In some states, courts are empowered to determine if an annexation may proceed, using guidelines and criteria established by the legislature. However, court review is usually limited to procedural or statutory compliance issues. Legislative guidelines usually do not empower courts to assess substantive issues, such as the prudence or equity of annexation, or to consider adverse fiscal impact on other local governments.^[10]

Legislative Determination

The second third party authorized to approve annexations is the state legislature. Because legislatures have the power to create annexation statutes, this is more accurately viewed as the legislature simply retaining this authority. This approach is found most often in states that are highly urbanized and, therefore, possess little remaining unincorporated territory. For example, the states of New England typically do not

have general statutes outlining the process for annexing unincorporated territory, because they lack unincorporated territory. Annexation in these states requires deannexing property from one municipality and annexing it to another.^[10] In addition, in states where annexation procedures are too burdensome or non-existent, municipal boundary changes may be made by special act of the state legislature.

Importantly, this approach provides a forum for other local government officials to participate in these decisions. Although the entire legislature must act on the annexation proposal, typically these local bills will not be put before the entire body until a consensus within the local delegation is achieved. The local delegation is unlikely to act, and therefore annexation will not occur, where there is conflict among the local governments about the proposed annexation.

Quasilegislative Determination

Finally, some states use independent, non-judicial boards or tribunals to determine whether a proposed municipal annexation will occur. This approach provides for a more direct role by other local governments in annexation policy. Most states that use this approach provide for local commissions that consist of various local officials from within a county or group of counties. A few states have state boards with members appointed by the governor.^[10]

PATTERNS OF ANNEXATION ACTIVITY ACROSS THE STATES

State annexation laws affect annexation patterns by providing incentives and disincentives for participation by key stakeholder groups.^[7] The specific procedural constraints outlined in these laws make some options harder to achieve than others, because some provisions necessitate the coordination of greater numbers of people. For example, rules requiring referenda and majority approval by property owners and/or city residents impose collective action barriers to effecting annexation, as proponents must circulate petitions and/or get their supporters to the polls. In contrast, where state law permits boundaries to be extended by the unilateral action of municipal officials, no collective action is required. The need to act collectively to expand municipal boundaries means that those groups better able to organize and sustain these actions will be favored in the process. Consequently, local boundaries will often be drawn to match the preference of those groups that have an advantage under the procedure. [Table 1](#) provides a listing of municipal annexation activity ranked by state.

Table 1 Municipal annexation activity ranked by state, 1990–1999

State	Total actions	State	Population annexed	State	Housing units	State	Area (mi ²) annexed
IL	9,973	NC	315,370	NC	113,490	MT	1,153.96
NC	4,763	TX	312,969	WA	85,525	CA	783.61
FL	4,123	WA	220,663	TX	80,013	TX	776.51
AL	3,816	IN	95,767	IL	40,867	FL	677.61
SC	3,620	IL	95,142	IN	35,093	AZ	608.53
GA	3,411	MS	69,811	TN	34,952	NC	520.03
TN	2,888	MO	68,244	OR	28,060	WA	475.72
TX	2,819	OR	67,941	FL	26,993	IL	417.66
WI	2,792	TN	66,326	MO	26,276	AL	259.97
MO	2,572	CA	54,974	NE	17,451	TN	250.58
OR	2,193	FL	51,036	AZ	16,499	MN	238.78
CA	2,164	AZ	43,670	MS	15,482	IN	223.68
OH	2,026	SC	43,604	CA	14,371	MO	204.90
IN	1,915	NE	43,280	SC	13,621	AR	195.40
CO	1,792	LA	36,516	GA	13,237	GA	193.82
KS	1,499	AL	33,652	CO	13,176	MS	191.34
MN	1,452	GA	32,696	AL	11,669	CO	183.48
WA	1,415	CO	30,757	LA	11,568	OH	174.48
IA	1,190	MN	24,937	OH	8,932	WI	162.40
ID	1,095	OH	22,183	UT	7,979	UT	149.60
UT	1,002	UT	21,147	AR	7,883	AK	138.46
AR	932	AR	20,602	KY	7,469	SC	135.98
KY	931	KS	18,283	ID	6,390	KY	109.16
LA	863	KY	18,152	MN	6,330	SD	107.01
MI	732	ID	16,395	KS	5,833	OK	105.78
NE	716	WI	11,206	WI	4,430	IA	100.80
OK	681	OK	10,347	OK	4,060	NV	95.06
AZ	609	VA	10,111	IA	3,201	KS	89.99
MT	602	IA	8,044	MD	2,907	LA	83.86
NV	460	NV	4,944	MT	2,256	NM	68.13
SD	440	NM	4,600	NM	2,038	OR	62.18
MD	320	SD	4,317	VA	2,037	ID	53.93
ND	302	MI	4,070	SD	1,775	VA	47.58
NM	291	WV	3,964	WY	1,350	NE	38.21
WY	257	MD	3,615	NV	1,332	MI	34.87
WV	180	WY	3,561	WV	1,315	MD	19.33
NY	167	MT	3,497	NY	1,298	WY	18.54
DE	160	NY	3,022	MI	1,274	WV	17.52
MS	84	ND	1,509	ND	822	ND	14.62
VA	77	AK	1,395	DE	605	NY	8.17
AK	21	DE	1,276	AK	443	DE	6.88
PA	10	PA	2	PA	1	PA	0.13
NJ	3	NJ	0	NJ	0	NJ	0
CT	0	CT	0	CT	0	CT	0

(Continued)

Table 1 Municipal annexation activity ranked by state, 1990–1999 (*Continued*)

State	Total actions	State	Population annexed	State	Housing units	State	Area (mi ²) annexed
HI	0	HI	0	HI	0	HI	0
ME	0	ME	0	ME	0	ME	0
MA	0	MA	0	MA	0	MA	0
NH	0	NH	0	NH	0	NH	0
RI	0	RI	0	RI	0	RI	0
VT	0	VT	0	VT	0	VT	0

From State annexation data compiled from the U.S. Census Bureau's Boundary Adjustment Survey.

Annexation as Strategic Action

Despite the widespread usage of the Sengstock typology to classify annexation laws, state annexation regimes often do not fit into a single Sengstock classification but, instead, have elements of two or more different types.^[10] For instance, many states permit annexation to be initiated by municipalities and property owners.^[10] Multiple state procedures create opportunities for stakeholders to manage the scope of conflict. As Schattschneider recognized in 1960, the outcome of any conflict is determined by the extent to which the audience becomes involved in it.^[11] “Every change in the number of participants is about something, that the newcomers have sympathies or antipathies that make it possible to involve them. By definition, the intervening bystanders are not neutral” (on page 4 of Ref.^[11]).

In the past few years, scholars of municipal annexation have begun to recognize the limitations of the Sengstock framework and produce studies focusing, instead, on determining the effects of specific provisions on annexation patterns. In 2006, Facer examined 15 separate provisions of state annexation law on seven different measures of annexation activity.^[12] He found that provisions designed to facilitate annexations were indeed linked to higher levels of annexation activity. Provisions permitting property owners to initiate annexation proceedings, those establishing special processes for preventing or eliminating “unincorporated islands,” and those enabling the annexation of non-contiguous lands had the largest stimulating effects on the level and extent of territory and population annexed.^[12]

However, these studies show that provisions such as requirements for referenda on annexation proposals, which are expected to raise the costs of collective action and therefore reduce annexation activity, instead, appear to increase the likelihood of these actions. Carr and Feiock's 2001 study showed that states with referenda requirements had substantially more annexations than the states not imposing this

constraint.^[7] Similarly, Facer's analysis showed that states using a referendum requirement had almost twice as many annexations as states not requiring a referendum on annexation proposals.^[12] One explanation for these findings is that because many states have created multiple procedures for municipal annexation, those that seek annexation are able to strategically select the arena (i.e., annexation procedure) in which they are most likely to be successful. In other words, they are able to choose the annexation path of least resistance, given their resources and objectives. Most often, the path of least resistance is annexation on a small scale.

For example, many states have created provisions that enable annexations to occur without a referendum if all the property owners in the targeted area agree to the annexation.^[10] These procedures enable municipal officials to exclude city residents, county officials, and all the other suburban property owners in the immediate area from the annexation decision. Consequently, efforts by municipal officials to manage the scope of conflict have led cities to annex literally property owner by property owner. Although large-scale annexations do occur, they are comparatively rare. Not only do annexations effected under these procedures exclude other stakeholders, but also, because of their small scale, these actions go relatively unnoticed in the wider community.

The fact that annexation proponents act strategically underscores the importance of thinking about municipal annexation not only in terms of measures of scale, such as population and territory annexed, but also in terms of the frequency of these actions (Table 2). For years, an overly narrow focus on measures of scale led scholars to mistakenly conclude that the annexation law in a state was far more restrictive than it was in practice. Requirements for referenda were viewed as making annexation nearly impossible; yet, recent studies show that this conclusion is wrong.^[7,14] It appears that many annexations occur through processes that are able to circumvent the more restrictive provisions.

Table 2 Municipal annexation activity per action ranked by state, 1990–1999

State	Average population	State	Average housing units	State	Average area annexed (sq. mi)
MS	831.08	MS	184.31	AK	6.59
WA	155.92	WA	60.44	MS	2.28
VA	131.31	TX	28.38	MT	1.92
TX	111.02	AZ	27.09	AZ	1.00
AZ	71.71	VA	26.45	VA	0.62
AK	66.43	NE	24.37	CA	0.36
NC	66.21	NC	23.83	WA	0.34
NE	60.45	AK	21.10	TX	0.28
IN	50.01	IN	18.33	SD	0.24
LA	42.31	LA	13.40	NM	0.23
OR	30.98	OR	12.80	AR	0.21
MO	26.53	TN	12.10	NV	0.21
CA	25.40	MO	10.22	MN	0.16
TN	22.97	MD	9.08	FL	0.16
AR	22.11	AR	8.46	OK	0.16
WV	22.02	KY	8.02	UT	0.15
UT	21.10	UT	7.96	KY	0.12
KY	19.50	NY	7.77	IN	0.12
NY	18.10	CO	7.35	NC	0.11
MN	17.17	WV	7.31	CO	0.10
CO	17.16	NM	7.00	WV	0.10
NM	15.81	CA	6.64	LA	0.10
OK	15.19	FL	6.55	TN	0.09
ID	14.97	OK	5.96	OH	0.09
WY	13.86	ID	5.84	IA	0.08
FL	12.38	WY	5.25	MO	0.08
KS	12.20	OH	4.41	WY	0.07
SC	12.05	MN	4.36	AL	0.07
MD	11.30	IL	4.10	MD	0.06
OH	10.95	SD	4.03	KS	0.06
NV	10.75	KS	3.89	WI	0.06
SD	9.81	GA	3.88	GA	0.06
GA	9.59	DE	3.78	NE	0.05
IL	9.54	SC	3.76	ID	0.05
AL	8.82	MT	3.75	NY	0.05
DE	7.98	AL	3.06	ND	0.05
IA	6.76	NV	2.90	MI	0.05
MT	5.81	ND	2.72	DE	0.04
MI	5.56	IA	2.69	IL	0.04
ND	5.00	MI	1.74	SC	0.04
WI	4.01	WI	1.59	OR	0.03
PA	0.20	PA	0.10	PA	0.01
NJ	0.00	NJ	0.00	NJ	0.00
CT	0.00	CT	0.00	CT	0.00

(Continued)

Table 2 Municipal annexation activity per action ranked by state, 1990–1999 (*Continued*)

State	Average population	State	Average housing units	State	Average area annexed (sq. mi)
HI	0.00	HI	0.00	HI	0.00
ME	0.00	ME	0.00	ME	0.00
MA	0.00	MA	0.00	MA	0.00
NH	0.00	NH	0.00	NH	0.00
RI	0.00	RI	0.00	RI	0.00
VT	0.00	VT	0.00	VT	0.00

From State annexation data compiled from the U.S. Census Bureau's Boundary Adjustment Survey.

MUNICIPAL ANNEXATION RESEARCH: KEY QUESTIONS FOR THE TWENTY-FIRST CENTURY

Three aspects of annexation policy are likely to receive significant attention in the future. First, all the actors affected by annexation, whether city officials, suburban residents, or interested third parties, have various mechanisms at their disposal to achieve their service-delivery objectives. In previous years, suburban residents could obtain municipal-level services only by joining a municipal government, either through annexation to an existing municipality or through the formation of a new city government. Many states now empower counties and special districts to offer services directly competing with municipal governments.^[15] Similarly, municipal governments once had to annex suburban properties to provide services to them, but now many states permit municipal governments to provide services to properties beyond their borders.^[13] Others permit one or more local governments to enter into agreements that effectively transfer land from one unit to another for a specified period of time, thereby obviating the subject of annexation.^[14] Thus, studies of annexation policy must recognize that the benefits of annexation may be met in a variety of ways and conceptualize annexation as one instrument among many.

Second, it is likely that scholars of annexation will produce better and better analyses of the consequences of state annexation laws on the patterns of annexation observed in the states. As already noted, annexation policy is a complex topic involving multiple state laws and strategic behavior by participants. Policymakers and academics alike have an important stake in understanding how key provisions such as referendum requirements, service plans, and the use of boundary commissions effect annexation patterns.

Finally, concerns over the effects of urban sprawl on the fiscal health of central cities have led to renewed calls for regional government in our urban areas. Advocates of annexation, such as Rusk, argue that cities able to annex suburb properties are fiscally stronger than

those cities unable to expand their boundaries.^[16] They contend that annexing land is an important route to economic health and development of urban areas. However, for much of the latter part of the 20th century, many state legislatures created annexation laws giving property owners greater control over annexation decisions.^[10] This trend coincided with a more general movement, aimed at providing suburban property owners with more choices in service provision. The intense debate in the United States over the appropriate role of annexation policy in the fiscal health of our urban communities will likely continue in the 21st century.

REFERENCES

1. Edwards, M. Annexation: a 'winner-take-all' process? *State Local Gov. Rev.* **1999**, 31 (3), 221–231.
2. Murphy, T.P. Metropolitan reorganization and black political power. *Urban Interest* **1979**, 1 (2), 29–34.
3. Feiock, R.C.; Carr, J.B. *African Americans and the Political Economy of Municipal Annexation*; American Political Science Association: San Francisco, 1996.
4. Feiock, R.C.; Carr, J.B. *Lines and Color: The Role of Race in Local Government Annexation Decisions*; American Political Science Association: Washington, DC, 1997.
5. Liner, G.; McGregor, R.R. Institutions and the market for annexable land. *Growth Change* **1996**, 27, 55–74.
6. Clingermyer, J.C.; Feiock, R.C. *Institutional Constraints and Policy Choice: An Exploration of Local Governance*; State University of New York Press: Albany, 2001.
7. Carr, J.B.; Feiock, R.C. State annexation 'constraints' and the frequency of municipal annexation. *Polit. Res. Q.* **2001**, 54 (2), 459–470.
8. Sengstock, F.S. *Annexation: A Solution to the Metropolitan Area Problem*; University of Michigan Law School: Ann Arbor, 1960.
9. Feiock, R.C.; Carr, J.B. Incentives, entrepreneurs, and boundary change: a collective action framework. *Urban Aff. Rev.* **2001**, 36 (3), 382–405.
10. Palmer, J.; Lindsey, G. Classifying state approaches to annexation. *State Local Gov. Rev.* **2001**, 33 (1), 60–73.

11. Schattschneider, E.E. *The Semisovereign People: A Realist's View of Democracy in America*; Holt, Rinehart and Winston: Fort Worth, 1960.
12. Facer, R. Annexation activity and state law in the United States. *Urban Aff. Rev.* **2006**, *41* (5), 697–709.
13. Krane, D.; Rigos, P.; Hill, M. *Home Rule in America: A Fifty State Handbook*; CQ Press: Washington, DC, 2001.
14. Bassett, E. “This land is *our* land?” An analysis of land-use planning and cooperation under Michigan’s conditional land transfer act. *State Local Gov. Rev.* **2006**, *38* (1), 23–33.
15. Carr, J.B. Whose game do we play? Local government boundary change and metropolitan governance. In *Metropolitan Governance: Conflict, Competition, and Cooperation*; Feiock, R.C., Ed.; Georgetown University Press: Washington, DC, 2004; pp. 212–239.
16. Rusk, D. *Cities Without Suburbs*, 3rd Ed.; Johns Hopkins University Press: Baltimore, MD, 2003.

Appearance of Impropriety

Stuart C. Gilman

Ethics Resource Center, Washington, District of Columbia, U.S.A.

INTRODUCTION

One of the most controversial concepts in public service ethics is the notion of the appearance of impropriety. Whereas few would question the evil of impropriety in carrying out one's public duties, many have great difficulty with the modifier "appearance." After all, some might argue, cannot any action be falsely perceived? Does not the appearance standard leave anyone open to attacks by the most excessive elements of the community? Is not appearance an excuse for subjective extremism? Even if all these questions are answered affirmatively, reality for those who serve the public is that the appearance of impropriety is a critical standard by which to evaluate public officials. Yet, at the end of the day, it might be the most important standard by which to judge the actions of public officials.

Appearance standards are not limited to public administrators, but also extend to legislators and judges. In fact, many professional associations, such as the American Society for Public Administration, the International City/County Managers Association, and the Association of Government Accountants, have codes that have appearance as one element. Recently, the appearance of impropriety has even been raised for private-sector companies whose multiple business lines seem to conflict with each other (e.g., public auditing companies and their consulting firms).

Appearance standards are all derived from the English Common Law Standard of the Common Man Rule, or in its modern incarnation the Prudent Person Rule:

Historically known as the prudent or reasonable man rule, this standard does not mandate an individual to possess exceptional or uncanny investment skill. It requires only that a fiduciary exercise discretion and average intelligence in making investments that would be generally acceptable as sound.^[1]

The Prudent Person Rule provides a model test: Would a reasonable person armed with all the facts conclude that an action or inaction was inappropriate?

TOWARD A DEFINITION

Appearance of a conflict of interest, appearance of the loss of impartiality, and appearance of impropriety

often are used interchangeably. Although some might claim that it is valuable to refine the differences among them, for our purposes these concepts are synonymous. A claim can be made that conflicts of interest are a subset of impropriety; however, in applied, practical cases it is almost impossible to support the distinction. Andrew Stark has argued that appearance of impropriety can either take "a factual or a normative caste (or sometimes both). Either (a) it *looks* like the official did something wrong or (b) the official did something that *looks* wrong" (Ref.^[2], p. 328). In other words, an official's action is questioned; however, no law was broken and no broadly agreed-on mores were violated. Instead, people feel there should have been a law against the action (or result), or it should be defined as an actually improper action (or result).

On a number of levels such judgments seem unfair. They are prejudiced (decided before the facts are actually known) and also seem to smack of ex post facto laws (laws made after the action). Following Stark's argument, in civil matters such standards of legal fairness are excessive because they assume protections inherent in criminal prosecutions (e.g., innocence until proven guilty). In the vast majority of cases, appearance issues do not rise to criminal standards. Most often, they are punished either politically or administratively, with the most extreme penalty being removal from office. In addition, we assume that public officials will be held to supererogatory standards, thereby justifying the public's trust.

Another argument against an appearance standard is that it rests on knowing one's intention. Former Senator Alan Cranston, reprimanded by the U.S. Senate for his role in the Keating Five Case, claimed that no one knew what he was thinking when he acted.^[3] The only real person who can judge is the senator himself. "You were not there. I was. And I know that what I knew at the time ... convinced me that my [actions] were appropriate."^[4] However, there are always mixed motives in anyone's actions. Simply saying the behaviors were not viewed as intended is not enough. Therefore, as Dennis Thompson pointed out, "because appearances are often the only window that citizens have on official conduct, rejecting the appearance standard is tantamount to denying democratic accountability."^[4]

In a democracy, it seems reasonable to hold public officials to standards that average citizens might not apply for their own behavior. An official has an obligation to do the “right thing” and also to sustain the confidence of the people in governmental institutions. It is important that officials not do one or the other, but do both. This is the double standard to which most democracies hold their officials. It might be argued that this is true of all types of government. Because information is arguably the one most jealously guarded resource in non-democratic societies, it is almost impossible for citizens to find out about impropriety, much less distinguish between real impropriety and its appearance.

WHY IS “APPEARANCE” A PROBLEM?

Unlike other unethical actions, “appearance” issues seem to violate the usual regulatory standards of clarity and objectivity. However, if we understand appearance not in its juridical form, but as a way of clarifying values, it seems to take on a wholly different character. Where a law, regulation, or rule would emphasize a “bright-line” test of whether you violated the precept, a values approach requires balancing various principles and interests. So, for example, under a regulatory regime a senior official might be limited to a \$25 gift each year from a single source. If an individual receives 1000 gifts worth \$25 or less, it would not violate the rule. However, a reasonable claim can be made that the individual would have violated the *spirit* of the rule. In this sense, the notion of spirit is the view that there are essential values that are a predicate to the rule itself. In this case, individuals seem to be intentionally exploiting the rule for their own benefit and, even more egregiously, consciously circumventing the letter of the law in a way that seems to be purposely abusive.

An additional dimension of appearance is that it is dynamic, often varying with historical tides, scandals, and changing societal mores. For example, the press and public often ignored sexual peccadilloes in the nineteenth century when “affairs of the heart” seemed to be part of the privilege of the ruling class. Well into the twentieth century, both Franklin Roosevelt and Dwight Eisenhower were purported to have long-term relationships with mistresses. Today, the slightest hint of such a relationship would be front-page news.

Other issues that now seem to be clear impropriety, in other historical contexts, seemed to be a mere appearance problem. As late as the early twentieth century, George Washington Plunkitt could claim that the Tammany Hall Democrats were more ethical than the Philadelphia Republicans. Although the Republicans had stolen the zinc roofs from poor houses to enrich themselves, the Democrats only took advantage of

buying land (and making a profit) when they knew some public project, such as a building or bridge, might be built on it. Plunkitt used these examples to distinguish between *honest* and *dishonest* graft.^[5]

The impact of these actions, even if legal at the time, can have a devastating impact on the public confidence in government. Public confidence is always fragile in democracies. Inevitably, revolutions, military coups, and other methods of destroying democratic institutions are justified by the lack of confidence in public institutions. In this way, a reasonable claim can be made that appearance of impropriety can have a far more pernicious impact, because collectively a series of unrelated appearance issues can undermine the most robust democratic institutions. This “death by a thousand cuts” by the appearance of impropriety can lay claim to the destruction of untold numbers of political regimes since the 1800s.

GENERAL RULES

Notions of “conduct unbecoming a civil servant” or “conduct that leads to embarrassment of the service” are some of the more general rules used by government. Without mentioning the word appearance, these general rules can often give wide latitude when it comes to enforcement against actions that are perceived as unethical. However, they are often criticized as overly broad and as license to carry out political or personal vendettas against an individual or a group of individuals. In Great Britain, the Nolan Committee on Standards in Public Life created a series of standards based on either “actual or apparent” violations. These broad principles of behavior expected from government officials include *selflessness, integrity, objectivity, accountability, openness, honesty, and leadership*. Although these are aspirational in terms of any enforcement capability of the committee, they are capable of being administratively enforced within ministries or departments.

In the United States, the courts have supported agency actions against employees for activities that appear to be improper, yet do not violate a specific law. One of the clearest examples comes from the federal court case *Wild v. U.S.*^[6] At the time, the executive branch was operating under the Lyndon Johnson’s *Executive Order 11222* (1965), which allowed each department or agency to develop its own specific standards of conduct, based on the model standards found in the executive order. The U.S. Department of Housing and Urban Development (HUD) had such a set of standards that forbid employees from undertaking activities that could appear to conflict with their jobs.

Lawrence Wild was an employee of HUD who had been dismissed for violating HUD’s code of conduct.

Among other elements, the code prohibited the appearance of official impropriety. Wild had violated this norm through his ownership of private rental units. These were so badly deteriorated that local newspapers noted his official position as a manager with HUD and as a landlord of slum dwellings:

In his defense, Wild had argued that as a matter of fact, the newspaper's impressions were inaccurate: he had either demolished or rehabilitated that vast majority of units in question, and had plans to do the same with the remainder. And in any case, Wild urged, as a normative proposition the appearance standard is "so vague as to fail to provide adequate notice" as to what kind of "off-duty conduct places [an official's] job on the line" (Ref.^[2], pp. 326–327).

The court recognized that the point of the appearance standard is to make certain conduct, not forbidden by regulation, a grounds for removal from federal service.

The appeals court went further in its opinion by arguing that whether an employee actually engaged in misconduct is not as important as whether *he could be perceived of having done so*. The appearance norm itself "adequately conveys its concern that an employee not conduct himself in a way likely to bring public obloquy on HUD."^[6] As the court concluded, the irony of Wild moonlighting as a slum lord while serving as a professional employee of HUD should have allowed him to conclude that his actions would have reduced the public's confidence in the department.

This case illustrates both the appearance of impropriety and some of the appearance concept's fundamental problems. The case allows us to distinguish between actual behavior and the perception of behavior. The perception that an official of an agency committed to upgrading the quality of housing bought slum dwellings was enough for the court to argue that HUD had legitimately fired Wild. Whether his personal actions were ethical, the mantle he carried as a HUD official required him to consider not only his own personal actions, but also the impact of those actions on the public's confidence in his agency's work.

Even if one accepts this as a reasonable line of argument, it is obvious that there are extraordinary impositions on an official's civil liberties and personal freedom. Many government officials are attracted to their agencies because of a passionate interest a particular policy arena: environment, forestry, children's health, etc. One can easily imagine the many circumstances that would lead to appearance issues. Yet, the potential of apparent conflicts of interest might prevent a public employee from buying stock in a particular pharmaceutical company that deals with children's health issues, purchasing public forest land, or even

accepting payment for giving speeches against the current government air pollution policies.

As with all exercises in clarifying values, the dilemma with the appearance standard is that there is a constant balancing of ethical principles. One can imagine cases where the upholding of a fundamental value trumps the issue of appearance. In a hypothetical example, an official responsible for disease prevention in prisons might be an advocate for provision of contraceptives to prevent AIDS. This disease is rampant in the prison system, with hundreds of inmates infected and dozens dying each year. In this particular state, state law bans purchase of such contraceptives by the state. The employee could create a non-profit organization to provide contraceptives free of charge and become a spokesperson for that group. Perhaps a newspaper accuses this employee of coercing inmates into using contraceptives forbidden by the state and therefore encouraging homosexual sex among inmates. Under the Wild standard, a judge could find it reasonable to fire this person because it would "bring public obloquy" on the prison system.

SPECIFIC RULES

Attempting to clarify the concept of appearance of impropriety, numerous governments codified the concept in the 1990s. Although this is true for a number of federal, state, and local governments, as well as Canada and Australia, the paradigm case is the executive branch of the U.S. federal government. In 1989, then President George H. W. Bush issued *Executive Order 12674*, which articulated 14 general principles of government conduct, including "employees shall endeavor to avoid any actions creating the appearance that they are violating the law or the ethical standards promulgated pursuant to this order."^[7] In the executive order, President Bush further ordered the U.S. Office of Government Ethics (OGE) to promulgate regulations to clarify and allow for enforcement of these standards, including appearance.

The regulations appear in Subpart E of the regulations (5 CFR 2635.501–503) entitled *Impartiality in Performing Public Duties*.^[8] There are two detailed provisions that deal with different aspects of avoiding "an appearance of a loss of impartiality." The first issue focuses on appearance issues that would be created by personal or business relationships. For this, the OGE used the traditional English Common Law Standard (5CFR2635.501) "an employee should not participate . . . if he determines that a reasonable person with knowledge of the relevant facts would question his impartiality in the matter." The guidance actually provides a "test" (2635.502) for an employee who is unsure of how to answer the question of appearance

by detailing types of relationships, as well as types of actions that would constitute an appearance of impropriety. This expands the prohibition against actual conflicts of interest in the criminal statutes (18 USC 208) that forbids actions that would benefit an employee, his or her spouse, or their minor children. In effect, this regulation extends the appearance of a conflict of interest to any member of the employee's household or someone with whom the employee has a personal relationship (e.g., parent, lover, uncle, sister, companion, personal friend). Interestingly, it also allows an independent source (authorization by an agency designee) to use an additional series of factors related to the impact on the agency to make an appearance determination. For the executive branch of the federal government, this determination would be usually made by the ethics official in the agency.^[9]

The regulation also provides several vignettes to illustrate how to apply the regulation, for example:

An employee of the Internal Revenue Service is a member of a private organization whose purpose is to restore a Victorian-era railroad station and she chairs its annual fund raising drive. Under the circumstance, the employee would be correct in concluding that her active membership in the organization would be likely to cause a reasonable person to question her impartiality if she were to participate in an IRS determination regarding the tax-exempt status of the organization.

What the regulation attempts to do is to provide some guidance to evaluate the circumstance oneself. If that does not resolve the question, the regulation provides a set of standards through which an independent, third party can make a determination.

The second part of the regulation deals with the appearance problem of receiving extraordinary payments (defined as that is \$10,000 or more in value) from a private organization when joining the government. In effect, there is a 2-year disqualification (from the date of receipt of the payment) from dealing with any specific matter before the government in which a prior employer has an interest. Unlike the first section of the appearance standard that involves various tests and balancing of interests, this second standard suggests a clear-cut standard of appearance.

THE FIDUCIARY RESPONSIBILITY

It is worthwhile noting that appearance of impropriety or losses of impartiality are not isolated to the public service. There are many notable examples in the non-government arena where chief executive officers, as well as others, have been fired because of acts that

appeared improper or unethical. Indeed, there is a fiduciary responsibility in publicly traded companies to stockholders and other stakeholders. However, there is a significantly broader sense of fiduciary responsibility for public servants, and this is often ignored when one analyzes appearance problems in public service.

There are several dimensions to the breadth of fiduciary responsibilities for public servants, including the constitutional undergirding of public service, the limits to privacy, and the growth of transparency. Most constitutional regimes put extraordinary constitutional obligations on public employees. For instance, in the United States, all public employees (including the vast majority of state and local officials) take an oath to uphold the Constitution. They do not swear to support their immediate supervisor, the head of the agency, or the president. Rather, they swear to uphold a document that begins "We the people . . ." Although this oath takes many forms, it is worthwhile noting the similarity in most ethics regimes as reported to the Organization for Economic Cooperation and Development (OECD).^[10]

Although privacy concerns are a growing arena for constitutional debate in many countries, one of the few areas of contraction can be found in terms of public servants. In the United States, the use of public financial disclosure is an obvious example of the importance of appearance over the value of privacy for public servants. Currently, the OGE accounts for more than 20,000 public financial disclosure forms in the executive branch of the federal government alone. Because ethics officials review these forms and actions to eliminate real or apparent conflicts of interest, they must be taken before they are permanently filed. Available to anyone on request, the primary purpose of the forms is to eliminate any appearance problem. The growth of disclosure systems since the early 1990s, as noted in the OECD report mentioned previously, is nothing short of remarkable. This is true worldwide, including countries as diverse as Argentina, South Africa, and Slovakia. It is also true in most states in the United States and provinces in Canada.^[11]

Some of these systems are confidential and not public. Confidential systems also rely on the citizen's faith in the independence and objectivity of offices reviewing such disclosures. The Canadian Ethics Counsellor is responsible for reviewing financial disclosure and eliminating conflicts.^[12] In a recent address, the Ethics Counsellor, Howard Wilson, commented that Canadians had faith in his office, and the Canadian people's respect for privacy would never put up with a public disclosure system for their public officials. One wag responded that the Canadian system was simply one scandal away from public disclosure.

Transparency, or the general openness of government, is another major force expanding fiduciary

responsibility. From procurement to contracting, and from hiring personnel to making major policy decisions, government transparency is designed to not only eliminate impropriety, but also to eradicate the appearance of impropriety. Perhaps the best legal example of this is the Freedom of Information Act. Although many in the media complain about redacted or denied documents, the law places severe limitations on what an agency can keep out of public scrutiny. Generic requests can place overwhelming claims on an agency's personnel and time. For example, one request from a conservative non-government organization for any document with the name "Hillary Rodham Clinton" required a search of every paper file, electronic file, and e-mail. The search, gathering, review, and release, all under legal supervision, took hundreds of personnel hours of labor for what was little more than a fishing expedition. Whatever one thinks of such processes, they are a stark testament to the broad fiduciary responsibilities to which we hold public servants.

CONCLUSION

The appearance of impropriety is both a vexing and illuminating concept. It is vexing because, although it appeals to fundamental values, it also undermines fairness. It is illuminating because it points to the extraordinary set of responsibilities of public servants, and the growing expectations around those responsibilities held by the public. In a sense, the appearance of impropriety is the leading edge of the organic expansion of public service ethics. Impropriety acts as a sort of tidewater mark, to which laws and regulations ultimately rise. Dennis Thompson called this phenomena mediated political corruption.^[4] Appearance of impropriety rises to inform and define the definition of political corruption. For this reason, understanding the shape and volatility of the appearance of impropriety is vital for political leaders and the public service, in general.

As Judge Frank Nebeker, then Director of the U.S. Office of Government Ethics, wrote in an opinion excoriating former Reagan Attorney General Ed Meese:

Avoiding criminal conduct is not the mark of public service. The duties imposed by non-criminal standards are far harder to discharge. They may even be strange and seem overly restrictive to some joining the government for the first time ... [But] the vast majority of officers and employees are hard working and loyal individuals who make every effort to adhere to the

high ethical standards the public has a right to expect from them and that they expect from themselves.^[13]

Therefore, although the appearance standard has a legal foundation, only emphasizing jurisprudence would ensure that we would miss its political weight and importance. The appearance standard is also grounded in political values and is an expression of democratic accountability. The basis of appearance is the notion that civic employees are servants of the public, and therefore must do everything in their power to assure citizens that government officials are working for the *public good*.

REFERENCES

1. Prudent person rule. In *West Encyclopedia of American Law*; Farmington Hills, March 1998.
2. Stark, A. The appearance of official impropriety and the concept of political crime. *Ethics* **January 1995**, 105.
3. The Keating Five case involved accusation against five U.S. Senators had illicitly used their influence to help Charles Keating, a savings and loan financier whose default significantly contributed to the Savings-and-loan disaster of the 1980s.
4. Thompson, D. Mediated political corruption: the case of the Keating Five. *Am. Polit. Sci. Rev.* **June 1993**, 87 (2), 376.
5. Riordan, W. *Plunkitt of Tammany Hall: A Series of Very Plain Talks on Very Practical Politics*; Putnam: New York, 1963. (Originally published in 1905).
6. *Wild v. U.S.* 692 F. 2nd 112; 1982.
7. George, H.W. Bush Administration. *Executive Order 12674*; 1989.
8. Although the Bush administration promulgated the initial executive order, two years later when asked to review the regulations balked at the title "Appearance of Impropriety" and instead compromised on the word "Impartiality."
9. In the administrative structure of the ethics system in the executive branch of the U.S. government, each department and agency must (under the Ethics in Government Act of 1978) appoint a Designated Agency Ethics Official (DAEO) who works conjointly with the U.S. Office of Government Ethics. In turn, many DAEOs have large staffs that provide guidance to employees. See www.usoge.gov for a detailed description of this program.
10. OECD. *Trust in Government*; PUMA: Paris, 2000.
11. See www.cogel.org.
12. Canadian Ethics Counsellor. Available at: <http://strategis.ic.gc.ca/SSG/oe00001e.html>.
13. 88X1. *Informal Advisory Opinions of the U.S. Office of Government Ethics*; Washington, DC.

Appreciative Inquiry

Thomas W. Kent

Department of Management and Entrepreneurship, College of Charleston,
Charleston, South Carolina, U.S.A.

INTRODUCTION

This article describes the nature of appreciative inquiry (AI) and what is involved in it. It is, essentially, an organizational intervention that is based on the premise that every organization has strengths and a history to be proud of. The process brings those strengths and proud moments to the fore, as a way to help regenerate people's optimism and hope, as they enter a significant planning process. The article describes this planning process in overview. For an in-depth description of the four stages, the reader should see the resources mentioned at the end of the article.

OVERVIEW OF AI

Much of the impetus for organizational change comes from negative forces. We do not like the way something works and we want to change it, we see problems that constantly reoccur and we want to fix it once and for all, competitors are eating our lunch and we need to do something, and we have personnel issues that are vexing and must be resolved. These are some of the drivers of organizational change as well as examples of negative forces.

It is arguable, but a solid case can be made that this is not the best starting point for a successful future. Whitney and Trosten-Bloom, for example, suggest that when you start from the negative, you set the stage for raising employees' doubts and fears and for increasing resistance to change.^[1] Frequently, as a result, the leaders of the change effort end up having to force the change through the organization, while losing employee commitment and cooperation as they do. In addition to that, working from the negative establishes a "fix it" mentality. When we fix something, we generally like to repair it back to something close to its original condition or working order. Or, we try to patch over a problem so that the situation is workable. Neither of these latter processes really enables organizational improvement. Let us look at a more positive approach to organizational improvement or transformation.

AI has been around as an organizational intervention for about 20 yr. It has been used in all kinds

of organizations—large and small, governmental, sales, media, community councils, and high tech.^[2–7]

Some of the most significant work using AI has occurred in community settings. "Imagine Chicago" may be the best example.^[8] This was a 10-yr effort, designed to create citywide learning communities and economic relationships that promote justice vs. deepen the divide between rich and poor. It began in 1992 and continues today, involving thousands of the citizens of Chicago. A small design team and 50 interviewers marked the beginning of the project, which eventually involved a wide range of individuals and institutions, including grassroots leaders, public schools, teachers, immigrant and faith communities, school children, and parents. This project took place in three stages: *understanding* what exists by focusing on the best of what exists; *imagining* what can be; and *creating* what will be. Below is a more descriptive account of the usual AI steps.

AI starts from the basic belief that every organization has identifiable strengths and assets that can serve as the starting point for change. With a clear picture of these strengths and assets in mind, the process creates a positive vision of the future—one that draws employees to it vs. something negative they might run from. This idea is like the planning methods of Ackoff when he suggests doing "ends planning" by creating an idealized design of the organization and then planning your way to that end.^[9] In practice, the AI process takes several days and involves all stakeholders. The process can involve many people, some times hundreds and, on occasion, thousands.^[10]

Cooperrider et al. call an organization's strengths and assets the "life-giving forces" within the organization.^[11] He believes that the job of the organization's leaders is to create an alignment of positive forces, so that the negative aspects of the organization become irrelevant.^[12] The AI process seeks to identify these forces and use the energy they engender to ignite the collective imagination of organization members to help them dream about what could be in their future as an organization.

The process involves four steps or phases: discovery; dreaming; designing; and destiny. This is what Cooperrider calls the "4-D" cycle (Fig. 1).

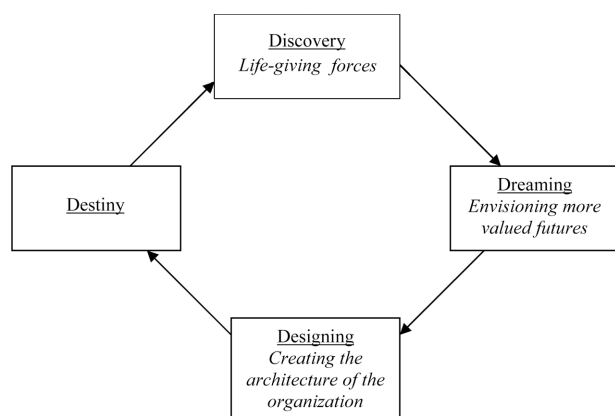


Fig. 1 Four stages of the appreciative inquiry change process.

DISCOVERY

The Discovery phase (or the *understanding* stage in the “Imagine Chicago” project) seeks to identify the best of what currently exists within the organization/community. This process generates positive excitement among the participants, and this allows them to envision new possibilities and new actions. The participants of the AI process are asked two questions:

- What factors give life to this organization when it is and has been most alive, successful, and effective?
- What possibilities, expressed and latent, provide opportunities for more vital, successful, and effective (vision-and-values congruent) forms of organization? (on page 32 of Ref.^[9])

During this phase, participants share their experiences and memories of significant accomplishments, momentous achievements, meaningful relationships, and other memories that they most value about the organization that they want to see sustained, enabled, and brought into the future. During this stage, people “appreciate” their organization more, and they are drawn together in an atmosphere of hope and anticipation.

Out of this process, a topic or set of topics is selected for further discussion and planning. The topics are selected by a microcosm of the organization—people from all functions and levels of the organization or community. Customers, clients, citizens, suppliers, and others may be a part of this microcosm. This team serves as a “steering committee” for the process and part of their role is to select the topics that will be carried forward through the process. It is important that all members of this team have an equal voice. School principles, managers, community leaders, and other power figures must not control the discussions.

DREAMING

The first stage—the discovery stage—naturally leads to images of what might be possible in the future. In the “Imagine Chicago” project, this stage was called *imagining*. This is the objective of the dreaming stage—to imagine a compelling future that is also practical. It is practical because it is anchored in the past, in the sense that the images of the future come from discovery-stage discussions of great moments in the past. During the dreaming stage, stakeholders imagine and discuss inspiring possibilities and great achievements that can contribute to the organization’s success and to the world.

The dreaming stage includes all the people involved in the previous stage. It is this, huge in some organizations, group that does the dreaming and creates the vision of what the organization is to become. The advantage of this process is that it is the “workforce” that creates the vision. Unlike a vision that is the brain child of the top administrator or the top management team, this vision belongs to the individuals whom the vision will most directly affect—employees, community members, etc. They carry it out in their heads when they leave the meeting. The top administrator does not have to hold numerous meetings to share the vision and to convince people that this is a good thing.

People pursue their own dreams; they do what they see in their minds. As a result of the dreaming stage, a common vision occurs to a large group of organizational members. Consequently, we automatically—even if we do nothing else—have a large number of people pursuing the same end.

DESIGNING^a

In the third stage—designing—the stakeholders create images that serve as the architecture of the new organization. The architecture includes new or changed structures, processes, and systems. They are based on the recollections of the moments of greatness that were shared in Stage 1 of the process. Some of the elements that might be designed could include partnerships, planning processes, environmental positions, societal relationships, results, leadership, training, communications, and hundreds of others. The participants do not actually redesign existing systems or processes but—through discussion—identify the systems, structures, and processes that will be designed in detail later and set the model, goals, and boundaries for the actual designs.

^aThese two steps were combined into one in the Imagine Chicago project and were called “*creating*.”

DESTINY^a

This stage is also called the “delivery” stage, because its goal is to ensure that the dream is brought to reality. In this stage, groups volunteer and appoint themselves to carry out the actual design and implementation work that must go on within each of the design elements chosen in the previous step. This stage is an ongoing stage, since implementation takes time. During that time, the large group acts as a massive design team, as the entire group reconvenes from time to time to receive updates, approve designs, lend support to volunteer or implementation teams who may be struggling with implementation questions, redirect efforts that have gone astray, etc.

It should be noted that the large group’s role is not that of monitoring progress. Its actual job is to sustain and build energy by recreating the culture of appreciation that was reached in the early stages and by infusing creativity into the implementation process. In a sense, each time the large group reconvenes, the energy and optimism that were built in the first step of the process are recreated. Thus, the large group’s work at this stage is to maintain optimism and commitment within the small groups that are doing the hard work of implementation.

The challenges during this stage are mainly two: One is to sustain the high level of energy that initiates the process through to the end of the process and the other is to maintain the level of creativity and innovation that is fostered in the beginning stages of the process.

The process starts off in Stage 1 with a very large group—typically. Some call this the Conference Group. The positive nature of the work done in the early stages builds tremendous excitement and energy. As time goes on during the fourth stage, this energy and excitement would naturally ebb. The key is to keep the large group—the Conference Group—engaged and energized. This group must be reconvened throughout all the stages. In this, the destiny stage, instead of analyzing results, critiquing actions, checking time tables and commitments, etc., the large group describes positive changes that have occurred, exciting moments of implementation, success stories, and so on. They do what is necessary to maintain the momentum and to keep the energy positive.

Among the important things the Conference Group does for the implementation teams is to create a reinforcing culture. By recognizing and elevating the status of innovative thinking, “out of the box” ideas, and creative action, they make it OK, even valued, to be innovative and daring. In one case, in a very large chemical

plant, an accountant, who had never been in the operations portion of the plant to any significant extent, simply wondered out loud to the implementation team if something would work in the operations area. The team explored the idea and saved the company tens of thousands of dollars each year. This story was told at a Conference Group meeting, and people went wild with enthusiasm. Not only had the company saved a lot of money, but an accountant came up with the idea (imagine that), and accounting and operations were actually working together (you are kidding—unheard of). The event sparked enormous amounts of fun, enthusiasm, optimism, and innovative ideas. These kinds of stories tell people that it is a good thing to go beyond the ordinary.

TRANSFORMING EFFECTS

The AI process has the potential to change the organization in a number of ways. Some outcomes are very tangible; others are quite intangible. What most reports talk about are the tangible outcomes—improved productivity, greater efficiency, etc. But no less important or significant than these changes are the intangible types of changes that organizations report.

Here are a few of the areas in which leaders see and report changes in their organizations. The organization’s members become more positive and optimistic. As a result, people begin to experiment with unconventional ways of doing things. They are encouraged by what goes on within the process to risk change and to bring about new ways of doing things. Innovation expands, as people begin to think about all sorts of new ways to do the work of the organization. Finally, the spirit of cooperation and of sharing thoughts as a way to spark creativity grows. In the “Imagine Chicago” project, hundreds of learning communities were created, and future (now current) neighborhood leaders were created. They grew from the school children who participated in the early stages to positive leaders today. These are not particularly measurable outcomes, but they are observable. Furthermore, they are changes at the heart of the processes that drive organizational success.

CONCLUSIONS

AI is an organizational intervention that has been available for a couple of decades. It can lead to substantial change both in terms of the creation of new methods, systems, and structures and in terms of new or improved processes that are at the heart of organizational and community performance. There are, after 20 years, a number of excellent sources that

^aThese two steps were combined into one in the Imagine Chicago project and were called “creating.”

can help the reader understand AI and how to use it. It is likely that leaders who wish to embark on this journey may wish to use outside facilitators to help with the process—especially if the system the leaders are attempting to change is large. Many of the authors of the references given below can assist leaders, and many universities in their schools of business or departments of public administration have faculty members who are able facilitators of this process.

REFERENCES

1. Whitney, D.; Trosten-Bloom, A. *The Power of Appreciative Inquiry*; Berrett-Koehler: San Francisco, 2003.
2. Watkins, J.M.; Mohr, B.J. *Appreciative Inquiry: Change at the Speed of Imagination*; Jossey-Bass/Pfeifer: San Francisco, 2001.
3. Bright, D.S.; Cooperrider, D.L.; Galloway, W.B. Appreciative inquiry in the Office of Research & Development. *Public Perform. Manage. Rev.* **2006**, *29* (3), 285–306.
4. Skinner, S.J.; Kelley, S.W. Transforming sales organizations through appreciative inquiry. *Psychol. Market.* **2006**, *23* (2), 77–93.
5. Berrisford, S. Using appreciative inquiry to drive change at the BBC. *Strategic Commun. Manage.* **2005**, *9* (3), 22–25.
6. West, D.; Thomas, L. Looking for the bigger picture: an application of the appreciative inquiry method in Renfrewshire Council for Voluntary Services. *Int. J. Inform. Manage.* **2005**, *25* (5), 429–441.
7. Miller, M.G.; Fitzgerald, S.P.; Murrell, K.L.; Preston, J.; Ambekar, R. Appreciative inquiry in building a transcultural strategic alliance. *J. Appl. Behav. Sci.* **2005**, *41* (1), 91–110.
8. Browne, B.W. Imagine Chicago: a methodology for cultivating community. *J. Commun. Appl. Soc. Psychol.* **2004**, *14*, 394–405.
9. Ackoff, R.L. *Creating the Corporate Future: Plan or Be Planned for*; Wiley and Sons: New York, 1981.
10. Ludema, J.D.; Whitney, D.; Mohr, B.J.; Griffin, T.J. *The Appreciative Inquiry Summit: A Practitioner's Guide for Leading Large-Group Change*; Berrett-Koehler: San Francisco, 2003.
11. Salopek, J.J. Appreciative inquiry at 20: questioning David Cooperrider. *Training Dev.* **2006**, *August*, 21–22.
12. Cooperrider, D.L.; Whitney, D.; Stavros, J.M. *Appreciative Inquiry: Handbook*; Lakeshore Publishers: Bedford Heights, OH, 2003.

Arbitrage

Susan MacDonald

California State University, Stanislaus, Turlock, California, U.S.A.

INTRODUCTION

Arbitrage refers to the simultaneous sale and purchase of identical securities or other financial instruments on different markets to take advantage of price differentials for profit. When applied to the public sector, arbitrage involves state and local government use of proceeds from tax-exempt bonds on investments that yield a higher rate of return than the interest on the bonds themselves. The difference between the two rates is considered profit or arbitrage, and since 1986 must, with few exceptions, be rebated to the IRS.

In the private sector, true arbitrage is completely hedged. In other words, both sides of the transaction are guaranteed at the time the position is taken so there is no risk of loss. If Security X is selling in New York for \$50 per share and in Chicago for \$49.50, the arbitrageur would purchase shares in Chicago and sell them simultaneously in New York, making a profit of \$0.50 per share. The price differential or profit is also known as the spread. Transaction costs must, of course, be deducted from the spread and they may include commissions and interest, if money is borrowed to purchase the shares. Arbitrage differs from traditional investing in that profits are made by the trade itself, not from the appreciation of a security. In fact, holding securities long enough for them to change in value is generally considered a risk by the arbitrageur.

Efficient markets do not, by definition, afford many opportunities for profit making through this type of trade, and arbitrage has been credited with contributing to market efficiency and “the law of one price.” This does not mean that efficient markets afford *no* opportunities for arbitrage; it does suggest, however, that arbitrageurs have had to modify their approach as they look for new opportunities in increasingly efficient markets. There are essentially five types of arbitrage in the private sector: simultaneous, risk, index, pair trading, and technical trading. In the public sector, there is only public debt arbitrage, which is essentially a form of simultaneous arbitrage—although it does not exactly occur simultaneously. Several factors have been instrumental in changing the nature of arbitrage over time; these include new market opportunities; new technology, especially in telecommunications and data processing; and advances in

mathematical and statistical theory. This essay will briefly trace the evolution of arbitrage—largely in the United States,^a and conclude with some observations about new venues for arbitrage and potential ramifications for public administration.

SIMULTANEOUS ARBITRAGE

Little was written about arbitrage until the early 1970s, although it had been practiced for almost 50 years in the United States. Guy Wyser-Pratte^[2] was one of the first to explain the details of risk arbitrage, and he traced the practice back as far as 13th century Italy when Venetian merchants took advantage of differences in currency exchange rates in order to make a profit. Einzig,^[3] writing in 1970 about foreign exchange markets, speculated that arbitrage in silver coins might be traced as far back as 500 B.C. and the ancient Persian Empire. In any event, profiting from differences in currency exchange rates is perhaps the most basic type of arbitrage and is still widely practiced today. Textbook examples and exercises can be readily found on the web. An example from the Association for Computing Machinery and University of Virginia explains:

Arbitrage is the trading of one currency for another with the hopes of taking advantage of small differences in conversion rates among several currencies in order to achieve a profit. For example, if \$1.00 in U.S. currency buys 0.7 British pounds currency, £1 in British currency buys 9.5 French francs, and 1 French franc buys 0.16 in U.S. dollars, then an arbitrage trader can start with \$1.00 and earn $(1 \times 0.7 \times 9.5 \times 0.16)$ 1.064 dollars thus earning a profit of 6.4 percent.^[4]

Later, the author cautions that the trader must *begin and end* with the same currency; otherwise the profit remains only hypothetical. This is consistent with the idea of hedging, or both sides of the transaction being known in advance and acted upon, thus without risk.

^aFor a comprehensive examination of cross-national arbitrage opportunities, see Ref.^[1], which provides a thorough and accessible comparison of markets, stock exchanges, and taxing circumstances affecting arbitrage opportunities around the world.

Several aspects of currency arbitrage are applicable to arbitrage in general: first is the timing. Currency exchanges take place at virtually the same time, so there is little risk of exchange rate volatility. Second, cash resources need not be tied up for long periods so there are no opportunity costs.^b Third, transaction costs must be deducted from the profit. This was probably less a concern in the 13th century than today, but arbitrage often involves small price differentials and is therefore financially more viable for those who can trade with few transaction costs. Fourth, currencies are convertible in two directions (francs can be exchanged for dollars, just as easily as dollars can be exchanged for francs). This increases the flexibility of currency conversion over other types of arbitrage. Similarly, there are often two components, or legs, to arbitrage transactions that must be balanced for profits to be realized; in the case cited above, conversion back to dollars was necessary for the profit to be realized. Finally, price differentials are normally short lived. Competition in the arbitrage arena is limited, but fierce. Actors must be able to act quickly. In sum, the arbitrageur takes advantage of small differences in price by trading *large quantities* at very *little cost* and in *very little time*. In the 20th century, arbitrageurs have largely been *institutional* traders whose profession is the prompt identification and agile execution of complex trades. The arbitrageur is, therefore, very different from most individuals involved in financial markets—he or she looks for price differentials within and across markets and engages in trade to exploit these differences. They have highly specialized skills and *substantial* resources—both in terms of dollars and state of the art technology—at their disposal.

EARLY FORMS OF ARBITRAGE AND TENDENCY TRADING

Historically, opportunities for arbitrage were found largely in markets dominated by government issues—whether currency, bills of exchange, or public bonds. The legendary story^c of the Baron von Rothschild's actions around the Battle of Waterloo provides a useful illustration. In 1815, the head of London's Rothschild bank sent carrier pigeons to the Battle of Waterloo so that he could learn the results of the battle before the news became public. Upon the pigeons' return, Rothschild ostentatiously sold British bonds on the London Stock Exchange, giving the impression that the British had lost to Napoleon. This led others

to sell their bonds, causing a drop in price. Rothschild then surreptitiously repurchased the bonds for a profit before human messengers arrived with news of Wellington's victory—and a recovery in the price of the financial instruments. Poor communication, lack of information, and timing played important roles in this arbitrage—as did a little deception. Of course these conditions were short lived.

Throughout the 19th century, European markets provided similar opportunities for arbitrage. As communication across markets improved, it became easier for traders to take advantage of differences in the price of securities on the London, Paris, Amsterdam, and Frankfurt exchanges. The simultaneous trading of identical securities kept arbitrage risks low and the trade straightforward, although currency fluctuations might complicate an arbitrageur's calculations. As traders became more experienced with equity price differentials, the spreads narrowed and opportunities for arbitrage lessened. This led the more adventuresome to begin *tendency trading*, or selling Security A in a temporarily upward-moving market and buying Security A in another market where it appeared to be declining. Such trading modified our definition of arbitrage by *expanding the time frame* for completing transactions. It also increased the risk. In the United States, traders found opportunities for arbitrage in railroad restructurings and similar corporate events that were common in the 1930s.

RISK ARBITRAGE AND CORPORATE EVENTS

Market circumstances in the 1930s provided new opportunities for arbitrageurs in the United States. During the Great Depression, many railroad companies found themselves undercapitalized and forced into bankruptcy. Those who were able, reorganized, recapitalizing their operations and providing new opportunities for arbitrage. A merger usually involves two corporations—one stronger, let us say A, one weaker, let us say B. A offers to take over B by using A's stock or cash, or a combination of the two, and converting B's shares into A's at a time certain in the future and at a specified ratio, usually less than 1 : 1. Negotiations between the two sets of corporate leaders take place in secret. However, once the *corporate event*^d has been made public, a time frame for completing the merger is announced. During this time, shareholders of both

^bIn economics, opportunity costs refer to alternative uses of financial resources that might yield greater profits.

^cThe story is told in Ref.^[5].

^dThe term "corporate event" is often used instead of merger, but its application is far broader and includes any corporate action such as mergers, acquisitions, hostile takeovers, as well as corporate decisions to convert preferred stock issues or convertible bonds into common shares. See especially Ref.^[6].

corporations must vote on the merger; and government agencies, like the Department of Justice, must act on the proposal. Denial by any one of these bodies will delay the process at best, or scuttle it altogether, at worst. For the arbitrageur, either circumstance may not only destroy his or her chances of profit, it may result in substantial losses.

The arbitrageur makes money from mergers, acquisitions, and other corporate events through a speedy yet astute assessment of the proposal, access to substantial amounts of cash (often borrowed at very low rates), and privileged treatment on the exchange, such as that reserved for large institutional traders. Once the intended merger has been made public, the arbitrageur will begin gathering details about the arrangement. Some of this information will be available in public documents; some will be available only by word of mouth among insiders. The arbitrageur is particularly interested in the business soundness of the proposal. Once this has been established, the arbitrageur considers problems that might prevent the merger from occurring. These include actions of regulatory bodies like the FTC, FCC, or ICC; tax decisions by the IRS; antitrust investigations by the U.S. Department of Justice, as well as the votes of shareholders in both corporations. Again, time is crucial in assessing the merits of the opportunity, as others will be examining the merger for their own benefit. Finally, the arbitrageur's own circumstances, including tax consequences, access to capital, and available interest rates must be factored into the equation before she enters into the process.

Basically, what the arbitrageur tries to do is situate himself in such a way as to take advantage of the anticipated boost in value that *B* will experience once the merger has been consummated. The arbitrageur does this by selling *A*'s shares short—that is, borrowing shares of *A* and then selling these borrowed shares on the market. The lender will specify when the shares are to be returned and the conditions under which the loan is made, including interest rate, securitization requirements, etc. The advantage for the arbitrageur is that he will be able to *return the shares*, not cash, at the time specified. The arbitrageur then uses the cash he has obtained from selling *A*'s shares to purchase *B*'s shares long. In other words, he buys *B*'s shares at their current value and holds them until the merger has been consummated. Because *B*'s shares will be converted into *A*'s according to the ratio specified in the agreement, the arbitrageur can calculate in advance how much profit to count on—assuming, of course, that the merger materializes in the time projected. Any change in either condition will affect the trade's profitability. Holding shares beyond the negotiated date converts the arbitrageur into an investor, making him susceptible to all of the market's myriad vagaries with

few of its benefits. It is because of the extended time frame and uncertainty that this is known as *risk arbitrage*.

Mergers were not the only corporate events that provided arbitrageurs with an opportunity for profit. Other events included the divestiture of public utility companies following passage of the 1935 Public Utility Holding Company Act, corporate acquisitions—both friendly and hostile, price disparities between common shares and convertible stock, convertible bonds, or purchase warrants issued by the same corporation, and cash-tender offers.^e The specifics of each of these transactions vary, but the merger principle remains the same: Corporate securities that share a common underlying value were temporarily out of sync. Arbitrageurs took advantage of the disparity, which they realized was short term. In the process, they often provided the liquidity necessary for the corporation to complete its transaction successfully.

Interest in arbitrage has waxed and waned since the 1930s, but the veil of secrecy surrounding it was lifted during the 1970s. This was due in part to authors like Wyser-Pratte and Evans, Jr., whose works have already been mentioned. However, it was also due to the extensive publicity surrounding arbitrageurs like Ivan Boesky, whose notoriety both attracted new investors and eventually served as a cautionary note of the risks involved. In 1977, Boesky boasted of annual profits approximating \$30 million; within 10 years his net worth was estimated at \$250 million. But in 1987 Boesky was also convicted of conspiring to file false trading papers and sentenced to three years in jail. Three years later, Michael Milken, another notorious arbitrageur, received a 10-year sentence for insider trading. Arbitrage could make one wealthy, but its risks appear psychological as well as financial.

INDEX OR ABSOLUTE ARBITRAGE

Index arbitrage is “the simultaneous sale and purchase of an *index future* and of a *portfolio* replicating the same index.”^[8] Like the example involving corporate mergers, two valuations are being compared; unlike the merger example, it is not immediately clear whether the index future or the portfolio should be bought or sold. To make this assessment, the definition requires further unpacking.

A *futures contract* is an agreement to buy or sell a commodity, currency, interest rate, or equity for a specific price at a specified time in the future. An *index future* is “a futures contract on a stock

^eArbitrage involving these other corporate events are covered extensively in Ref.^[7].

index ... [that is,] an agreement between a seller (short position) and buyer (long position) to a *cash settlement* based on the *change in the stock index's value* [emphasis added] between the date the futures contract is entered by the two parties and some future date.”^[9] Thus, simply put, we have four monetary values: today's futures (F_{T1}), today's stock index (S_{T1}), tomorrow's futures (F_{T2}), and tomorrow's stock index (S_{T2}). Of these, only the value of today's stock index is certain. Today's futures valuation (F_{T1}) is known, but that number includes an estimate of the stock index's value at *expiry*—the date in the future when the contract will expire and a final accounting will be made. Because estimates of the *index future* change regularly, the value of futures contracts will fluctuate. However, the closer one gets to expiry, the less fluctuation there will be. Indeed, at expiry, *the index future will equal the stock index*; it is just that the two indices do not mirror each other precisely in the intervening time. Because the two indices are identical at expiry, the transaction is *said to be* without risk.

In sum, an *index futures contract* anticipates a given exchange's value at a specific time in the future. Purchasing this type of derivative^f can be useful for people or institutions that own diverse portfolios on a given exchange, and that want to hedge their investment position. If they assume the market will go up, they may purchase an *index futures* predicting that the exchange will decline. If the index does indeed decline, the holder of the *index futures* can offset losses in his or her portfolio with gains from settling the *index futures*, once it has expired. This type of derivative instrument was introduced by the Kansas City Board of Trade in February, 1982 but the Chicago Mercantile Exchange (CME) followed suit in April that same year, trading futures contracts based on Standard & Poor's 500 common stocks. By 2000, 90% of all futures trading in the U.S. was conducted on the CME.

Financial instruments like stocks, bonds, futures, etc. are traded on exchanges in a bidding process; their values are negotiated by sellers presenting their *ask* price and buyers presenting their *bid*. To manage the results of these negotiations, futures contracts are settled daily, while all futures contracts (on the S&P 500) expire quarterly—on the third Friday of March, June, September, and December. The last trading days are the Thursdays prior to *expiry*. To appreciate what is entailed in the daily settlement process, it is helpful to review some hypothetical numbers. Clearinghouses like the CME require traders to make initial deposits, settle their accounts daily, and postpone full payment until expiry. Futures contracts themselves

are denominated in *points* that represent a fraction of their total value. On the CME, the initial deposit is \$23,438 *per point* and *each point is worth* \$250. Thus, if an investor buys 100 futures at \$1000 each and they close that day at \$1005 she will have made \$125,000, but will have to pay the clearinghouse:

\$2,343,800	(100*\$23,438) Initial deposit required at end of Day 1
<u>−\$125,000</u>	((1005 − 1000)*100*\$250) unrealized gains, in the trader's account
\$2,218,800	Net deposit required at the end of Day 1

The value of the starting position is much higher than the deposit. Because each point is worth \$250, the starting position is (100*250*1000) or \$25,000,000. Each day, both realized and unrealized gains and losses are tabulated and an accounting is made with the clearinghouse. In addition to the deposit, the CME requires a minimum reserve or “maintenance requirement” of \$18,750 per unit. If the trader in the example above had *lost* \$19 per share at the end of Day 1, instead of gaining \$5, her losses would have totaled (\$19*100*\$250) or \$475,000. This would then have been *subtracted* from the initial deposit of \$2,343,800, leaving her with a deposit of only \$1,868,800 or \$6200 shy of the required minimum of \$1,875,000. Had this occurred, she would have had to give the CME an additional \$6200 to continue trading, or settle her account completely at a substantial loss. Of course, daily losses *can* be made up on subsequent days, depending on the market. These daily tabulations are zero-sum—for every loser there is a winner, and vice versa. Excluded from the clearinghouse's daily accounting is interest, if cash or other financial instruments are borrowed—unless, of course, the loan is arranged through the clearinghouse itself, which typically charges higher rates than the market. Many arbitrageurs consider interest payments a cost of doing business that must eventually be deducted from profits. At expiry, the balance left in the trader's reserve is returned.

To get a sense of the numbers involved in this type of arbitrage, Reverre^[10] estimates that approximately one dozen institutional investors engage in this activity, and each has approximately \$5 billion in cash at their disposal. Finally, he estimates that annual profits per arbitrageur are around \$40 million.

How, then, is it possible for arbitrageurs to make these profits? Even a somewhat simplified answer requires us to return to the four values mentioned earlier: today's futures (F_{T1}), today's stock index (S_{T1}), tomorrow's futures (F_{T2}), and tomorrow's stock index (S_{T2}). As mentioned, we know the value of S_{T1} ; it is today's value of the stock exchange index. S_{T2} is calculated by projecting the cost of holding the stock

^fA derivative is a financial instrument whose return is *derived from* a formula or something other than interest.

portfolio, in its current configuration, until the expiration date. To calculate this, we estimate the *carrying costs* of the portfolio and then add these costs to the current value of the stock index. Assuming that the underlying value of the stock will remain constant, we estimate the interest required to borrow money for the portfolio and deduct revenues anticipated from dividends. We also know that the value of the futures contract (F_{T2}) will, by definition, equal the value of the stock portfolio (S_{T2}) at the time of expiry. If today's price of the futures contract exceeds the projected value of the stock portfolio that has just been calculated, ($F_{T2} > S_{T2}$), the arbitrageur will conclude that the futures contract is overpriced. He will then consider selling the index futures (taking a short position) and buying the stock portfolio (taking a long position). On the other hand, if the estimated value of S_{T2} exceeds today's price of F_{T2} ($F_{T2} < S_{T2}$), he will consider buying the index futures and selling the stock portfolio. Of course, transaction costs must be calculated to ensure that the arbitrage is worthwhile. Once again, profit margins are small, so it is imperative that transaction costs be kept as low as possible and that the volume involved is high. Only a few institutional traders can meet both conditions.

When index arbitrage was introduced in the early 1980s, computers facilitated many calculations, but were not sufficiently powerful to calculate the changing values of *all* stocks on a given exchange in the time required by arbitrageurs. Rather than base their calculations on all stocks, then, many institutional traders selected *representative* portfolios, which they traded in lieu of the entire index. As computers, and their software, became more sophisticated, arbitrageurs were able to rely on increasingly precise and sophisticated calculations. This made their jobs at once easier and more complicated, as trading itself once again moved into new territory.

PAIR TRADING AND SPECULATIVE ARBITRAGE

Pair trading, or hypothetical arbitrage, resembles the simultaneous trading of identical stocks on different exchanges for a profit. With pair trading, however, the sale and purchase can take place on the same exchange because the stocks are *similar*, rather than identical. What constitutes "similarity" is open to debate or judgment. What the arbitrageur looks for are *two stocks that normally trade at the same price*, but for some reason unrelated to their underlying value are trading at different prices. This will enable the arbitrageur to short the higher priced stock and to buy the stock selling at the lower price—making a profit when the two come back into alignment or equilibrium.

Speculative arbitrage, also described as technical trading, involves the same techniques developed in index arbitrage and pair trading, but takes advantage of advances in statistical probability, supercomputers, and mathematical modeling to identify arbitrage opportunities on a given exchange. Discussion of speculative arbitrage has exploded in the academic literature—it has become the new *terra incognita* for the most mathematically gifted in business schools. Indeed, the Association for Computing Machinery and University of Virginia web site that discussed currency exchanges, mentioned earlier, was set up as a *programming contest* to determine whether different currency exchange rates offered an opportunity for arbitrage. The program can be used as a prototype for any price comparison, including stock prices. Despite the popularity of speculative arbitrage in the academy, it has been largely prohibited on stock exchanges.

PUBLIC DEBT ARBITRAGE

Since 1986, public finance officers have had to become increasingly familiar with public debt arbitrage. Arbitrage, in this context, refers to any "profits" earned by state and local governments from investing their bond proceeds in financial instruments that yield more than the bonds themselves pay. Congress did not prohibit arbitrage in the Tax Reform Act of 1986 (P.L. 99-514); it merely required state and local governments to *rebat*e or tax arbitrage profits *in full* to the IRS. This requirement was applied to almost *any profits* earned between the first bond sale and the date the debt was retired, which could extend several years depending on the life of the project and debt restructuring. Some governments had issued bonds with the sole purpose of profiting from the interest differential—arbitrage—but the 1986 legislation applied to *any portion* of the proceeds (with limited exceptions) from the debt issue.

Historically, Congress neither prohibited such investment, nor taxed the surplus generated by these transactions. However, it began restricting the use of tax-exempt bonds as early as 1968, when Congress eliminated the tax exemption on any public debt used for *nonpublic* purposes. Indeed, if more than 10% of the proceeds from any public bond issue were used for *private* purposes, the entire bond's tax-exempt status would be revoked. In 1969, Congress extended the loss of tax-exempt status to bonds whose proceeds were used, in whole or in part, for investment in higher yielding instruments, or arbitrage. Although Congress allowed some exceptions, this legislation had the *effect* of increasing state and local governments' debt costs. Investors are attracted to tax-exempt bonds, whose returns are generally *less* than other financial

instruments, because their tax-exempt status allows investors to make money without incurring additional tax liability. If the bonds lose their tax-exempt status, investors lose that tax advantage and then expect the issuers to offer bonds with a higher yield, or interest payment—otherwise there is little financial incentive to purchase the lower yielding debt instruments. Increasing the interest payments not only costs the issuing agency more, it reduces the price differential or amount that can be earned through arbitrage. Nevertheless, many states and municipalities still found it worthwhile to forfeit the tax-exempt status of some bond issues in order to earn a higher yield. In the mid 1980s, before additional tax reform, Georgia's State Auditor, G.W. Hogan, estimated that as many as 40,000 state and local governments were able to add at total of \$729 billion to public coffers through arbitrage; the state itself earned almost \$40 million per year in this way.^[11]

In 1982, Congress passed the Tax Equity and Fiscal Responsibility Act (TEFRA), which tightened controls on municipal bonds by requiring agencies to *register* their bonds in order to secure tax-exempt status. (Arbitrage bonds were omitted because they had already been declared non-tax-exempt.) Registration was intended to reduce the fraudulent issuance of tax-exempt bonds. South Carolina officials believed the legislation violated both the states' Tenth Amendment reserved rights and their immunity from intergovernmental taxation, established in the 1895 Supreme Court decision, *Pollock v. Farmers' Loan & Trust Co.* Although *South Carolina v. Baker* was not decided until Spring 1988, the U.S. Supreme Court distinguished between states' rights to issue debt and Congress' ability to determine which debt instruments would be accorded tax-exempt status. As long as Congress did this in a nondiscriminatory way, the Court concluded, there was no infringement on the Constitution. The Court also addressed the question of taxing state bond interest.

By 1986, the federal government was facing budget deficits of historic proportions and members of Congress were making a concerted effort to eliminate as many tax loopholes as possible. Those who favored additional restrictions on state and local bonds argued that it was an inefficient use of federal tax dollars to ignore the revenues that states and localities were earning through arbitrage. They proposed that agencies *rebat*e their entire arbitrage profits to the IRS, in effect imposing a 100% tax on the spread, *in addition to* having removed the bonds' tax-exempt status. Rebate opponents argued, as in *South Carolina v. Baker*, which had not been decided at the time, that the federal government had no authority to tax state and local governments. Despite objections, the tax measure passed, although some exceptions were incorporated

into the final version. The following year, the U.S. Supreme Court heard *South Carolina v. Baker* and issued its decision in April 1988 (485 U.S. 505). The Court explicitly overturned *Pollock*, declaring that Congress *could* impose a nondiscriminatory tax on bond holders by removing the tax-exempt status of certain bond issues. It also upheld Congress' right to tax states on any bond interest earned. Presumably, this would apply only to registered, tax-exempt bonds as unregistered municipal bonds would not be tax-exempt and therefore trade closer to market rates.

The Supreme Court's actions have left opponents to focus primarily on the law's exceptions. In general, arbitrage restrictions apply to all registered, tax-exempt bonds issued after August 31, 1986 and these restrictions continue to apply to all bond proceeds until the debt has matured or the bonds have been redeemed. Congress allowed exceptions for small issues or bonds whose entire proceeds are spent within a designated, and limited, period of time. For example, municipalities issuing no more than \$5 million annually in tax-exempt bonds for public purposes do not have to rebate their positive earnings. This amount was increased to \$10 million in 1997 and to \$15 million in 2001 for public school construction, but remains at \$5 million for all other purposes. A second exception involves bond proceeds for construction where expenditures are completed within two years and kept to an agreed upon spending schedule. Any investment profits earned in the interim are exempt from rebate. Other exceptions involve Tax and Revenue Anticipation Notes (TRANS), debt service funds, and tax-exempt investments, but in each case, the investment period is short and the use of funds is narrowly defined.^g

The difficulty for bond issuers, and their finance officers, is that each debt issue, and the proceeds from each set of bond investments, must be tracked individually and in detail. Commingled funds, bond refinancing, and separate escrow accounts all complicate this tracking of bond proceeds for arbitrage-rebate purposes. Payment to the IRS is required every 5 years, but the agency has only 60 days after the due date to submit 90% of the full amount owed. Similarly, full payment is due to the IRS within 60 days of the last bond's retirement. This includes bonds that have been refinanced. Determining if and when investments meet these criteria is time consuming and costly.

In 1999, the Government Finance Officers Association (GFOA) released a survey of state and local government officials, that it conducted jointly with the National Association of Counties, and found that the average amount of time spent on arbitrage tracking

^gFor further information on arbitrage exceptions and requirements, see Ref.^[12].

was 303 hours per year, and the average annual cost was \$33,305. Costs ranged, however, as high as \$409,000 with the number of hours reaching 3360.^[13] These costs and staff expenditures were made regardless of whether arbitrage was involved. Indeed, bond tracking *must* occur if only to demonstrate to the IRS that the issuing agency has not engaged in arbitrage or that it meets the criteria for exemption.

GFOA has been working with Treasury staff and Congress to raise the \$5 million nonschool construction limit to \$25 million for municipal bonds. It has also been urging legislators to allow public entities without taxing authority to issue tax-exempt bonds so that cooperative ventures can be financed. More recently, GFOA has been promoting the idea of safe harbor legislation that would simplify both the accounting requirements and consolidate exceptions into one package.

NEW VENUES FOR ARBITRAGE

Since 2000, arbitrage has spread in various directions. Financial arbitrage has become more prevalent in hedge funds and large investors, like pension funds, are reported to be increasingly involved in arbitrage to offset market declines. It is also being used to take advantage of temporary price differentials that can occur when new regulations are phased in. Capital structure arbitrage exploits price differentials that occur when a company's debt and equity are out of sync. Arbitrage is also being used metaphorically. Articles now include topics like information arbitrage or product arbitrage where consumers take advantage of differences between the old, physical market and the internet market.

CONCLUSIONS

In the market, arbitrage is not only a source of profit for the powerful few, it has historically provided liquidity for corporate actions, contributed to market efficiency, and played an equalizing role through the law of one price. Arbitrage has played a different role in the public sector. In both sectors, arbitrage involves a sale and a purchase—at least nominally. But in the

public sector, state and local governments borrow money (sell bonds) and then invest the proceeds (purchase financial instruments)—there is no equivalent sale and purchase of an identical financial instrument, as in the private sector. Also, in the private sector, arbitrage serves the broader community as well as those directly involved in producing it; by contrast, in the public sector, the more powerful allow arbitrage but tax it at 100% and, in the words of Chief Justice Marshall, thus destroy it.

REFERENCES

1. Reverre, S. *The Complete Arbitrage Deskbook*; McGraw-Hill Companies, Inc.: New York, 2001.
2. Wyser-Pratte, G.P. *Risk Arbitrage*; Bulletin [of the] C.J. Devine Institute of Finance; New York University, Institute of Finance: New York, 1971; 74–75.
3. Einzig, P. *The History of Foreign Exchange*; Macmillan and Company, Ltd.: London, 1970.
4. *Arbitrage*; <http://acm.uva.es/p/v1/104.html> (accessed July 2004).
5. Wyser-Pratte, op. cite., 7 pp.
6. Evans, M.D., Jr. *Arbitration in Domestic Securities in the United States*; Parker Publishing Company, Inc.: West Nyack, NY, 1965.
7. Evans, M.D., Jr. (op. cite.); Wyser-Pratte, G.P. *Risk Arbitrage II*; Monograph Series in Finance and Economics. Salomon Brothers Center for the Study of Financial Institutions, Graduate School of Business Administration; New York University: New York, NY, 1982; 3–4.
8. Reverre, S. op. cite. 371 pp.
9. Santoni, G.J.; *Index Arbitrage and Stock Price Volatility*; Chicago Mercantile Exchange (n.d.) www.cme.com/files/indexarb.pdf 3 (accessed July 2004).
10. Reverre, S. op. cite. 389 pp.
11. [n.a.] Carpetbaggers: Southern battle over taxes. *Time* **1987**, 130 (11), 63.
12. (a) Eichhoff, G. Arbitrage compliance: a primer for finance officers. *Govern. Finance Rev.* **1999**, 15 (4), 25–28; (b) Nutter, S.E. Tax-exempt governmental bonds, 1991–1995—statistical data included. In *Statistics of Income Bulletin*; U.S. Department of Treasury: Washington, DC, 2000 (accessed in Iôksmart September 2004).
13. Government Finance Officers Association. In *Issue Brief: Arbitrage Rebate*; 1 pp. Updated May 2003.

Asset Pricing

Greg Chen

School of Public Affairs, Baruch College, City University of New York,
New York, New York, U.S.A.

INTRODUCTION

This article discusses asset pricing, its theories and applications. It starts by reviewing portfolio theory and the concept of diversification. It then introduces the centerpiece of asset pricing model, the capital asset pricing model (CAPM). This is followed by a review of empirical testing results of the model. The failure of the model in explaining and predicting market data motivated the modification and development of new and more complex models. The article, subsequently, surveys and assesses selected alternative models, including the arbitrage pricing theory (APT), other multifactor models, conditional CAPMs, and CAPM with higher order comoments.

PORTFOLIO THEORIES

Portfolio theory is the foundation for modern CAPMs. Portfolio theory was laid out by the groundbreaking work of Markowitz^[1] and Tobin.^[2] Markowitz developed the notion of efficient frontier of portfolios. Each portfolio on the efficient frontier offers the maximum possible expected return for a given level of risk. Tobin laid the foundation of the separation theorem. The theorem states that a portfolio can be formed by separating the problem into first finding an optimal combination of risky securities and then deciding whether to lend or borrow, depending on your attitude toward risk.

In portfolio theory, investors face two kinds of risks, namely, diversifiable (unsystematic) and non-diversifiable (systematic). Unsystematic risk is the component of the portfolio risk that can be eliminated by increasing the portfolio size. Risks that are specific to an individual security such as business or financial risk can be eliminated by constructing a well-diversified portfolio. Systematic risk is associated with the economy or the overall movements in the general market, and therefore is often referred to as the market risk. The market risk is the component of the total risk that cannot be eliminated through portfolio diversification.

CAPITAL ASSET PRICING MODEL

Building upon the portfolio theory, Sharpe^[3] developed a computationally efficient, single factor, ex ante model to relate risks to expected returns, the CAPM. Mathematically, CAPM is expressed as follows:

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f] \quad (1)$$

where $E(R_i)$ is the expected return on asset i , R_f is the risk-free rate, $E(R_m)$ is the expected return on the market portfolio, β_i is the beta on asset i , and $E(R_m) - R_f$ is the market risk premium.

The beta in the model represents an investment's non-diversifiable risk relative to the market risk. In other words, the beta of an investment measures the comovement of the investment's expected return with the market's expected return. The beta for a stock is defined as follows:

$$\beta = \frac{\text{cov}(R_i, R_m)}{\sigma_m^2} \quad (2)$$

where $\text{cov}(R_i, R_m)$ is the covariance between the returns on asset i and the market portfolio and σ_m^2 is the variance of the market portfolio.

The CAPM was developed under strong simplifying assumptions. They can be summarized as:

- Investors are price takers and have homogeneous expectations.
- One period.
- Presence of a riskless asset.
- Perfect market, i.e., no taxes, transaction costs, regulations, or short-selling restrictions.
- Returns are normally distributed or investor's utility is a quadratic function in returns.

The CAPM is an elegant, but very simplified representation of reality. The extent to which the model reflects the real world needs empirical examination.

Security Market Line

Graphically, the CAPM can also be expressed by the security market line (SML). This is illustrated in [Fig. 1](#).

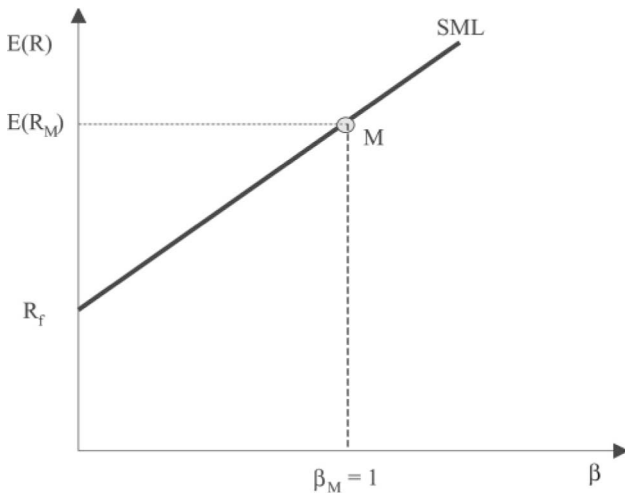


Fig. 1 Security market line (SML).

The slope of the SML is equal to $[E(R_m) - R_f]$ which is the market risk premium and that the SML intercepts the y-axis at the risk-free rate.

In capital market equilibrium, the required return on an asset must equal its expected return. Thus, the SML equation can be used to determine an asset's required return given its beta. This can also be used to assess if an investment is correctly priced. If an investment's return falls on the SML, the investment should be considered correctly priced because the expected return of the investment matches the one required by the CAPM. However, if the expected return of the investment differs from the one as required by the CAPM, the investment should be judged as either underpriced or overpriced.

Estimating Beta

The CAPM is an ex ante model relating risks to expected returns. To use the model, one needs to know how to estimate the beta of a stock or portfolio. Beta is typically estimated by regressing the realized excess return on the individual assets on the realized excess returns on the market, following the regression of Eq. (3). Five years of monthly return data are often used in the estimation of beta and many financial institutions provide beta at a nominal or no cost.

$$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + \varepsilon_i \quad (3)$$

where α_i is the intercept and ε_i is the error term.

A typical data set of realized excess returns and the corresponding excess market return is illustrated in Fig. 2. The regression line superimposed into the data points is called security characteristic line (SCL). The slope of the SCL is the estimate of the beta of the security.

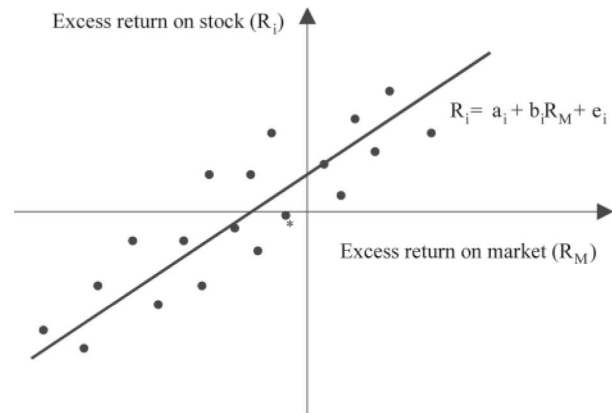


Fig. 2 Security characteristic line.

Empirical Test of CAPM

The empirical tests of the CAPM focus on the predictions about the intercept and slope in the relation between expected return and market beta.^[4,5] The approach is to regress a cross-section of average asset returns on estimates of asset betas. The model predicts that the intercept in these regressions is the risk-free interest rate, R_f , and the coefficient on beta is the expected return on the market in excess of the risk-free rate, $E(R_m) - R_f$. This prediction plays a prominent role in tests of the CAPM. In the early work, the main method of the test is cross-section regressions.

The results of the empirical test are mixed. Although earlier investigations confirm the model, the last 20 years have witnessed disappointing results. The general consensus at the present time is that the static one-period CAPM is unable to explain or to predict satisfactorily the cross-section of average returns on stocks.

There are many potential reasons for the inconsistent results of CAPM from theoretical and technical viewpoint. For example, factors, other than the single market index portfolio, may have contributed to the determination of returns;^[6] the use of realized return as a proxy for the expected returns may have produced biased results due to the aggregation of positive and negative market excess returns;^[7] the static CAPM does not allow the change of beta or market premium, conditional on existing information; and returns can be non-normal, violating a key assumption of CAPM.

Prompted by the disappointing results of the single-factor CAPM studies, extensive effort has been made to improve on the one-factor CAPM model. The key alternative models include the multifactor models, conditional CAPM, and CAPM with higher-order comoments. The following section surveys the alternative models.

ARBITRAGE PRICING THEORY AND MULTIFACTOR MODELS

One problem with CAPM resides in its assumption that the non-diversifiable risk of the portfolio (in relation to the market risk) is the sole determinant of its return. The arbitrage pricing theory (APT) models introduced by Ross^[8] extends the CAPM model to include other factors. Similar to the CAPM, the APT assumes that there is a relationship between the risk and return of a portfolio. Unlike the CAPM, the APT recognizes that other systematic factors affect the return of a portfolio too.

The APT is based on the no arbitrage principle (or law of one price). This principle states that any two identical investments cannot be sold at a different price. In other words, the theory states that market forces will adjust to eliminate any arbitrage opportunities, where a zero investment portfolio can be created to yield a risk-free profit.

The APT preserves the linear relationship between risk and return of the CAPM but abandons the single measure of risk by the beta of the portfolio. The APT model is a multiple factor model, which uses factors such as the inflation rate, the growth rate of the economy, the term structure of interest rates, in addition to the beta of the portfolio in determining the return of the portfolio. As with the CAPM, the APT can also be modified to determine the return of an individual investment.

The APT model has been empirically validated. In limited cases, the APT has been reported as outperforming the CAPM in terms of within-sample explanatory power.^[9] Moreover, APT requires fewer assumptions in its derivation which make the model more general than CAPM.^[8]

The key problem with the APT is that the factors are not well specified a priori. There is no consensus on what factors should be included in the model. Another perceived problem has to do with APT's macro perspective. Individual stock properties, such as size and earning are not normally considered in its factor selection.

A more general, and somewhat atheoretical approach is the multifactor models. These multifactor models postulate that microvariables such as size, ratio of book-to-market value, macroeconomic variables and the price to earnings ratio account for a sizeable portion of the cross-sectional variation in expected returns. Fama and French^[6] observed that the two nonmarket risk factors, SMB (the difference between the return on a portfolio of small stocks and the return on a portfolio of large stocks) and HML (the difference between the return on a portfolio of high-book-to-market stocks and the return on a portfolio of low-book-to-market stocks), are

useful factors when explaining a cross-section of equity returns.

CONDITIONAL CAPM

Another problem in CAPM has to do with conditionality. While the real world is dynamic, the model is static in nature. The conditional CAPM model is developed to address this problem. The conditional CAPM model provides for the variation of the beta, given the available information at the time. The idea is that CAPM holds period by period, although not consistently so over time. Conditioning on information at time, the conditional CAPM allows for the change of beta and market premium over time.

The empirical test results of conditional CAPM are promising, however, mixed. Some studies suggest that the conditional CAPM might hold, period-by-period, and that time-varying betas can explain the failures of the simple, unconditional CAPM.^[10] Other studies, however, show that conditional CAPM does not explain asset-pricing anomalies like book to market and momentum.^[11] They claim that conditional CAPM performs nearly as poorly as the unconditional CAPM.

HIGHER-ORDER COMOMENTS CAPM

The rationale for higher-order comoments CAPM is that investors need compensation for taking risks on skewed returns, and that skewness and kurtosis cannot be diversified away by increasing the size of portfolios.^[12,13] It is reasoned that if investors know that asset returns have conditional coskewness before hands, expected returns should include a component attributable to conditional coskewness. Harvey and Siddique^[12] reported that conditional skewness explains the cross-sectional variation of expected returns across assets and is significant even when factors based on size and book to market are included.

Kurtosis was also tried in higher order models in addition to skewness. Empirical test shows that systematic kurtosis contribute to the risk premium of an asset over and above that of system covariance and systematic skewness. Investors are generally compensated for taking high risk as measured by high systematic variance and systematic kurtosis. Investors also forego the expected returns for taking the benefit of a positively skewed market.^[14]

One identified problem with higher moments model is that ex ante skewness is difficult to measure.^[13] The issue needs further investigation, as precision in measurement is often a key or confounding factor in model formulation and validation.

CONCLUSIONS

This article reviews portfolio theory and discusses existing and evolving asset pricing models. The paper described the most popular model of CAPM, and assessed the validity of the models based on empirical test results.

The empirical record of the model is disappointing. The general consensus is that the static one-period CAPM is unable to explain satisfactorily the cross-section of average returns on stocks. CAPM is not consistent with data from the real world.

This is not totally surprising given the strong assumptions that CAPM made in its derivation. Capital asset pricing model is a highly simplified and stylized version of the reality. More complex and realistic models are needed for real world prediction. Many new and improved models have been developed over the past 50 years with qualified success. Further study on asset pricing is both essential and likely rewarding, given the central importance of asset pricing in modern finance.

REFERENCES

1. Markowitz, H. Portfolio selection. *J. Finance* **1952**, 7, 77–91.
2. Tobin, J. Liquidity preference as behaviour towards risk. *Rev. Econ. Stud.* **1958**, 26, 65–86.4.
3. Sharpe, W.F. Capital asset prices: a theory of market equilibrium under conditions of risk. *J. Finance* **1964**, 19, 425–442.7.
4. Fama, E.F.; MacBeth, J.D. Risk, return and equilibrium: empirical tests. *J. Polit. Econ.* **1973**, 81, 607–636.
5. Campbell, J.Y. Asset pricing at the millennium. *J. Finance* **2000**, 55, 1515–1567.
6. Fama, E.; French, K. Size and book-to-market factors in earnings and returns. *J. Finance* **1995**, 50, 131–156.
7. Pettengill, G.N.; Sundaram, S.; Mathur, L. The conditional relation between beta and returns. *J. Financial Quant. Anal.* **1995**, 30, 101–116.
8. Ross, S.A. The arbitrage theory of capital asset pricing. *J. Econ. Theory* **1976**, 13, 341–360.
9. Groenewold, N.; Fraser, P. Share prices and macroeconomic factors. *J. Bus. Finance Acc.* **1997**, 24, 1367–1383.
10. Jagannathan, R.; Wang, Z. The conditional CAPM and the cross-section of expected returns. *J. Finance* **1996**, 51, 3–53.
11. Lewellen, J.; Nagel, S. The Conditional CAPM Does Not Explain Asset-Pricing Anomalies, NBER Working Paper 2003, No. w9974.
12. Harvey, C.; Siddique, A. Conditional skewness in asset pricing tests. *J. Finance* **2000**, 55, 1263–1295.
13. Kraus, A.; Litzenberger, R.H. Skewness preference and the valuation of risk assets. *Journal of Finance* **1976**, 31, 1085–1100.
14. Fang, H.; Lai T-Y. Co-kurtosis and capital asset pricing. *Financial Rev.* **1997**, 32, 293–307.

Association of Southeast Asian Nations (ASEAN)

Richard K. Common

The Business School, University of Hull, Hull, U.K.

INTRODUCTION

Despite the diversity across a myriad of policy areas, the nation states that comprise the membership of the Association of Southeast Asian Nations (ASEAN) have developed a mechanism for “cooperation” in matters pertaining to public administration and policy across its member states. ASEAN now provides a regional rival to the older and arguably more important organization for public administration affairs in Asia, the Eastern Regional Organization for Public Administration (EROPA). Although ASEAN was established in 1967, it was not until the second meeting of the ASEAN Standing Committee in 1980 that the ASEAN Conference on Reforms in the Civil Service (ACRCS) was established as an ASEAN activity. The ACRCS was renamed the ASEAN Conferences on Civil Service Matters (ACCSM) in 1987, with a stronger emphasis on collaboration between member countries. Clearly, much of the impetus surrounding these developments was based on a desire among member states to facilitate economic integration in addition to positioning the region as an attractive place for foreign direct investment underpinned by strong administrative and policy capacities. A fundamental founding principle of ASEAN is cooperation between member states, and the ACCSM is regarded as one of its key cooperation initiatives. This entry will aim to evaluate the extent of cooperation between administrative systems in the region and the nature of the reforms that are being promulgated by the ACCSM.

PUBLIC ADMINISTRATION IN ASEAN

It is beyond the remit of this entry to describe in any detail the direction and nature of administrative reforms in each member nation of ASEAN (see other entries in EPAP). However, it is useful at this stage to pick up some major themes and points of difference within these countries. There are roughly three groupings within ASEAN based on the extent of modernization within the various administrative systems where there is a strong correlation between political stability and economic development.

Brunei Darussalam, Malaysia, Singapore

In this group, there have been considerable attempts by stable political leaderships to orientate administrative systems close to Western models.^[1] In particular, Malaysia and Singapore have come the closest to replicating the New Public Management observed in western nations, particularly Britain and New Zealand. However, even here it is difficult to generalize. For example, Malaysia’s “Look East” policy explicitly emulated the example set by Japan and to a lesser extent, Korea, in terms of economic and human resources policy. Brunei’s civil service has consistently looked to the United Kingdom for inspiration despite its unique monarchy.^[2] Arguably, these three nations are also the most economically successful in ASEAN.

Indonesia, Philippines, Thailand

This grouping has experienced varying degrees of political turmoil when compared to the first grouping, in addition to a similar variance in economic development. Both Indonesia and Thailand suffered particularly damaging consequences as a result of the Asian economic crisis of the late 1990s, and had to seek International Monetary Fund (IMF) loans. For Thailand, in particular, the program of economic reform that constituted the IMF response in 1997 included civil service reform. Economic progress in the Philippines has also been exacerbated by political instability. However, attempts have been made to modernize administrative systems in all three countries, but here generalizations remain difficult. For instance, Thailand’s system of government remains highly centralized when compared with that of Indonesia and the Philippines.

Cambodia, Laos, Myanmar, Vietnam

The third group of countries represent the most recent members of ASEAN, with Cambodia being the newest, joining ASEAN in 1999. However, with the exception of Cambodia, these countries are characterized by authoritarian rule and command economies, although Vietnam and Laos are engaged in cautious reform

strategies. For the purposes of this entry, what is clear is that administrative modernization has not been high on the agenda of any of these countries until very recently.

KEY COMPONENTS OF ACCSM

Under the auspices of ASEAN, it is proposed here that ASEAN acts as a forum for learning in terms of policy and administrative development. Comparisons with the European Union (EU) are largely futile in this area as a key component of the “ASEAN way” is to rely “on national institutions and actions, rather than creating a strong central bureaucracy.”^[3] However, the 1992 declaration of an ASEAN free trade area (AFTA) and the Asian economic crisis created pressures for “greater coordination and institutionalization.”^[3] Thus the ACCSM is pivotal to understanding developments in public policy and administration within ASEAN.

The main forum for discussing developments in relation to public administration policy occurs within the ACCSM biannual conferences. At each conference, technical and country papers are presented. The rest of this section will briefly review developments from the 8th ACCSM in Manila mainly because it marked the beginning of ASEAN enlargement and it signaled the start of closer cooperation and institutionalization of public administration practice within ASEAN member states.

At the 8th ACCSM in Manila in 1995, an Action Plan was endorsed entitled “Building Dynamic and Responsive 21st Century Civil Services” outlining the following objectives:

- Human resources development.
- Recruitment and selection.
- Information exchange.
- Research and development.

These objectives were to be met by resource sharing and through the use of ICT.^[4] For instance, in relation to human resources development, it was declared that member countries who have a “special interest in a particular field of public administration” set up ASEAN Resource Centers (ARCs) to disseminate information to other ASEAN members. The establishment of the Resource Centers has emerged as a key plank in the strategy of strengthening cooperation in public administration within ASEAN to allow the exchange of best practice within ASEAN member states.^[5] The ASEAN Covenant, “Toward Building Better Bureaucracies” formalized the establishment of the ARCs “to optimize the movement of goods and services across traditional boundaries.”^[5] What

is apparent is that the ARCs are concerned with the micromanagement aspects of public administration and policy only.

The 9th ACCSM was held in 1997 in Singapore, where it was proposed that each ACCSM was to be held on a biannual basis. Thus it was the 10th ACCSM, in 1999 in Thailand under the theme of “Good Governance: A Challenge for Economic Revitalization and Democracy Development,” that further consolidated the Covenant of the 8th ACCSM. In addition, it was the first ACCSM to be attended by all the current members of ASEAN. The conference also took place under the backdrop of the repercussions of the Asian financial crisis.

At the 11th ACCSM in Hanoi in 2001, the theme was “ASEAN Civil Services for Dynamic and Sustainable Development,” with the focus shifting to the post-1995 new member states. Both Cambodia and Laos proposed the establishment of ASEAN resource centers on Capacity Development of Civil Servants and Civil Service Performance Management, respectively. Only Myanmar has yet to submit a proposal.^a In addition, the ACCSM prescribed a role for member states with more developed administrative systems to guide others in terms of human resource development.

The theme of the last and 12th ACCSM in 2003 in Brunei was “E-Government: An Opportunity for National Development and Public Sector Modernization.” This represents macro-level cooperation within ASEAN and builds upon the creation of an e-ASEAN task force to develop a regional ICT strategy, linked to wider “E-Government” initiatives.^[7]

OBSTACLES TO COOPERATION

What is the likely outcome of closer cooperation between the ASEAN nations with regard to public administration? The sheer diversity of the nation states in question immediately militates against any generalizations. In terms of modernizing public administration and policy making within ASEAN, a strong “colonial legacy” could be a basis to predict whether a nation’s reform program will replicate those being undertaken in the West. Turner and Hulme^[8] identify the common features of the colonial legacy including a tendency toward centralization, ambiguity about the roles and relationships of public administrators and politicians, and a tradition of appointing “generalist” administrators to senior positions.

Although these features are readily identified in the majority of ASEAN countries, the colonial imprint is uneven across ASEAN. Britain colonized Brunei,

^aFor details of ASEAN resource centers, go to Ref.^[6]

Malaysia, Myanmar, and Singapore and likewise France in Cambodia, Laos, and Vietnam. The Dutch administered Indonesia until 1949, while its neighbors, the Philippines, were under the control of the United States until 1946, and Spain before that, until 1898. Thailand is the only ASEAN country never to be under the administration of a foreign power. However, Thailand's hierarchical society and bureaucratic polity, although unexposed to modernization early in its development, is typical of most countries in ASEAN. In the Philippines, U.S. colonial administration was influential, which helped some aspects of modernization but "the American emphasis on political and civil, rather than bureaucratic, action did not result in a highly trained Filipino officialdom."^[9] In Indonesia, the Dutch relied on local officials, which sided with the Dutch during the nationalist revolution, so they were politicized after independence.^[9] Overall, the "indigenization" of systems of public administration that followed colonization has served to preserve bureaucratic systems, which in turn, has made them resistant to reform.^[10]

Secondly, the elite domination of administration and politics in the ASEAN countries would suggest that ACCSM has important symbolic value. As Turner and Hulme^[8] have argued, bureaucracies dominated by technocrats tend to welcome reforms that give "legitimacy or the appearance of substance to the regime and state." Moreover, political legitimacy in the ASEAN countries requires strong economic performance; thus the appearance of administrative modernization has a strong appeal.^[10] For instance, the Philippines' policy making is "subject to the *particularistic* demands of a wealthy elite or oligarchy" supported by the bureaucratic management of the economy that compromises national economic development.^[11] The same is true of Indonesia.

In Vietnam, the major turning point for administrative reform was the Eighth Plenum of the Central Committee in January 1995, which focused on the reform of state institutions, the reform of administrative procedures, and the creation of a corps of administrative officials. It is the latter that addressed the civil service in particular as it was reported that only a minority of civil servants "were fully qualified to do their jobs... The reform involved the design of a new system and code, a new salary structure, and a new recruitment process based on examination and the retraining of old cadres."^[12] This development happened to coincide with accession to ASEAN the same year.

The newer group of ASEAN nations have had to quickly adjust to the Association's initiatives. Market reforms, implicit in the thrust of ASEAN initiatives, have also challenged the newer group in terms of bureaucratic management of the economy. Vietnam's historically bureaucratically managed economy has

been "progressively dismantled."^[13] For Laos, joining ASEAN was traumatic, as it "struggles to pay its ASEAN fees and lacks enough English-speakers and skilled diplomats to make an impression at the 300 yearly meetings of the organization and its various committees."^[14] In Myanmar, the problems are different and possibly more acute. Any colonial impact by Britain was negated by military rule, which "drove most of the trained civil officials from office" in 1962.^[9] The subsequent excessive corruption was recently exacerbated by a freeze on civil servants' salaries.^[15] Thus of all the ASEAN countries, it is probably Myanmar's administrative structure that is least developed. In a stark warning by Matthews,^[16] "there is simply no infrastructure to sustain a modern nation." Finally, Cambodia has had to respond to economic initiatives within ASEAN, which have required new institutional arrangements. This has demanded modernization of the civil service system in addition to developing appropriate human resources, developing appropriate incentive structures, and rooting out corruption.^[17] In addition, Cambodia faces similar problems to Laos when engaging with ASEAN. There is concern about a general lack of competence in English, the language of ASEAN, and the requirement to upgrade technical knowledge to be able to represent the interests of Cambodia within ASEAN.^[18]

CONCLUSION

Although ASEAN has explicitly promoted cooperation among its members in terms of policy, the overall picture tends to be one of divergence rather than convergence. This is partly explained by the political context as for most of ASEAN rule by military or civil bureaucracy or both, rather than rule by party is the norm. Convergence between such contrasting economies and political elites through strategies such as the e-ASEAN ICT task force is unlikely.^[7] However, the main objective of ASEAN, which is to enhance regional cooperation, as outlined in the Bangkok Declaration of 1967, looks more likely to be achieved through the ARCs and the dissemination of "best practice," rather than through any conscious attempt at the impossible task of converging administrative systems.

REFERENCES

1. Common, R.K. *Public Management and Policy Transfer in Southeast Asia*; Ashgate: Aldershot, 2001.
2. Halligan, J.; Turner, M. *Profiles of Government Administration in Asia*; Australian Government Publishing Service: Canberra, 1995.

3. Tay, S.; Estanislao, J. The relevance of ASEAN: crisis and change. In *Reinventing ASEAN*; Tay, S., Estanislao, J., Soesastro, H., Eds.; Institute of Southeast Asian Studies: Singapore, 2001; 3–24.
4. Civil Service Commission, Republic of the Philippines. *ASEAN Conference on Civil Service Matters*; 1995. <http://www.csc.gov.ph/asean1.html> (accessed October 2003).
5. ACCSM. *The ASEAN Covenant: Towards Building Better Democracies*. National Institute of Public Administration: Kuala Lumpur, 1995. <http://dls.intan.my/accsm/accsm.htm> (accessed October 2003).
6. <http://aseansec.org/12953.htm>.
7. Wescott, C. E-government in the Asia-Pacific region. *Asian J. Polit. Sci.* **2001**, 9 (2), 1–24.
8. Turner, M.; Hulme, D. *Governance, Administration and Development*; Macmillan: Basingstoke, 1997.
9. Crouch, H.; Morley, J. The dynamics of political change. In *Driven by Growth*, Revised Ed.; Morley, J., Ed.; Institute of Southeast Asian Studies: Singapore; M.E. Sharpe: New York, 1999; 313–354. (see previous data).
10. Common, R.K. *Public Management and Policy Transfer in Southeast Asia*; Ashgate: Aldershot, 2001.
11. Hutchison, J. Crisis and change in the Philippines. In *The Political Economy of South-East Asia*, 2nd Ed.; Rodan, G., Hewison, K., Robison, R., Eds.; Oxford Univ. Press: Oxford, 2001; 42–70.
12. Vasavakul, T. Vietnam: sectors, classes, and the transformation of a Leninist state. In *Driven By Growth*, Revised Ed.; Morley, J., Ed.; Institute of Southeast Asian Studies: Singapore; M.E. Sharpe: New York, 1999; 59–82.
13. Gates, C. The ASEAN economic model and Vietnam's economic transformation. In *ASEAN Enlargement: Impacts and Implications*; Than, M., Gates, C., Eds.; Institute of Southeast Asian Studies: Singapore, 2001; 322–361.
14. Mallet, V. *The Trouble with Tigers*; HarperCollins: London, 1999.
15. Steinberg, D. Burma/Myanmar: under the military. In *Driven By Growth*, Revised Ed.; Morley, J., Ed.; Institute of Southeast Asian Studies: Singapore; M.E. Sharpe: New York, 1999; 25–58.
16. Matthews, B. Burma/Myanmar. Round Table January **1999**, (349), 77–97.
17. Kato, T. Cambodia's accession to AFTA. In *ASEAN Enlargement: Impacts and Implications*; Than, M., Gates, C., Eds.; Institute of Southeast Asian Studies: Singapore, 2001; 164–199.
18. Zasloff, J. Emerging stability in Cambodia. *Asian Aff. Am Rev.* **2002**, 28 (4), 187–201.

Attractors, Strange Attractors and Fractals

David Ogula

Valley Stream, New York, U.S.A.

INTRODUCTION

The conduct of complex dynamic systems has been the focus of research across disciplines. The intense scrutiny of such systems has led to the unearthing of such fascinating features as attractors, strange attractors, and fractals. Examining the behavior of dynamic systems provides a sense of the direction in which the system is headed, as well as allows us to determine patterns of behavior in systems that are otherwise unpredictable. The study of complex systems, of which attractors and fractals are a component, has provided the basis for a better understanding of the nature of chaos and turbulence.

Attractors provide an understanding of the behavioral results of a complex dynamic system, while strange attractors depict the gyrations between chaos and turbulence; that is, the notion that unstable motions result in a geometrically strange object. Fractals, on their part, make it possible for us to measure and define shapes in infinite detail.

The following review provides a basic understanding of some of the well-known attractors, fractals, and associated types. It also correlates with attractors, complex behavioral patterns, observable in organizations.

ATTRACTORS

The literature on chaos theory shows that chaotic systems manifest specific bounds that are peculiar to each system. Such systems are known as attractors. Simply stated, an “attractor” is a state into which a system settles. Chaotic systems are known to dissipate, which means that in the long term, such a dissipative dynamical system may settle into an “attractor.”^[1] Attractors and associated types such as strange attractors, describe the long-term conduct of a dynamical system.

Although there is still some controversy among researchers as to the appropriate definition of an “attractor,” most scholars accept the definition of an “attractor” as “a set in the phase space that has a neighborhood in which every point stays nearby and approaches the attractor as time goes to infinity.”^[1] An attractor could also be described as “... a model representation of the behavioral results of a system.” The attractor is not a force of attraction in the system,

it simply provides a sense of the direction in which the system is headed, based on rules of motion in the system.^[2] Others define an attractor as “a set for which a positive measure of initial conditions in a neighborhood are asymptomatic to the set.”^[1]

Nonlinear systems have a wide array of attractor types—point attractor, cyclic attractor or oscillating attractor, etc. The point attractor and cyclic attractor manifest when a system’s behavior repeats itself in precisely the same manner. A point attractor behaves like a free-falling object that comes to rest at a point—the point of attraction. Thus, a steady behavioral state corresponds to a fixed-point attractor, where all trajectories beginning from the appropriate “basin-of-attraction” eventually meet at a single point. Fixed-point attractors are said to be the only possible form of attractor in linear dissipative systems.^[3] On the other hand, a cyclic attractor or oscillating attractor behaves like a yard security light that comes on or goes off according to the amount of daylight.

According to Merry, attractors can also be simple linear systems, which can be calculated precisely. Such calculations are obtainable in organizational behavior. However, organizational behavior that displays such characteristics is more typical of mechanical systems and could apply to individuals only if they are treated as machines. Merry describes five distinct behavioral stages of organizational experience, comparable to the different stages in a person’s life. Organizations experience progression from growth to maturity, through crisis, and transforms in much the same way as do people.^[4]

In the first two stages, organizational behavior is repetitive, meaning that organizational behavior repeats itself in exactly the same way. The behavior of the organization (point attractor), repeats itself like a pendulum that comes to rest at the same point or like a thermostat that maintains the temperature between two points (cycle attractor).

The study of adaptive systems has shown that organizations go through different stages of order and chaos. That is to say, organizational behavior mirrors a cyclical stage, where the organization’s behavior changes and repeats itself, passing through cycles of complex forms—the cyclic attractor.

In the third stage, which corresponds to the dissipative stage, change occurs in organizations in ways that

cause it to break apart, causing entirely new structures to emerge. The organization at this stage is seen as a dynamic system capable of reconfiguring itself.^[5] In Kiel's view, this dynamic instability allows the organization to alter its basic structures as it responds to pressures for change.^[5]

At stage four, instability in the behavior of the organization becomes more pronounced. Bifurcation and strange attractors occur—a stage where organizational behavior bifurcates into a wide array of forms. Bifurcation denotes uncertainty, unpredictability, and reduction in the ability to control. The progression of change is “from similar, uniform behavior to a variety of different behaviors . . . there is a division of dissimilar behavioral patterns.” The notion here is that people will react differently to similar circumstances. Even when people are exposed to similar situations, their behavior breaks into a wide range of possibilities. Evidence of this can be seen in individuals, departments, and teams in organizations. Differences in behavior can be caused by a small change in organizational action. For example, the introduction of a simple work procedure such as setting a specified time for turning in employee time records, in a week, can create far more complicated reactions from employees. Furthermore, disciplinary action taken by management against one employee may be a signal for other employees to take their duties more seriously, or it may reduce morale, cause anxiety, or create an atmosphere of insecurity.

Although such organizational behavior is unpredictable, it is possible to delineate the pattern that is the framework of the dynamics of such behavior over time. For instance, there are situations in which one may be aware that a company's property may be subject to theft but cannot predict when or by whom (Ref.^[4], pp. 6–8). With regard to property damage suffered by the Verizon corporation during the August 2000, workers' strike, company management may have been aware that its telephone cables may be sabotaged, but they could not have been able to predict when and by whom, specifically. Merry reinforces this point, stating that in much organizational behavior that falls within this stage, the organizational behavior and relationships are based on a chaotic pattern strange attractors. One can know that a major transformation is going to take place in the organization, but it is impossible to foretell exactly how it will affect your team or work. He notes that, though at first glance, chaotic behavior seems to be the antithesis of organizational behavior, a second look shows that the variety and irregularity allow resilience and creativity, which are necessary for organizational learning—a condition necessary for an organization to survive.^[4]

In the fifth behavioral stage, the organization moves from “chaos to a new order.” This stage is referred to

as that of “deep chaos.” At this stage, the limits of the behavioral process disappear, and randomness becomes the order of events. However, it should be noted that in adaptive systems, deep chaos is a transitional phase in the system, where the old order breaks and disintegrates, allowing a new order to emerge, which replaces the old. Merry notes that “complex systems reach this state after internal gradients have turned into deep cleavages and external perturbations have agitated the system beyond a critical point.”^[4] The state of the former Soviet Union in its last stages of break up, and in a far more catastrophic way, the breakup of Tito's Yugoslavia, illustrate this state.

Notwithstanding the complex nature of attractors, its and related concepts such as strange attractors, basin boundaries, period-doubling, bifurcations, and the like, can easily be understood by those who have no mathematical or scientific background.^[6] Another example of an attractor is a limit cycle, which is a periodic orbit that attracts. Limit cycles attract as well as repel.

Attractors also produce basin boundaries. When two or more attractors are present, basin boundaries occur. Attractors in complex dynamic systems have their basins similar to a river with a watershed basin that empties into it, and each basin has its own boundary.^[7] It is common for basin boundaries to have unstable chaotic sets. The basin boundaries are pairs that separate different basins and easily reveal unstable chaotic pairs, which wind up as basin boundaries with intricate fractal structures.^[8]

The boundary between two or more attractors in a dynamical system serves as a portal of a kind that seems to govern so many ordinary processes. One behavior that becomes evident when the system settles is that of a chaotic attractor (or an attractor that is chaotic). On the other hand, other systems can wind up as a “steady-state” non-chaotic behavior but with several steady states. Where the initial condition is specified with precision, it becomes difficult to decide in which basin it lies, especially in cases where the boundaries are fractals (Ref.^[7], p. 233). The complicated structure of basin boundaries makes it difficult to predict long-term behavior.

STRANGE ATTRACTORS

Strange attractors are a class of fractals that depicts the long-term behavior of dissipative dynamical systems. Although all strange attractors are fractals, not all fractals originate from strange attractors. When initial conditions are drawn to a special type of attractor, that attractor is called strange attractor.^[9] Strange attractors describe the gyrations between turbulence and order. A strange attractor is said to emerge in the realm

of chaos, where everything is supposed to fall apart—instead, order emerges and not chaos (Ref.^[10], p. 117). Chaos and strange attractors require non-linearity. The term “strange attractors” was introduced by David Ruelle and Floris Takens in their presentation, “On the Nature of Turbulence,” in 1970. They remarked that when periodic motion goes unstable, the typical product will be a geometrically strange object.^[11] When applied to organizations, it becomes evident that even organizations often exhibit a pattern of behavior that is difficult to predict. They exhibit a pattern of behavior that can be described as chaotic and can take the form of the strange attractor, with gyrating or irregular twists and turns that give it the name of “strange.” But, such “behavior does have a pattern that throughout time can be identified and the form does have boundaries that set limits to behavior.”^[4]

Researchers have used computers programs to simulate the evolution of strange attractors. These computer programs can display each moment of a system’s chaotic behavior as a point of light on the screen and enable one to observe how the system evolves. When observing such systems on a computer screen, one would see that the system twists and meanders unpredictably across the screen, hardly appearing in the same spot again. But, as one observes, this chaotic behavior knits into a pattern, and order emerges on the screen. The result is that the chaotic movements of the system have formed into a shape, which is a “strange attractor,” and what appears on the screen could be described as the order inherent in chaos (Ref.^[10], p. 116). On the whole, strange attractors evoke feelings of wonder and astonishment for most people who watch them.

There are various well-studied strange attractors, the most famous of which is Lorenz’s attractor. This attractor emerged from the study of equations used in the prediction of weather.

Ed Lorenz, in his 1963 presentation, determined that simple systems of three differential equations can have complicated attractors. The Lorenz attractor, which has the shape of butterfly wings, graphically illustrates the notion of sensitive dependence as a representation of chaos. Lorenz demonstrated that his attractor was chaotic, because it exhibited sensitive dependence. Moreover, his attractor was also

“strange,” which made it a fractal.^[11] The term strange attractor generally refers to most chaotic attractors. However, some argued that the term “strange attractor” should be reserved for attractors that are “geometric” strange, e.g., fractals. There are instances in which chaotic attractors are not strange.^[12]

FRACTALS

Fractals deal with shapes of infinite detail. They can be used to illustrate geographical boundaries of a nation, the bifurcation or dividing of a river’s delta, cloud formations, and so on. A fractal “allows us to define and measure accurately the property of roughness.”^[13]

Benoit Mandelbrot, who has conducted extensive research on fractals, defines a fractal as “a rough geometric shape that can be subdivided in parts, each of which is a reduced size of the whole.”^[14] Mandelbrot introduced the term fractal as a means of dealing with the problems that arise while studying scale in the real world.^[15] While exploring the iteration of complicated process equations with mathematical tools such as sines, square roots, etc., Mandelbrot devised a set of complex mathematical objects that became known as the Mandelbrot set.

To Benoit Mandelbrot, a fractal is a way of coping with problems of scales. He described a fractal as “any curve or surface that is independent of scale—a property described as *self-similarity*,” which implies that any portion of the object, if magnified, “would appear identical to the whole.”^[15] Self-similarity is an even arrangement across scale. Its images are identifiable everywhere, in culture, in the reflection of a person standing between two mirrors, or in the sketching of fish eating a smaller fish. It entails patterns within patterns. Objects that display self-similarity exhibit details in such fine scale as well as produce constant measurements (Ref.^[7], p. 103).

The notion of self-similarity in fractals is a way of “seeing” infinity, which could easily be illustrated as iterations of a scaling process as shown in Fig. 1, or by taking the middle of an equilateral triangle and attaching a new triangle that is one-third in size with identical shape. This will increase the number of points in the object thus formed, from three to six. When the

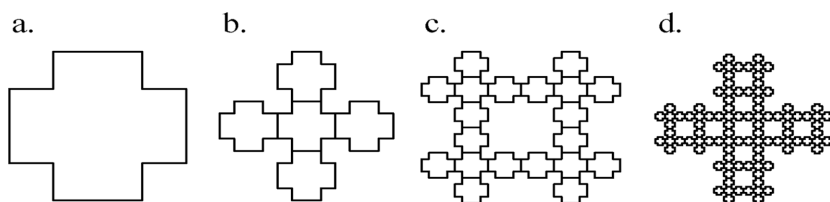


Fig. 1 Illustrates self-similarity by a simple iteration of a cross-shaped object. Each segment of the object from b to d will look identical to a, if magnified. Constructed by David Ogula ©.

process is repeated by attaching a smaller triangle, the points increase from six to 12 and so on to infinity. The resulting shape is similar to a snowflake.

Fractal shapes such as those shown in Fig. 1 and the well-known Koch curve, show self-similarity in that any portion of the object will look identical to the whole when enlarged. While Euclidean one-dimensional lines fill or occupy no space, the outline of the above shape or the Koch curve, which has a finite length crowding into finite area, fills space. In fact, a fractal literally means self-similarity (Ref.^[7], pp. 98–100).

The pioneering work of Mandelbrot and others gained wide acceptance over time, leading to the application of fractals in describing shapes, scales, and dimensions. Mathematicians, engineers, and other scholars began using fractals to graphically illustrate the capacity for “self-reference”—a key concept in the study of complex dynamic systems. In such systems, instead of spinning off in different directions, each part of the system maintains harmony with itself and with all other parts of the system as it changes, maintaining a deep relationship between individual activity and the whole.

In recent years, there has been tremendous interest in the field of fractals. They have been applied to complex physical shapes from mathematical shapes, to illustrations in the social sciences, management and administration, and even the sound of music. Some musical artists are currently using fractals to generate basic melodies in their compositions.^[16]

One of the key concepts in the study of fractals is that of “fractal dimension,” which provides a way to measure the roughness of fractal curves. As more scholars studied fractal dimensions, they began to wonder about the dimension of coastlines, mountains, and national boundaries. Mandelbrot, on his part, argued that any coastline is, in a sense, infinity or, in another sense, depends on the length of the measuring instrument.

Fractal dimension can be viewed as a way of measuring qualities that apparently have no clear definition, i.e., the degree of roughness or irregularity of an object.^[8] The dimension of a fractal curve is a number that characterizes the way in which the measured length between given points increases as scale decreases. There is evidence that fractals have also been used to store photographic quality images in a tiny fraction of the space ordinarily needed.^[17] Thinking of fractal dimensions makes one wonder how the twists and turns of a coastline could be measured (Fig. 2). For example, one begins to wonder how long is the coastline of Britain, or how long is the coast of the Eastern United States from Maine to the Florida panhandle.

If we directly measure the eastern coastline of the United States, we would find discrepancies with existing estimates of the length of the coastline. The eastern



Fig. 2 The fractal nature of a shoreline, showing the degree of irregularity of the Eastern coastline of the United States. Constructed by David Ogula ©.

seaboard of the United States, like most shorelines and national boundaries, goes up and down, with rocks, hills, cracks, and curves. A measuring device that is straight and hard like a ruler would give a value of length different from another device that is made of material that can bend and curve over such uneven surfaces. Thus, “fractals can be characterized by the way in which representation of their structure changes with changing scale.”^[16]

Another concept in the study of fractals is Fractal Geometry, which deals with “geometric forms used for building man-made objects”—made up of “lines, planes, rectangular volumes, arcs, cylinders, spheres, etc.” These objects originate from Euclidean geometry, and they “can be classified as belonging to an integer dimension, either 1, 2, or 3.”^[14]

As the literature on fractals shows, there are different types of fractals, but fractals are not just necessarily curves. There are important differences between fractal curves and the ideal curve that is normally applied to natural processes.

REFERENCES

1. Meiss, J.D. *What is An Attractor*; 1995–2000 [2.8]; <http://amath.colorado.edu/appm/faculty/jdm/faq.html> (accessed September 16, 2000).
2. Conveney, P.; Highfield, R. *Frontiers of Complexity*; Fawcett: Columbine, 1995.
3. Ilachinski, A. *Fractals* 1996; <http://www.cna.org/isaac/Glossb.htm#fractalss> (accessed September 11, 2000).
4. Merry, U. *Nonlinear Organizational Dynamics: The Scientific Revolution*; <http://pw2.netcom.com/~nmerry/art2.htm> (accessed January 10, 1998).

5. Kiel, L.D. *Managing Chaos and Complexity in Government*; Jossey-Bass: San Francisco, 1994; 9–15.
6. Lorenz, E.N. *The Essence of Chaos*; University of Washington Press: Seattle, 1993; pp. ix, 5.
7. Gleick, J. *Chaos, Making a New Science*; Penguin Books: New York, 1988; 95–235.
8. Grebogi, C.; Ott, E.; Yorke, J.A. Chaos, strange attractors, and fractal basin boundaries in non-linear dynamical systems. *Science* **1987**, *238*, 585.
9. Sprott, C.J. *Strange Attractors—Creating Patterns in Chaos*; The University of Wisconsin: Madison, WI, 1999; 5–7.
10. Wheatley, M.J. *Leadership and the New Science—Discovering Order in a Chaotic World*, 2nd Ed.; BK Publishers: San Francisco, 1999; 117 pp.
11. Meiss. *What is a Strange Attractor* [2.12]; <http://amath.Colorado.edu/appm/faculty/jdm/faq-2.html> (accessed September 16, 2000).
12. Elert, G. *Strange Attractors*; 1997–2000; <http://hypertextbook.com/chaos/21.shtml> (accessed October 6, 2002).
13. Ramsden, S. *Fractals, Feedback and Chaos—A Brief History*; <http://online.anu.edu.au/ITA/ACAT/contours/docs/fractals-history.html> (accessed September 11, 2000).
14. Burke, P. *An Introduction to Fractals*; 1991; <http://astronomy.swin.edu.au/pbourke/fractals/fracintro/> (accessed September 6, 2000).
15. Green, D. *Fractals and Scale*; 1995; <http://life.csu.edu.au/complex/tutorials/tutorial3.html> (accessed September 15, 2000).
16. Swickape. *Fractals, Chaos and Music*; <http://www.geocities.com/SoHo/Square/7921/Math.html> (accessed September 15, 2000).
17. Connors, M.A. *Exploring Fractals*; 1994–1997; <http://www.math.umass.edu/~mconnors/fractals/intro.html> (accessed September 21, 2000).

Auditing

Jean Harris

*Department of Accounting, Pennsylvania State University at Harrisburg,
Middletown, Pennsylvania, U.S.A.*

Ann Pushkin

*College of Business and Economics, West Virginia University,
Morgantown, West Virginia, U.S.A.*

INTRODUCTION

Auditing is fundamental to providing assurance that information is creditable, and creditable information is fundamental for accountability. An audit is an examination. Although there are many types of audits, this explanation is primarily about financial statement and financial-related audits of state and local governments. In these audits, assertions are examined on a test basis, and accounting principles and estimates are assessed. Based on the collection and evaluation of evidence, the auditor renders an opinion about the financial statements and, for governments, sometimes issues additional reports. Fair presentation of financial statements is judged relative to established criteria set forth in generally accepted accounting principles (GAAP). For many governments, additional reports address the adequacy of internal controls and compliance with policies, contractual agreements, grant provisions, regulations, and statutes.

AUDITS

Purpose of Audit and Types of Opinions

A financial statement audit provides assurance that an entity's financial statements are presented fairly. The financial statements are prepared and owned by the entity being audited or auditee, and embody the assertions of that entity. By applying appropriate audit processes and observing auditing standards, the auditor evaluates the auditee's assertions relative to established criteria. From the assurance provided by the auditor of fair presentation, the users of reports may hold the auditee accountable.

From conducting an audit, the auditors may disclaim an opinion on the financial statements or issue one of three types of opinions: unqualified, qualified, or adverse. An opinion is disclaimed when an auditor does not have sufficient evidence to express an opinion. An unqualified opinion indicates that financial statements are fairly presented in accordance with GAAP.

If an entity's record of financial performance is poor, it may receive an unqualified opinion if its financial statements fairly reflect its poor financial performance. Many rules guide the distinction between a qualified opinion and an adverse opinion. In general, a qualified opinion indicates that, except for a deviation from GAAP, the financial statements taken as a whole are presented fairly. An adverse opinion indicates the financial statements taken as a whole lack of fair presentation.

Responsibility for Fraud

To reach an opinion, the auditor must plan and execute a financial audit to obtain reasonable assurance that management's assertions are free of material misstatement, whether due to error or fraud. Even when audits are executed in accordance with due professional care, there are no guarantees that all fraud will be discovered, particularly when management sets out to deceive the auditor, and others. However, if an allegation of fraud has been made, a fraud audit should be performed; this is a separate type of engagement from a financial audit.

Concept of Entity

The concept of entity is central to financial reporting and auditing. The bounds of entity establish the activities reported in financial statements. State and local governments are familiar reporting entities. The financial statements of a governmental reporting entity include the activities of the primary government, such as a city, and of component units that have financial accountability to the primary government, such as a city school board. All the activities reported in general purpose financial statements (GPFS) are subject to examination under a financial statement audit.

Reporting entities and accounting entities may have different bounds. Any entity for which a distinct set of financial statements is prepared is an accounting entity. Thus, a discretely presented component unit

or a fund, such as the general fund, is an accounting entity. In contrast, the GPFS of a reporting entity may include data from many accounting entities.

A reporting entity's GPFS may be issued separately or as part of the entity's comprehensive annual financial report (CAFR). Supplemental and statistical information in the CAFR, that is additional to the financial statements, is not subject to audit.

AUDIT ACTIVITIES

Differences Between Auditing and Accounting

Although financial statement auditors must understand accounting, it is the reporting entity that is responsible for its accounting and preparation of its financial statements in accordance with GAAP. For state and local governments, the primary authoritative bodies for promulgating GAAP are the Governmental Accounting Standards Board and the Financial Accounting Standards Board.

Because auditors assess the fair presentation of financial statements, they must have a sound knowledge of GAAP. But unlike accountants, auditors also must have a strong knowledge of auditing standards and procedures that set forth the means for conducting an audit.

Types of Audits

Generally accepted governmental auditing standards (GAGAS) address two types of audits: financial audits and performance audits. Financial audits include financial statement and financial-related audits. Financial statement audits pertain to the fair presentation of financial position, results of operations, and cash flow in conformity with GAAP. In contrast, financial-related audits pertain to the disclosure of information, adherence to compliance requirements, and suitability of internal controls.

Performance audits include economy and efficiency audits and program audits. The objectives of an

economy and efficiency audit are to assess the economical and efficient acquisition, protection, and use of resources; to identify inefficient and uneconomical practices; and to determine if the entity has complied with relevant laws and regulations. In contrast, the objectives of a program audit are to evaluate the extent to which desired results and benefits are achieved, to assess the effectiveness of operations, and to determine compliance with applicable laws and regulations.

Performance audits are directed to evaluating efficiency and economy of programs or to evaluating the effectiveness of program results. Thus, auditees are frequently selected on the bases of potential for saving costs or improving results. Legislative bodies often have agencies that perform performance audits at the request of members or committees. At the federal level, the General Accounting Office (GAO) serves this function. In addition, the GAO includes standards for performance audits in *Government Auditing Standards* (Yellow Book). Table 1 shows the ways that performance audits differ from financial statement and financial-related audits.

Audits in Public and Private Sectors

Public sector refers to governments at all levels. Audited financial statements and information, along with the independent external auditor's report, contribute to the accountability imposed on government officials for the resources entrusted to them. Public officials, legislators, creditors, and citizens use the audited information and audit reports to determine whether government assets are properly accounted for and have been expended in compliance with relevant laws and regulations. Also, in the public sector, performance audits indicate whether government organizations, programs, and services are achieving their purposes and are operating economically and efficiently. Accordingly, information and audit reports for public sector auditing have a much broader scope than in the private sector. Results of both financial audits and performance audits in the public sector are available to the public.

Table 1 Performance audits

Performance audits	Financial statement and financial-related audits
Audits are preformed at irregular intervals	Audits are performed annually
Assertions may be vague or undefined	Assertions are well defined
A broad range of evidence is considered	Well-defined types of evidence are sought
The focus tends to be program specific	Financial statements reflect the entity as a whole
Auditors tend to have diverse educational backgrounds with strong knowledge of programs	Auditors tend to have strong backgrounds in accounting

Source: From Ref.^[3].

Most large organizations have an office or department of internal audit. Internal auditors are employed by the auditee, and generally report, by separate channels, to the highest level of management or administration within the auditee's organization. Internal auditors are expected to maintain an objective perspective and to provide assurance regarding the reliability of systems and data within the auditee's organization to its top officials. In contrast, external auditors work for public accounting firms or for special audit agencies, such as the office of a State Auditor or, at the federal level, the GAO. Internal and external audit functions are linked. The presence of an internal audit function serves as an internal control to enhance the reliability of information. Thus, a strong internal audit operation tends to reduce the level of assessed risk and accordingly reduces the amount of audit resources required to conduct an external audit.

Private sector refers to business entities. However, the only businesses required to have financial statement audits by external independent auditors are those businesses that trade securities on an organized stock exchange and are regulated by the Securities and Exchange Commission (SEC). These businesses are called public companies because their stock trades on an exchange that is open to the public. Hence, the term, public companies, has a meaning distinct from public sector governmental entities. Likewise in the professional designation, certified public accountant (CPA), public refers to holding one's self out to render services to the public rather than being an employee of a public sector governmental entity. Results of financial statement audits of public companies within the private sector are public information and are used primarily by current and potential stockholders, financial analysts, creditors, and customers for business decision making. Regardless of sector, financial statement audits attest to the fair presentation of information rather than to the quality of management's decisions or the soundness of an investment.

Conduct of Audits

There are three basic phases in a financial audit—the planning stage, the execution of the plan along with evaluation of results, and report writing. Developing an audit plan is an extremely complex task requiring knowledge of prior audits and a thorough understanding of the auditee's operation and environment. With this knowledge, the auditor assesses the inherent or business risks associated with the auditee and structures an audit plan based on assessed risk.

In addition, the plan addresses any uncorrected recommendations made by the auditor in prior years for known material findings. Once the initial plan is

established, the auditor executes the plan and documents all evidence as a basis for drawing conclusions. As the audit progresses, each piece of new evidence is evaluated with respect to the audit objectives of the relevant procedure. Such evidence is compared with other evidence and conclusions drawn in preceding audit procedures. When this comparison produces conflicting evidence, resolution of the conflict may require an expansion of the initial audit plan. Thus, auditing is a reiterative process. When all steps in the plan are completed and the auditor is satisfied with the sufficiency and competency of the evidence to support conclusions, a report is written and communicated to the auditee.

SINGLE AUDITS

Purpose

Any government or non-profit organization that expends over a threshold amount in federal awards from more than one program must have a single audit. The single audit is not distinct from a financial statement audit, but is an extension to it. Thus, the organizationwide single audit serves as a basis of opinion on financial statements and as a basis for judging that federal awards are used in an appropriate manner in accordance with contractual and legal requirements.

The single audit approach was adopted as a way to improve financial management, conduct uniform audits, promote efficient and effective use of audit resources, and ensure reliable audits. Rather than being audited on a grant-by-grant basis for each separate grant, recipients are subject to a single organizationwide audit. This one single audit satisfies all federal agencies. Thus, the duplication of having many agencies separately auditing a common recipient on a grant-by-grant basis is avoided.

Standards and Reports

Single audits are conducted in accordance with generally accepted auditing standards, GAGAS, and single audit standards. [Table 2](#) indicates the standards, the standard setters, and primary source of guidance issued by standard setters. The American Institute of CPAs (AICPA) is a private national professional organization, the Office of Management and Budget (OMB) is a federal executive agency, and the GAO is a federal legislative agency. In addition to its roles in establishing GAGAS and conducting other activities, the GAO is responsible for audits of the financial statements of federal agencies.

Table 2 Auditing standards

Standards	Standard setter	Source of guidance
GAAS ^a	AICPA ^c	<i>Statements on Auditing Standards</i> (SASs); AICPA: New York, 2001. <i>Audits of States and Local Governmental Units</i> (audit guide); AICPA, New York, 1998.
GAGAS ^b	GAO ^d Comptroller General of the United States	<i>Government Auditing Standards</i> (Yellow Book); GAO: Washington, DC, 1994.
Single audit standards	OMB ^e	<i>OMB Circular No. A-133, Audit of States Local Governments, and Non-Profit Organizations</i> ; OMB: Washington, DC, 1997.

^aGAAS, generally accepted auditing standards.

^bGAGAS, generally accepted governmental auditing standards.

^cAICPA, American Institute of Certified Public Accountants.

^dGAO, General Accounting Office.

^eOMB, Office of Management and Budget.

Source: From Ref.^[4].

Under the rules for a single audit, six reports must be issued in addition to the auditor's basic opinion on the financial statements. The term, financial-related auditing, refers to the work that supports the additional reports required by a single audit. In contrast to the opinion on financial statements that is directed to the financial statements taken as a whole, the additional reports required in a single audit tend to be directed to major federal financial assistance programs. The extra reporting requirements of the single audit focus on the adequacy of internal controls and compliance with contractual, regulatory, and statutory requirements. Internal controls refer to the plans, procedures, and actions of an entity to ensure compliance with its policies, and with regulations and laws to which it is subject; to enable the evaluation of operations and personnel; to ensure reliable data; and to protect its assets. Although a single audit does not require testing for illegal acts, the auditor must report any illegal acts that are discovered. The auditing standards that apply to specific reports are shown in Table 3.

The Office of Management and Budget first implemented the concept of the single audit in 1979. Subsequently, it was adopted by statute with enactment of the Single Audit Act of 1984 and then amended in 1996. Single audits are administered by OMB, which issues circulars that provide guidance for performing audits. By statute, single audits must be conducted in accordance with GAGAS of GAO. Periodically, GAO issues amendments that are codified as part of its *Government Auditing Standards* (Yellow Book).

REGULATION OF AUDITS

Public Interest in Regulation

The strong public interest in auditing derives partially from a lack of direct economic exchange and, in a governmental setting, partially from the need for stewardship over public funds. In the acquisition of audit services, there is no direct exchange between the

Table 3 Single audit reports

Focal report	Applicable standards		
	GAAS ^a	GAGAS ^b	Single audit standards
Financial statements	X	X	X
Entity's internal control structure		X	X
Entity's compliance with laws and regulations		X	X
Schedule of expenditures of federal awards			X
Internal controls over federal financial assistance			X
Major federal financial assistance program			X
Compliance schedule of prior audit findings			X

^aGAAS, generally accepted auditing standards.

^bGAGAS, generally accepted governmental auditing standards.

Source: From Ref.^[4].

auditor and the external user of audit reports. It is the auditee who employs and pays the auditor rather than the external users of reports. Thus, to accept the creditability of audit report, the user must have a high level of confidence in the auditors, the standards they observe, and the procedures they employ. The public interest in stewardship is evidenced by the many single audit requirements that emphasize compliance with policies, regulations, and statutes. Report users who are dissatisfied with the work of an auditor may make their dissatisfaction known to the auditee, pursue a complaint with a state board of accountancy or federal oversight agency, or bring suit against the auditor in court. For a discussion of the role of auditing in society, see *The Audit Society*.^[2]

Stakeholders

There are many stakeholders with an interest in government audits. Two of the primary stakeholders are parties who participate in the municipal bond markets and federal agencies that disburse funds to auditees. Present and potential investors in municipal bonds are concerned about the receipt of bond interest and ultimate repayment of bond principal. These investors rely heavily on the judgment of bond raters who review financial statements, audit reports, and a lot of other information in making their evaluations. Federal agencies are concerned that federal funds expended by state and local governments are spent in appropriately authorized ways. Other stakeholders with an interest in financial information and audit reports include public administrators, legislators, taxpayers, and citizens, in general.

Role of Independence

Independence of the auditor is critical to the audit. The independent auditor must be independent with respect to financial connections and to managerial decision making. In particular, the independent auditor must not be involved with the auditee in a way that enables the auditor to make managerial decisions that subsequently are examined under audit. If the auditor is not independent from the auditee in mental attitude and in appearance, the creditability of the audit report and of the audited financial information is so severely damaged that neither the report nor financial statements may be regarded as legitimate.

The core of all independence requirements is that the external auditor be financially independent from the auditee and not involved in managerial decision making for the auditee. Auditors must comply with rules, regulations, and laws adopted by various authorities. These authorities include the GAO, state boards

of accountancy that license CPAs, the AICPA, and the SEC. As required by GAGAS, CPAs and other professionals who audit under the Single Audit Act must be free from personal and external impairments with the auditee, and must maintain both an independent mental attitude and independence in appearance.

Additional independence requirements apply to CPAs. When CPAs are performing an audit or other attest engagement with respect to financial statements, they must comply with the laws and regulations of the relevant state with regard to independence of the auditor from the auditee. For CPAs who are members of the AICPA, they must also comply with its rules regarding independence. When a CPA audits a business that is under the regulatory authority of the SEC, the CPA is said to practice before the SEC. To practice before the SEC, CPAs must show that they are in good standing with their state board of accountancy and in compliance with SEC rules regarding independence.

QUALIFICATIONS AND CERTIFICATION OF AUDITORS

Education of Auditors

Independent audits of financial statements may be performed by CPAs or by persons employed by independent audit agencies of governments. Most independent auditors have a Bachelor's degree in accounting or its equivalent. Maintaining a license as a CPA generally requires the completion of continuing professional education (CPE) on an annual basis. Specific CPE requirements are imposed for all auditors who participate in single audits.

Certification of Auditors

All CPAs are certified and licensed by state boards of accountancy. The requirements for certification vary among the states and may be obtained from the National Association for State Boards of Accountancy or a state board of accountancy. To be certified, an auditor must meet an educational requirement to take the CPA exam and pass the CPA exam. Many states have gone beyond requiring a bachelor's degree to require a minimum of 150 credit hours to sit for the CPA exam. An additional requirement for licensing is to obtain relevant experience of a certain type and duration. Maintaining a license then requires the completion of CPE.

Oversight and Peer Review

Auditors conducting a single audit must provide federal oversight officials with access to their working

papers that document the conduct of the audit. To ensure and promote quality work, members of the AICPA who have clients regulated by the SEC are subject to the peer review processes of the AICPA. In addition, some states require a peer review for licensing of CPAs.

TRENDS IN AUDITING

Influence of Information Systems

In today's environment, most financial information and statements are produced using complex electronic information systems. Thus, auditors must assess the reliability of these systems before accepting evidence from such systems as a basis for an opinion on the fair presentation of financial statements. That is, the validity of substantive evidence can be only as reliable as the system that produces such evidence. Effective systems must have physical and logical access controls to programs and data, processing controls, change controls, maintenance policies and procedures, and authoritative oversight. Accordingly, financial auditors must be technologically competent with respect to computerized financial systems and frequently must be able to engage information systems auditors to conduct effectiveness tests of system controls. Evidence on the effectiveness of the system will guide the financial auditor in planning the audit and in gathering substantive evidence on which to base an opinion on the fair presentation of financial statements.

Continuous Audits

Continuous audit refers to procedures and practices employed on a continual basis throughout the year

that are intended to provide ongoing assurance to decision makers about the quality of information they are receiving. Traditional financial statement audits are performed annually with much of the audit work done after the auditee has prepared a set of financial statements. On completion of the audit, a report or set of reports are issued at one point in time about one set of annual financial statements. In today's environment, the annual audit supplies a small part of the total information available to decision makers. However, most of the other information available to decision makers is not audited by independent auditors and is not regulated by authoritative bodies. Accordingly, regulators and standard setters believe that there is a demand for more timely information than that provided by the annual audit of annual financial statements. In addition, decision makers most likely will expect independent external auditors to provide continuous assurances on both financial and non-financial data and on the effectiveness of the systems that produce the continuous flow of information. For more information about this new approach to auditing, see *Continuous Auditing*.^[1]

RESEARCH METHODS AND ISSUES

Methods

Various methods are employed in the conduct of auditing research. Subjects tend to be auditors, auditees, and users of audit reports. Data are often collected from subjects by means of experiments, surveys, and interviews. Other rich sources of data are archival stores of financial statements, audit opinions, regulatory reports to the SEC, and stock and bond market information. Much research centers around one or more

Table 4 Auditing-related web sites

Abbreviation	Name	Web site
AAA	American Accounting Association	www.AAA-edu.org
AGA	Association of Government Accountants	www.agacgfm.org
AICPA	American Institute of CPAs	www.aicpa.org
FASB	Financial Accounting Standards Board	www.fasb.org
GAO	General Accounting Office	www.gao.gov
GASB	Governmental Accounting Standards Board	www.gasb.org
IIA	Institute of Internal Auditors	www.theiia.org
NASACT	National Association of State Auditors, Comptrollers and Treasurers	www.nasact.org
NASBA	National Association of State Boards of Accountancy, Inc.	www.nasba.org
NYSSCPAs	New York State Society of CPAs	www.nysscpa.org
OMB	Office of Management and Budget	www.omb.gov
SEC	Securities and Exchange Commission	www.sec.gov

hypotheses that reflect a possible association between or among variables. It is common to employ statistical methods in testing for associations.

Issues

The relationships suggested by hypotheses often connect auditing to the disciplines of psychology, economics, and finance. Some of the dominant questions in auditing research are as follows: What variables influence the behavior and/or judgment of auditors, of personnel and directors of the auditee, or of audit report users? What is the nature of the market for audit and other assurance services? What variables influence market pricing, the quality of audit services, and the response of financial markets to audit opinions? How effective are various audit procedures, regulatory processes, and management practices within auditing firms? Many accounting-related professional and academic periodicals publish auditing research. *Auditing: A Journal of Practice and Theory* (American Accounting Association), an academic journal, disseminates auditing research exclusively. Four important practitioner journals are *The CPA Journal* (New York State Society of CPAs), *Internal Auditor* (Institute of Internal Auditors), *The Journal of Accountancy* (AICPA), and the *Journal of Government Financial Management* (Association of Government Accountants).

Influence of Internet

The Internet has affected the performance of audits, expanded research resources, and introduced new means to improve the proficiency of auditors. Using the Internet during the conduct of audits, auditors may rapidly exchange information among themselves and with the auditee. Government policies, programs, and publications may be researched using the Internet. Many academic and professional periodicals are available electronically and may be searched easily. As a means of resolving problems, auditors may participate in discussion groups and listservs. One issue relating to the information via the Internet is whether the source is reliable and verifiable. Another issue is the confidentiality of transmission for certain information. Auditing-related web sites are listed in [Table 4](#).

REFERENCES

1. Canadian Institute of Chartered Accountants (CICA). In *Continuous Auditing*; CICA: Toronto, 1999.
2. Power, M. *The Audit Society*; Oxford University Press: Oxford, 1997.
3. Granoff, M.H. *Government and Not-for-Profit Accounting*, 2nd Ed.; John Wiley & Sons, Inc.: New York, 2001.
4. American Institute of CPAs (AICPA). In *Audits of State and Local Governmental Units*; AICPA: New York, 1998.

Balanced Scorecard for Public Administration and Policy

Peter J. Reed

Bannock Consulting, London, U.K.

INTRODUCTION

In an influential series of articles in the *Harvard Business Review*^[1–3] and best-selling books from Harvard Business School Press during the 1990s^[4,5] Robert S. Kaplan and David P. Norton introduced and developed the idea of the Balanced Scorecard. The essence of their argument was that financial accounting measures are no longer sufficient as true indicators of the worth or the health of companies or their corporate performance. Most financial measures were developed to measure the effectiveness of investment in tangible physical resources and assets. In today's information economy, and particularly in what can be termed "knowledge-based organizations," they are inadequate indicators to account for the true value (or growth in such value) of essential intangible assets such as customer satisfaction and loyalty, employee skills and motivation, and intellectual or information capital.

Most companies, and probably all public sector organizations, already monitored nonfinancial measures, but in general, Kaplan and Norton's contention was that managers tended to focus more heavily on aggregate financial measures as if, somehow, these could accurately reflect the efficiency or effectiveness of operations performed by staff at all levels of the organization. The Balanced Scorecard was designed, therefore, not just to introduce the importance of "balance," but also to draw attention to the usefulness of nonfinancial measures that inevitably drive long-term success. By the same token, employees that are exposed to the concept (for example, by participating in organizational self-assessment workshops, or even through the publication of indicators, targets, and results) will be able to see the financial consequences of their actions—especially those geared directly to the improvement of internal processes.

Public sector organizations necessarily have a different perspective concerning the customer value proposition. Their "stakeholders" are normally taxpayers and their stewardship of public money has to be with a view to maximizing efficiency and demonstrating this good management of resources through transparency and public accountability. There is no profit motive, and no requirement or expectation to produce an increase in share value, or a return on investment in conventional corporate terms. They were, therefore,

among the early adopters of the Balanced Scorecard, recognizing that it could provide both a strategic performance measurement system and a framework to capture and summarize results for publication. Having in many cases already espoused Quality Management principles, and award schemes for quality standards like Investors in People, they were quick to realize that this was a useful and complementary addition to their "toolkit" of methodologies for managing continuous improvement of service delivery.

THE ORIGINAL FOUR PERSPECTIVES AS APPLIED TO THE PUBLIC SECTOR

Learning and Growth

This perspective looks at the ability of public employees, the quality of information systems, and the effects of organizational alignment in supporting accomplishment of organizational goals for excellent service delivery. Processes will only succeed if adequately skilled and motivated public employees, supplied with accurate and timely information, are implementing them. This perspective takes on increased importance in public service organizations that are undergoing radical change. Even learning to regard the public as valued clients or customers has necessitated a major paradigm shift in many countries and cultures, both at the national and at the local levels of service delivery. In order to meet these changing requirements and customer expectations, employees may be asked to take on dramatically new responsibilities, and may require skills, capabilities, technologies, and organizational designs that were not available before. Many senior civil servants did not even think of themselves as managers, let alone "leaders" until quite recently.

Internal "Business" Processes

This perspective focuses on the internal process mechanisms, and the results of these that lead in turn to financial viability and satisfied customers just as they do in the private sector. To meet organizational objectives and customers' expectations, organizations must identify the key "business" processes at which

they must seek to excel. Key processes should then be (and now frequently are) monitored to ensure that outcomes will be satisfactory. Internal operational processes and the efficiency of their performance are the mechanisms through which the public's expectations concerning standards of service delivery are achieved. The problem often arises in civil and public services that the processes are perceived as being set in stone, only relevant internally, and not connected to customer convenience. This malaise is not, of course, confined to the public sector. In fact, in the public sector they are often set in print in weighty volumes with archaic and bureaucratic systems for annual amendment, which may or may not take place. Ministerial or departmental mandates, protocols, regulatory frameworks and procedures, are frequently tied up in "red tape." This is usually just because they always have been, and employees have traditionally been expected to see these "standard operating procedures" as somehow sacrosanct, rather than to question them or suggest improvements for greater efficiency or customer satisfaction.

Customer, Client, or User Focus

This perspective captures the ability of the organization to provide quality standards of services, effectiveness of their delivery, and overall customer convenience and satisfaction. In the government or public service model, the principal driver of performance is different from that in the strictly commercial environment. Here, customers and stakeholders take pre-eminence over financial results. This is not to say that financial prudence is not fundamentally important. In general, public organizations have a different, perhaps greater, stewardship, fiduciary responsibility, and focus than do private sector entities. Financial responsibility for sound fiscal and expenditure management policy and strategies should, of course, be taken seriously by bureaucrats. The public as stakeholders (like share- or stockholders in this respect) expect it; but it should never be at the expense of the quality of the services that are provided within budgetary constraints.

Ensuring that policies, strategies, and operations have a positive impact on society is, of course, more obviously a prime consideration for the public services whose very existence and *raison d'être* is predicated on this. Finding metrics by which to measure this, and the trend over time, should, therefore, be a concern for leaders in all such organizations. Client Service (or Citizens') Charters have gone a long way in demystifying the language of policy, and to clarify just what government is offering its stakeholders and to what standard they have a right to expect the delivery of services.

The Balanced Scorecard can be used effectively to benchmark these standards and enable political leaders to report honestly to the public on progress with policy intentions.

Financial or Key Performance Results

In the government arena, the "financial" perspective therefore differs from that of the traditional private sector. Private sector financial objectives generally represent clear long-range targets for profit-seeking organizations, operating in a purely commercial environment. Financial considerations for public service organizations have an enabling or a constraining role, but will rarely be the primary objective for "business" systems. Success for public service organizations should be measured by how effectively and efficiently they meet the needs of their constituencies. Therefore, in the government arena, the financial perspective emphasizes cost efficiency, i.e., the ability to deliver maximum value to (and as perceived by) the customer.

THE MALCOLM BALDRIDGE (USA), EFQM EXCELLENCE (EU), AND MODERN COMPTROLLERSHIP (CANADA) MODELS AND FRAMEWORKS

The Malcolm Baldrige Quality Award scheme in the United States and the Excellence Model from the European Foundation for Quality Management (EFQM)^[6] appeared on the scene at around the same time as Kaplan and Norton's publications. Indeed, the Malcolm Baldrige Award was first introduced in 1987, and so predated these. Rather surprisingly there is very little in the literature—either from EFQM or from the Malcolm Baldrige Award Foundation (or the Baldrige National Quality Programme—BNQP, or the National Institute of Standards and Technology—NIST) to acknowledge Kaplan and Norton's contribution, or vice versa; so it is difficult to be certain about the original source of the DNA.

In any event, both EFQM and the Baldrige movement in America quickly introduced refinements of the scorecard with more criteria—differentiating between a wider range of Enablers that drive improvement, and Results that augment measures of purely financial performance. Currently, the Baldrige model has seven criteria, and the EFQM Excellence Model has nine.

The Baldrige model's criteria or categories are:

- Leadership
- Strategic planning
- Customer and market focus
- Measurement, analysis, and knowledge management
- Human resources focus

- Process management
- Business results

The EFQM Excellence Model's criteria are divided into five Enablers and four Results Areas as follows:

Enablers

- Leadership
- Policy and strategy
- Human resources management
- Resources (and partnerships)
- Processes

Results

- Customer results (satisfaction)
- People (employee) results (satisfaction)
- Impact on society
- Key performance results (including financial performance)

BUILDING A SCORECARD

Public service organizations—just like their private sector counterparts—can choose whether to use one of the tried and tested framework models, whether to populate Kaplan and Norton's original four perspectives with appropriate measures for their operating environment, or whether to design their own scorecard based on the essential fundamental principles of balance, interdependence, and cause-and-effect linkages. If they decide to use either of the Baldrige or EFQM models (or the Common Assessment Framework derived from the latter) they will find a wealth of material available freely in the public domain, most notably prompts for a comprehensive list of self-assessment questions for management teams to ask themselves when assessing their current performance under the criteria headings. Managers should be aware, however, that a balanced scorecard is potentially as important for informing and mapping strategy (summarizing this diagrammatically in concise, convenient form) as it is as a benchmarking instrument.

If an organization's management team decides to use the original four-category model for their scorecard they need to do more than choose some measures and sort them into the four perspectives. Measures should be chosen in order to inform what are considered to be the key elements of strategy. Kaplan and Norton emphasize that the measures must also fit together, identifying three important principles that should underpin the design:

1. *Cause and effect*: The measures must relate to each other and tell a story that shows how

activities to enhance individual and organizational learning and growth can improve internal processes. This will strengthen the customer value proposition (leading to customer satisfaction), which, in turn, should ensure that financial and other key indicators of performance for achievement of the organization's mission are met (the financial or key performance results perspective).

2. *Lead and lag measures*: Two types of indicators should be chosen. Measures of customer satisfaction, for example, are useful but are after-the-fact indicators that do not necessarily tell you how to get better. To back up such *lag* measures, *lead* indicators, such as response times, quality measures, participation rates in staff training programs, and so on, are necessary because they cause, lead to, or predict customer satisfaction.
3. *Financial effectiveness*: Even in the public sector where "return on investment" is not measured in the same way as it is in business, managers (and ultimately politicians) need to be able to demonstrate that "return on assets employed" is as efficient, and the strategies employed as effective, as they can be to offer best value for money to the taxpayer.

In Canada the Treasury Board of Canada Secretariat has developed Modern Comptrollership. This is a management reform and philosophy involving public service values, sound risk management, integrated performance information, and appropriate control systems. The first step is to establish a baseline using a self-assessment framework diagnostic tool "Modern Comptrollership Capacity Assessment." The model for this is slightly more sophisticated than the CAF and the process is designed to be more time consuming. It has seven key criteria for assessment, for which the methodology provides detailed descriptions of awareness and actual management behavior reflecting five standards for approaches and results. These seven key areas (which, in turn, break down into 33 subcriteria) are:

<i>Core criteria</i>	<i>Enabling criteria</i>
Shared values and ethics	Strategic leadership
Mature risk management	Motivated people
Integrated performance information	Clear accountability
Rigorous Stewardship	

The Treasury Board's information in the form of a web brochure insists that this Capacity Assessment "is not a scorecard." This seems a strange assertion,

because, to all intents and purposes, it is, and a very good one too.^[9]

MULTIPLE SCORECARDS

Many management teams will quickly see the benefit of having at least one Balanced Scorecard to facilitate high-level strategic planning and performance benchmarking. If the organization is large enough, “federal” in structure, or decentralized, it may well be seen as appropriate for individual departments, units, or branches to have their own scorecards. In such cases, these should, of course, either be logically cascaded from the higher-level corporate scorecard, or the latter should be distilled from aggregation of the essential elements and outputs of such departmental tools.

INTANGIBLE ASSETS

In this context, it is important to stress that an organization’s intangible assets (such as knowledge and information) are at least as important as its tangible assets and probably more so. Kaplan and Norton contend that the secret of long-term value creation is the conversion of intangible assets into tangible outcomes by describing and measuring the critical few parameters that represent the strategy and managing these. In their view, it is important to attempt to identify measures for human capital, information capital, and organization capital, which describe exactly how the Learning and Growth Perspective can be strengthened.^[7,8] In aggregate, these could be seen as an “intellectual balance sheet” the “worth” of which is likely to be far more influential for value creation in a public service organization than bricks and mortar, desks, computer hardware, fleets of vehicles, sheer numbers of employees, or even financial resources per se.

Together, measures for these aspects of organizational capital constitute what we could call organizational readiness—for change, for the challenges of measuring performance improvement, and for achievement of the mission. Understanding and knowledge about these must be pushed down and out into the organization because the people who do the work need to know not only what they should be doing, but also to what standard, how they will know this, and how they will be recognized or rewarded for making the extra effort required.

There needs to be an understanding of the “flows” around a scorecard from box to box. To be effective it *must* be used to manage and not just to monitor. In the current climate this is what makes it difficult for companies (and perhaps even the public sector as well) to justify “returns on investment” (or efficiency savings)

that are less rapid or dramatic in scale than the investors (or politicians or taxpayers) had hoped for or expected. Nevertheless, many visionary firms do indeed outperform share price indices and always have done so by engaging in long-term capability building, so it is naive to suggest that markets simply do not allow companies and corporations to build for the future because of an all-pervasive short-term profits and earnings focus. By the same token, it should be possible for public service organizations to maintain commitment to the medium-term strategy (if this is well founded on the need for increased learning and growth) despite the pressure of continual short-term “value for money” assessments and audits.

One of the reasons that effective leadership is so important in such examples is that it is easy to manage by the scorecard in relatively easy times; but when the going gets tough, it takes leaders with real and extraordinary commitment and belief not to fly back to the financial box and do whatever is necessary to make savings or keep results good in the short term. It can help if the overall performance reward framework can be linked to all stakeholders in the scorecard (not just shareholder or financial stakeholder value, although using the scorecard should ultimately drive through to better results here too).

APPLYING A PROVEN METHODOLOGY

While it is perfectly possible to design your own bespoke scorecard, many organizations will prefer to adopt and perhaps adapt a tried and tested methodology. This approach will usually ensure the capture (and translation to measurement when required) of an organizational strategy across a wide variety of strategic and operational situations, as the models (like the basic scorecard) are essentially generic. The organization’s management will need to consider at this point whether they wish to use the balanced scorecard alone as their principal model to develop overall strategy (to be captured in a strategic or business plan). They may prefer to use the scorecard to ensure a comprehensive and holistic systems overview, and to identify areas for improvement, using a more conventional strategic planning model with which they may already be familiar to formulate the main strategic objectives. In either case, it is possible to use the scorecard regularly thereafter (perhaps at six-monthly or annual intervals depending on the pace of change) for actual “scoring” or benchmarking of continuous performance improvements in all chosen perspectives, once the team has determined the priorities for attention and action planning.

Arguably, balanced scorecards should be used frequently for monitoring and managing operational measures, especially with the role of translating

strategy into action. This will avoid the situation where the management team look at the financial indicators on a daily, weekly, and monthly basis, but consign the nonfinancials to an occasional check.

David Norton's own consulting firm, Renaissance Solutions Inc., which he cofounded in 1992, advocated a four-phased approach in designing a bespoke balanced scorecard measurement system for clients.

Step 1: Defining Suitable Measures

Because the Balanced Scorecard should reflect the strategy, an organization must develop a *distinct* strategy. A business strategy and a Balanced Scorecard that describes it are not random. Experience has shown that the architecture of a Balanced Scorecard has several dimensions. A good design process will recognize these dimensions and provide frameworks to guide the architect and the executive team in their thinking about the strategy. There are frameworks that describe the strategy and represent the foundation on which a complex design is based. For example, in the financial perspective, one could frame discussions about the three primary components of a financial strategy:

1. Revenue growth/mix
2. Productivity or efficiency
3. Asset utilization and expenditure management

Whether operating a private sector business firm or a public service organization, executive teams use this framework to anchor their financial objectives. Similar frameworks for the Customer, Internal Processes, and Learning and Growth perspectives will give the management team a basis from which to consider the setting of strategic objectives.

The Canadian Treasury Board has also produced a "Management Accountability Framework" with realistic expectations, indicators, and suggested measures that comprehensively integrate the many management-focused initiatives currently underway and the vision and ethics they share.^[10]

Step 2: Build Consensus Around Strategic Objectives

It is common and normal for a management team not to share full consensus on the relative importance of its strategic objectives. Even harmonious management groups in well-managed organizations will often lack a shared understanding about the overall strategy and the relative roles of different groups within the organization and this can keep the team from agreeing on priorities. The second step of the development process is designed to build consensus among the members

of the executive team around the long-term strategic priorities of the organization. To achieve this goal, each executive team member should attempt to capture his or her implicit and explicit strategies for the business. These personal visions are then synthesized into feedback that is reviewed at an executive workshop. During this session, the executive team learns about where there is and is not consensus about their strategy and discusses unresolved issues. Ultimately, a coherent group vision for the organization emerges in the form of 10 top priority objectives.

Step 3: Select and Design Measures

With the prioritized strategic improvement objectives agreed upon by the executive team, the next step is the selection of measures to track the achievement of these objectives. Subteam working sessions focus on the development of measures for a subset of the objectives, finalizing the wording of objectives and searching for measures appropriate for tracking each objective. At the end of this step, the subteams synthesize their recommendations into a united "strategic story." With agreement on the strategic objectives and measures, the Balanced Scorecard measurement system design is complete.

Step 4: Develop the Implementation Plan

For a Balanced Scorecard measurement system to create value, it must be integrated into the management system of the organization. The final step of the process entails three primary tasks:

1. Identifying the current practices in various management processes
2. Evaluating opportunities for integrating the BSC into the management process
3. Developing an implementation plan

This step typically reviews the organization's approach to data reporting and review, management meetings and decision-making, strategic learning, strategic communication, personal objective setting, and planning and budgeting.

UNLOCKING THE STRATEGIC BENEFITS

Companies and government organizations that have been using the Balanced Scorecard for some time and have successfully integrated the tool into their management processes are beginning to see significant financial and operational results. Managers describe the intangible benefits gained from being able to "understand

the drivers of business success” and “educate the organisation, deciding how, when and where to raise the bar.”

By developing a Balanced Scorecard that truly tells the story of the desired strategy, it is possible to set the foundation for a management system that is capable of driving dramatic improvements in performance.^[7,8]

REFERENCES

1. Kaplan, R.S.; Norton, D.P. The balanced scorecard: measures that drive performance. *Harvard Bus. Rev.* **1992**, January–February.
2. Kaplan, R.S.; Norton, D.P. Putting the balanced scorecard to work. *Harvard Bus. Rev.* **1993**, September–October.
3. Kaplan, R.S.; Norton, D.P. Using the balanced scorecard as a strategic management system. *Harvard Bus. Rev.* **1996**, January–February.
4. Kaplan, R.S.; Norton, D.P. *The Balanced Scorecard*; Harvard Business School Press: Boston, MA, 1996.
5. Kaplan, R.S.; Norton, D.P. *The Balanced Scorecard: Translating Strategy into Action*; Harvard Business School Press: Boston, MA, 1996.
6. European Foundation for Quality Management. <http://www.efqm.org/>.
7. Kaplan, R.S.; Norton, D.P. Having trouble with your strategy? Then map it. *Harvard Bus. Rev.* **2001**, January–February.
8. Kaplan, R.S.; Norton, D.P. *Strategy Map—Converting Intangible Assets into Tangible Outcomes*; Harvard Business School Press: Boston, MA, 2004.
9. http://www.tbs-sct.gc.ca/cmo_mfc/Toolkit2/cap_eval_brochure/cab_e.asp (accessed May 2005).
10. http://www.tbs-sct.gc.ca/cmo_mfc/MAF/IC/MAF-e-notes.asp (accessed May 2005).

BIBLIOGRAPHY

1. A comprehensive bibliography of papers on the use of balanced scorecards in public administration, web-links, and sources of information appears in a National Partnership for Reinventing Government paper: Balancing Measures—Best Practices in Performance Management. August 1999. Available at www.orau.gov/pbm/links/npr1.h. This paper also has useful links and references for the Malcolm Baldrige Model and Award Scheme.
2. David Norton is a founder of the Balanced Scorecard Collaborative—a Community of Practice for balanced scorecard users: <http://www.bscoll.com>.
3. The author of this article has explored the importance of leadership of the process of organizational change using balanced scorecard tools in the following book: Reed, P.J. *Extraordinary Leadership—Creating Strategies for Change*; Kogan Page: London, 2001 (reprinted 2002, 2003, 2004).
4. The website of the Baldrige National Quality Program (BNQP) can be found at <http://www.quality.nist.gov/>.

Balanced Scorecard Tools for Organizational Assessment in Public Administration and Policy

Peter J. Reed

Bannock Consulting, London, U.K.

INTRODUCTION

Traditional financial metrics (as summarized in a balance sheet) do not capture the full or the true value or worth of any organization, but they most obviously fail to do so in the field of public services. It is now widely recognized that they are also inadequate as sole arbiters of success in the private sector. This is most evident in “knowledge-based” organizations for which people (whether we see them as “intellectual capital” or “human resources”), technology, and an organizational climate of innovation and learning are essential to support any strategy.

The “Balanced Scorecard” is a proven approach to strategic management that was originally conceived to embed long-term strategy into the management system through the mechanism of measurement. The concept was originally developed in 1992 at Harvard by Robert S. Kaplan and David P. Norton and became immediately influential with managers of private sector companies and corporations.^[1-7] In fact, the public sector, worldwide, has been just as receptive in recognizing that effectiveness and efficiency were the keys to their particular stewardship of financial resources. The Balanced Scorecard translates vision and strategy into a set of prioritized objectives that effectively communicates strategic intent, and it motivates employees and tracks performance against the established goals. It has consequently become the foundation of a number of benchmarking systems to track both the degree and the rate of continuous improvement in standards of service delivery. Because of its holistic design concept, it can, however, prove to be an excellent framework model for all improvement initiatives. The process of using it to “score” performance and the score itself are actually secondary in importance to the understanding and insights that the model can generate for strategy formulation and strategy mapping through the encouragement of systems thinking.

This paper introduces the concept of the Balanced Scorecard as a self-assessment tool for organizational development in the public services. It describes how the use of one version of this, in its manifestation as the Common Assessment Framework (CAF), has helped governments in many countries to monitor

and evaluate the progress of their performance improvement initiatives.^[8] In the case of public services that are subject to modernization or reform programs, such self-assessment could follow on from structural and functional reviews. Alternatively, it could precede such a review and could inform restructuring where this is deemed desirable for continuous improvement in service delivery. The paper also touches briefly on key requirements for the success of such an initiative.

THE NEED FOR A HEALTHY BALANCE

Balance suggests a steadiness that results when all parts are properly adjusted to each other, when no one part or constituting force outweighs or is out of proportion to another.

—*Webster's Third New International Dictionary*

This definition of balance provides the essence of an overview of what constitutes a healthy organization. In the same way, performance management and measurement systems must achieve a balance that recognizes that there is more to life (and management) than financial results alone: a view that supports progress against predetermined objectives, without suboptimization. This view has encouraged an understanding that organizations are more organic than mechanical. They cannot be simply re-engineered. Rather, they benefit from an approach that is more akin to a holistic health and fitness regime. Continuous improvement toward excellence can be planned. It can be project managed and integrated into daily operational activities; its effectiveness can, and must, be measured if learning from it is to inform future strategy.

SUCCESSFUL PERFORMANCE MEASUREMENT AND MANAGEMENT

Various groups, including the National Partnership for Reinventing Government in the United States (See “[Bibliography](#)” at the end of this article) the Cabinet Office in the United Kingdom, and the

Innovative Public Service Group in Europe, have found that there are certain attributes that set apart successful performance measurement and management systems. The following is a summary of their common conclusions:

- A conceptual framework is needed for the performance measurement and management system.
- Effective internal and external communications are the key to successful performance measurement.
- Accountability for results must be clearly assigned and well understood.
- Performance measurement systems must provide intelligence for decision makers, not just compile data.
- Remuneration rewards and recognition should be linked to performance measurements.
- Performance measurement systems should be positive, not punitive.
- Results and progress toward program commitments should be openly shared with employees, customers, and stakeholders.

PERFORMANCE MANAGEMENT SYSTEM GOALS

The following are the reasons a leading-edge organization seeks to create an efficient and effective performance management system:

- To translate vision into clear measurable outcomes that define success and that are shared throughout the organization and with customers and stakeholders.
- To provide a tool for assessing, managing, and improving the overall health and success of business systems.
- To continue to shift from prescriptive, audit- and compliance-based oversight to an ongoing, forward-looking strategic partnership involving the organization and any partners with whom it transacts its business operations.
- To include measures of quality, cost, speed, customer service, and employee alignment, motivation, and skills to provide an in-depth yet holistic and predictive performance management system.
- To replace existing assessment models with a consistent approach to performance management.

THE BALANCED SCORECARD APPROACH

Leading organizations agree on the need for a structured methodology for using performance measurement information for the following reasons:

- To help to set agreed performance goals.
- To allocate and prioritize resources.

- To confirm or change current policy or program directions to meet those goals.
- To report on success in case of meeting those goals.

Governments and agencies throughout the developed world have adopted variations on the Balanced Scorecard methodology as their chosen methodology for informing and deploying strategic direction, communicating expectations, and measuring progress toward agreed objectives. The Balanced Scorecard is a conceptual framework for translating an organization's strategic objectives into a set of performance indicators distributed among the following four perspectives:

- Financial
- Customer
- Internal business processes
- Learning and growth

Some indicators are maintained to measure an organization's progress toward achieving its vision; others are maintained to determine the long-term drivers of success. Through the Balanced Scorecard, an organization monitors both its current performance (finance, customer satisfaction, and business process results) and its efforts to improve processes, motivate and educate employees, and enhance information systems—in fact, its ability to learn and improve.

THE COMMON ASSESSMENT FRAMEWORK

The CAF is a result of cooperation among the EU ministers responsible for public administration. On request from the Directors General of this field, the new version of the CAF has been developed by the Innovative Public Service Group. It is available in the public domain, and the European Institute of Public Administration has a section of its website devoted to providing support to users who can also post their results on a database administrated by The European Institute of Public Administration (EIPA), should they wish to do so for the purpose of international benchmarking. (See www.eipa.nl and go to Projects/CAF in the menu.) Acknowledgment is due to EIPA for the following description.

A pilot version of the CAF was presented in May 2000 during the First European Quality Conference for Public Administrations held in Lisbon. The present version is based on experience gained in implementing and using the first version of the CAF.

The CAF is offered as a tool to assist public sector organizations across Europe to use quality management techniques to improve performance. It provides a simple, easy-to-use framework, which is suitable for a self-assessment of public sector organizations.

The CAF has four main purposes:

1. To capture the unique features of public sector organizations.
2. To serve as a tool for public administrators who wish to improve the performance of their organization.
3. To act as a “bridge” across the various models in use in quality management.
4. To facilitate benchmarking between public sector organizations.

The CAF has been designed for use in all parts of the public sector, applicable to public organizations at the national/federal, regional, and local level. It may also be used under a wide variety of circumstances, e.g., as a part of a systematic program of reform or as a basis for targeting improvement efforts in public service organizations. In some cases, and especially in very large organizations, self-assessment may also be undertaken in a part of an organization, e.g., a selected section or department.

The structure of the CAF is based on the Excellence Model developed by the European Foundation for Quality Management (EFQM) and is illustrated below (Fig. 1).

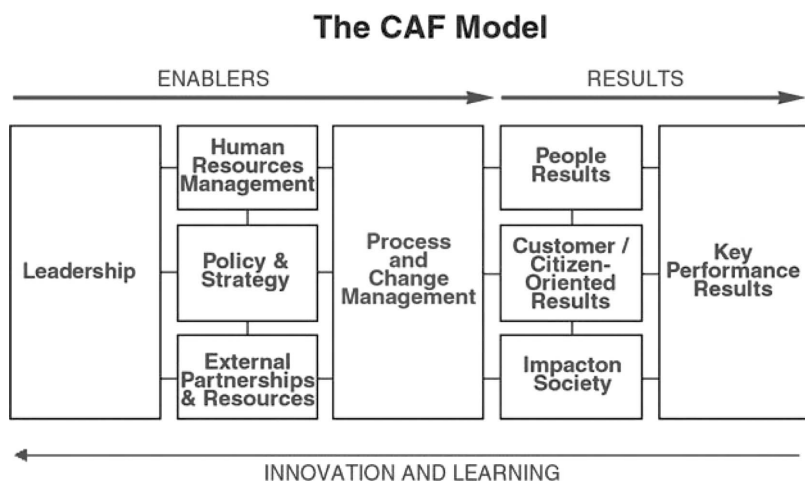
The nine-box structure identifies the main aspects requiring consideration in any organizational analysis. Within each of these boxes, a list of criteria is provided. The criteria identify the main issues that need to be considered when assessing an organization. The freely available CAF materials provide comprehensive lists of questions to prompt management teams to ask rigorous questions for each of the criteria. Using the CAF thus provides an organization with a powerful and well-proven framework to initiate a process of continuous improvement.

The CAF provides:

- An assessment based on evidence.
- A means to achieve consistency of direction and consensus on what needs to be done to improve an organization.
- An assessment against a set of criteria that have become widely accepted across Europe.
- A means to measure progress over time through periodic self-assessment.
- A link between goals and supportive strategies and processes.
- A means to focus improvement activity where it is most needed.
- Opportunities to promote and share good practice within different areas of an organization and with other organizations.
- A means to create enthusiasm among employees by involving them in the improvement process.
- Opportunities to identify progress and outstanding levels of achievement.
- A means to integrate various quality initiatives into normal business operations.

To summarize, self-assessment against the CAF model offers the organization an opportunity to learn more about itself: where it is now in terms of performance and standards of service delivery and how it believes it can improve in measurable ways.

The CAF has the advantage of being compatible with the EFQM model and may, therefore, be a first step for a public service organization wishing to go further with quality management. The materials available are essentially self-explanatory, but many organizations find it helpful and beneficial to have the first baseline assessment facilitated by an experienced consultant or practitioner who is thoroughly familiar with



(Based on the EFQM Excellence Model diagram)

Fig. 1 The CAF model.

the methodology, its uses, and the best way to present the team's findings to politicians or to the government.

Key Outputs of CAF Self-Assessment

A self-assessment exercise will normally be conducted by the management team and can be done comfortably in two working days. There will be two main outputs.

A Score. The first workshop will produce a baseline score against which future assessments can be benchmarked. The scoring system is internationally recognized and accredited. Indeed, there is an international database (held by the EIPA in the Netherlands), where it is possible to register results, and if desired compare them with others at a similar stage of development, or indeed with best practice organizations.

A List of Areas for Improvement. This is potentially more important than the score because it should inform strategy. Typically, around 100 areas for improvement (AFIs) will be identified across the board of the model (an average of 10–12 per criterion). When consensus is reached concerning which of these should be developed into priority improvement projects (maintaining the spread and systemic balance in the final selection if possible), a “vital few” should be incorporated into annual operational plans, and their process and project management monitored and used to improve organizational learning about what works best to improve service delivery standards.

Practical Benefits of the Methodology

Self-assessment using CAF can be conducted as a follow-up to a functional review or it can precede and inform such an analysis. It can form a very useful step in the process of conducting a situation analysis as a part of the strategic planning process for a department (or indeed any organization). For example, managers often find it difficult to frame a “picture” or comprehensive description of the desired medium-term future state of affairs. Often, the complexity of the current reality seems depressing and the problems all-pervading. In such cases, the model provides a useful framework and a well-tested set of prompts in the form of questions for the management team to help them formulate such a description.

1. Use in such a situation analysis or functional review (and gap analysis between a “picture of the future” and “current reality”) will typically identify 100+ areas for improvement—at least 10 per criterion that will help to:
 - a. Link outputs, processes, and inputs.

- b. Prioritize actions—within and across criteria.
2. It shows what good practice looks like and provides:
 - a. A broader systemic view of performance—highlighting interdependencies and the need to measure more than just budget compliance.
3. It involves staff in continuous improvement:
 - a. It improves internal communications—highlighting knowledge sharing.
4. It provides a management framework to:
 - a. Link political intent to local operations.
 - b. Add structure to other initiatives (especially any improvement or change initiative).

Key Strengths of the Methodology

The model's enablers inform:

- Sound broad-based strategy development and deployment.
- Good employment practice.
- Sound financial management.

The results criteria provide:

- Focus on client/customer perceptions and impact.
- Comprehensive financial measures.

OWNERSHIP AND RESPONSIBILITY

Self-assessment using CAF, or any other Balanced Scorecard model, is essentially a strategic exercise. It is designed to inform strategy and action planning. Therefore, just like other major change initiatives, it is important that the process is seen to be “owned,” sponsored, and led by senior line managers. In fact, so far as there is an important policy dimension—which is after all the foundation of any strategies that are to be developed—there is no reason why politicians (secretaries of state, ministers, or deputy ministers) should not take part, or at least take a serious interest. This sends out a message, if one is needed, that there is a sense of urgency, and that staff and stakeholders should note that this is an important and serious initiative designed to catalyze the changes necessary to improve performance and value for money.

There is certainly a vital component of the exercise that relates to staff development. Results, after all, can only be achieved through people. In this context, it is most welcome that many government departments are now upgrading the status and responsibilities of HR departments or units. Managers need to show, however, that they recognize and understand that HR units are not responsible for staff development—only for coordinating this. The responsibility for developing

people should rest squarely with line (senior) managers, and should be based on the emerging organizational priorities from strategic and operational plans, and from the CAF assessment.

Unfortunately, the concept of such line “managers” is new for the governments of many developing countries, and so HRM units may have to take the lead in the establishment of HRD strategies, and in identifying skills gaps that emerge from initiatives such as the self-assessment benchmarking process. The recommendations of any diagnostic studies—for example, training needs’ analyses—should be taken in the context of all the other conclusions, for instance, from functional review, and aligned with those relating to HRM that will emerge from the identification of AFIs when self-assessment benchmarking using the CAF commences.

SUMMARY—PERFORMANCE AND QUALITY MANAGEMENT USING BALANCED SCORECARD SELF-ASSESSMENT AS AN ORGANIZATIONAL DEVELOPMENT SYSTEM

- Performance management is a system within a larger system. To get maximum benefit, managers must pay attention to the whole process, not just a part of it. Organizational self-assessment is as important as individual appraisal.
- Performance management links up with strategic planning, budgeting, employee development, employee compensation systems, and quality improvement programs. The more links that are established between performance management and other processes in the organization, the better the return on investment. The linkages and interdependencies between the criteria “boxes” in the model and the information flows are, in fact, as important as the criteria themselves.

TEN KEYS TO SUCCESS

- Political or senior management “ownership”
- Honesty (about current reality)
- Openness (to achieve consensus)
- Making the exercise “blame free” (analysis and monitoring should simply track performance and improvement)
- Avoiding excuses/justification

- Showing cause and effect linkages where possible (or acknowledge that these are complex and nonlinear)
- Recognition of commitment
- Alignment with strategic and operating plans
- Selection of a vital few action priorities from AFI
- “Quick Win” projects to turn this analysis into action and results.

REFERENCES

1. Kaplan, R.S.; Norton, D.P. *The Balanced Scorecard: Measures that Drive Performance*; Harvard Business Review: Boston, MA, 1992.
2. Kaplan, R.S.; Norton, D.P. *Putting the Balanced Scorecard to Work*; Harvard Business Review: Boston, MA, 1993.
3. Kaplan, R.S.; Norton, D.P. *Using the Balanced Scorecard as a Strategic Management System*; Harvard Business Review: Boston, MA, 1996.
4. Kaplan, R.S.; Norton, D.P. *The Balanced Scorecard*; Harvard Business School Press: Boston, MA, 1996.
5. Kaplan, R.S.; Norton, D.P. *The Balanced Scorecard: Translating Strategy into Action*; Harvard Business School Press: Boston, MA, 1996.
6. Kaplan, R.S.; Norton, D.P. Having trouble with your strategy? Then map it. *Harvard Bus. Rev.* **2001**, Jan–Feb.
7. Kaplan, R.S.; Norton, D.P. *Strategy Maps—Converting Intangible Assets into Tangible Outcomes*; Harvard Business School Press: Boston, MA, 2004.
8. EFQM. <http://www.efqm.org/> for information on the original Excellence Model, and that of the EIPA at: <http://www.eipa.nl> for information on the CAF.

BIBLIOGRAPHY

1. Baldrige National Quality Program (BNQP). <http://www.quality.nist.gov/>.
2. National Partnership for Reinventing Government. Balancing measures—best practices in performance management. Aug 1999. www.ora.gov/pbm/links/npr1.h. (A comprehensive bibliography of papers on the use of balanced scorecards in public administration, web-links, and sources of information. This paper also has useful links and references for the Malcolm Baldrige Model and Award Scheme.)
3. Reed, P.J. *Extraordinary Leadership—Creating Strategies for Change*; Kogan Page: London, 2001(reprinted 2002, 2003, 2004). (The author has explored the importance of leadership of the process of organizational change using balanced scorecard tools).

Benchmarking in Public Administration and Public Policy

A. Gunasekaran

*Department of Management, University of Massachusetts Dartmouth,
North Dartmouth, Massachusetts, U.S.A.*

Krishan M. Gupta

*Department of Accounting and Finance, University of Massachusetts Dartmouth,
North Dartmouth, Massachusetts, U.S.A.*

INTRODUCTION

Changing markets, globalization of economies, technological pressures, and competitive positions are necessitating the need for innovative management practices not only in the private sector, but also in the public sector and its administration. Due to heightened public awareness and also as a matter of good public policy, public sector organizations are being increasingly challenged to achieve the highest levels of performance. Although the definition of performance may be difficult to establish for the public enterprises, they are increasingly expected to perform better, and to deliver more to their various stakeholders and constituents with lesser resources.

Since the 1980s, the private sector for-profit enterprise management has changed significantly, thanks in part to the various modern management practices, such as strategic management, Just-In-Time (JIT), business process reengineering, lean and agile manufacturing, total quality management (TQM), and benchmarking (BM). With a time lag, public enterprises have also implemented these strategies and techniques for improving their performance and efficiency.

Benchmarking is becoming an increasingly popular tool for transferring best practices into the public sector. It has also been used effectively as a tool to control and scrutinize the mission effectiveness and the underlying generic service processes of non-profit public enterprises (PEs). It facilitates a radical change in the organizational culture, and supports continuous and sustained efforts toward improvement. It is a positive and proactive process; however, it can easily turn into a negative and reactive process, especially for the PEs, if the controlling authorities use BM for measuring performance alone.

The main significance of BM lies in its ability to identify and unearth the best practices, from anywhere and anyone. The key emphasis is on learning from others and not trying to reinvent the wheel. Openness and ability to adapt and change are the major ingredients of the BM process.

Market-based economy is generally accepted as the appropriate model for global and national economic development. However, the public sector organizations will continue to deliver critical services and play a significant role in any economy. Just like the private sector, the PEs are equally driven by the concerns of cost, efficiency, and responsiveness toward their stakeholders. However, unlike the private sector, it is difficult to identify suitable measures of performance in PEs. Further, dependence on public-funded, fixed, short-term budgets creates its own peculiarities and political imperatives. Nonetheless, BM in the public sector has proved to be equally, if not more, successful. Considering the importance of BM in PEs, this article attempts to review some of the literature with the objective of understanding the major issues in performance improvement and discussing the application of BM in PEs. A framework for BM in PEs has also been outlined. It is hoped that this article will serve as background work for further research in this area.

BM AND PEs

American Productivity and Quality Center (APQC) defines BM as “the process of identifying, understanding, and adapting outstanding practices from organizations anywhere in the world to help your organization improve its performance.”^[1] Camp^[2] defined BM as “the continuous process of measuring products, services, and practices against the toughest competitors or those companies recognized as industry leaders. (It is a) search for the industry best practices that will lead to superior performance.” Thus, the BM process is focused on continuous performance improvement by identifying, understanding, and adapting the best practices both inside and outside the organization.

Benchmarking improves overall efficiency of the organization. A typical large business enterprise can have significant payback within the first year of BM.^[3] It integrates easily into the strategic management of the organization through its emphasis on

continuous improvement of quality and learning culture. BM facilitates the organization in achieving its strategic objectives and makes them more knowledge based. It forces organizations to move away from internal performance measurement and toward adaptation and learning. Although BM is well integrated with TQM and continuous improvement, it does not conflict with the other ongoing efforts of performance improvement.

The continuous process of documenting and sharing BM efforts promotes the core value of learning and openness necessary for continuous improvement and growth. Further, it forces organizations to become more outward looking and to seek new challenges with the spirit of overcoming new challenges. It improves the overall competitiveness of the organization not only by making its business processes more efficient, but also by actively promoting the learning and continuous improvement culture, which is much needed in the rapidly changing environments. Exposure to the outside world and interaction with BM partners also helps PEs to identify new areas of opportunities, improve their overall mission effectiveness, and achieve cost reduction and efficiency.

Benchmarking is equally applicable to different service processes within any organization—profit or non-profit, private or public. The approach has been successfully applied in manufacturing, sales, service, training, and educational organizations. In the private sector, top managers are always aware of the environment and are constantly comparing (consciously or unconsciously) their organizational performance with others. Benchmarking formalizes this process for PEs and changes the primary management focus from “outcomes” to the enablers of the outcomes. In the public sector, BM has been used successfully as a tool for measuring organizational performance, bringing in the best managerial practices, and learning from other organizations. However, there remain some skepticism and resistance to organizational change, along with a certain level of mistrust in the public sector about the ready adaptability of private sector management practices to the PEs.^[4]

It is quite likely that the industry best practice may not be easily revealed due to transaction and other search costs. Only relative or local optimum benchmarks may be established to bridge the perceived gaps in performance.^[4] Benchmarking can also be achieved against well-established industry practices, instead of searching for the external best practices. For example, DIN ISO 9000-9004, quality awards, and Citizen Charters can be equally effective in BM in PEs. Quality award competitions promote superior performance by fostering innovation and quality. The generic administrative practices of participating agencies become more transparent and open to public scrutiny,

thereby forcing them to strive for superior performance. Citizen Charters, quite popular in Europe, focus on the idea of accountability in PEs. They are used for assessing and rewarding quality in PEs. It is a quality checklist having the sanction and authority of the state with the stated objective of improving the delivery of public service to the citizens. However, BM is considered a superior approach as compared with the traditional target-setting approach for managing performance.^[2]

The BM process is a paradigm shift requiring a sea-change in the organization's practices, culture, and structure. The focus is on adaptation, not on adoption or mere imitation. The classic example in BM is that of Henry Ford, who adapted the Chicago slaughterhouse practices to develop the assembly line approach for automobile manufacturing. It is most effective when applied as a tool of continuous improvement within the context of a larger change initiative.

Although an easy case can be made for growing “managerialism” in the public sector, with the stated goal of transferring the best managerial approaches/practices of the private sector to the public sector, it may not be all that easy. It is easy to apply bottom-line and customer-driven approaches in the management of private business enterprises. However, similar measures for PEs are difficult to identify. Even the customer satisfaction approach may be questionable in some cases. Here, the uniqueness of BM stands out. The private sector is driven by competition, and BM introduces an element of cooperation with the BM partners. Yet, in the context of PEs, BM introduces an element of competition—they are forced to compare themselves with others and to improve.

IMPLEMENTATION OF BM PROJECTS

Total commitment, support, and understanding of the top management for BM is the most critical ingredient for successful implementation of the BM process. Besides providing support to the BM team and encouraging involvement at all levels, top management also needs a good understanding of internal business processes and the process of BM itself.

Benchmarking becomes more powerful when combined with change initiatives and used as a tool for continuous improvement. It has yet to become a core value in itself for the organization. However, there is a clear understanding that the best practice is a moving and ever-evolving target. There is no single best—it is relative and contextual.^[5]

Benchmarking requires an organizational culture that encourages openness to new ideas and eagerness to seek best ideas from any source. It requires active willingness to seek BM partners within and outside

your own industry. Obviously, it is a two-way street and requires openness to share your own practices with your partners. In terms of attitude, the BM approach can be summed up as, “We want to meet and beat the best known performance in any process.”

Adequate time and resources need to be allocated for BM projects. However, there has to be a strict cost–benefit analysis at every stage to ensure significant positive contributions and to not waste resources on marginal issues. Not everything can be benchmarked, and the PEs have to choose carefully the areas for BM process.

Like any implementation barriers, BM has its own impediments during the course of performance measurements and identifying best practices. Some of them include inertia and resistance to change, cost, response time, temptation to copy, focusing on performance measurement rather than the processes and practices that result in superior performance, lack of involvement by top management, and organizational structure and culture that punishes mistakes made while innovating. Lack of adequate support, time, and resources are identified as primary obstacles in successful implementation of BM.^[1]

A FRAMEWORK FOR BM IN PEs

Performance measurement in the public sector is more difficult than in the private sector because simple concepts of financial measurements cannot be applied. It is not feasible to apply the usual measures of profitability to the public sector. At the same time, it is equally difficult to operationalize the concept of “consumer satisfaction.” The problem of scarce resources in public sector administration necessitates a certain degree of insensitivity to consumer demands. In other words, consumer satisfaction cannot be the only, or dominating, dimension in performance measurement in the public sector.^[4]

Technology, globalization, competition, and collaboration have completely transformed business processes in the public sector. This calls for a different approach to BM for PEs in terms of its responsiveness to mission objectives. The BM champions also have to take advantage of the possibilities offered by technology in terms of communication, conferencing, and search. However, at the same time, one has to be careful about the reliability of the search results because the barriers to entry in the information dissemination are much lower. Further, by carefully defining and limiting the scope of the project, BM can deliver quicker results.

An effective way to control cost is to select only a few processes at a time for BM. One need not benchmark all the areas at once. The step-by-step approach will also help to leverage the experience and results gained from the initial phases.

Under the present environment, there is a greater need for analyzing the service processes before launching a BM project. First, the process of internal service process mapping itself will reveal several ideas that can make the process more efficient. Second, a thorough understanding of the internal service processes will reduce the total project time by focusing on core service processes.

Mission Effectiveness

The goal of any PE is to fulfill its mission or charter, whereas the business enterprises seek profitability. Therefore, the key performance metric for a PE at the strategic level is *mission effectiveness*.^[6] This concept can be compared with the strategic competitiveness in the corporate sector. The mission effectiveness for the PEs is not fixed, it is relative and contextual. The mission and underlying sub-missions may change their tone and dominance, depending on the political and social pressures on the PEs.

Political Issues

The existence of a business enterprise is solely determined by its economic performance. However, for a PE, its existence is often determined by a political process, which in turn depends on various sociopolitical factors at a given time. Further, some public agencies in the area of public safety or law and order might even receive a higher dose of resources, even when not able to provide the expected deliverables. Thus, the life of a PE may be much more dependent on its sociopolitical relations with its environment than on improved or worsened performance. At the same time, public pressures and fiscal constraints may impose certain performance targets (Citizen Charters in Europe) on PEs. However, BM is considered to be a better approach to manage performance than the traditional target-setting approach.

Cost Reduction and Efficiency

Profitability measures, so common in business enterprises, cannot be readily applied to the PEs. However, some financial measures such as cost reduction and efficiency can be used to establish best practices. Efficiency measures can also be financial or non-financial. For example, capacity utilization ratios, total passenger journeys per year per staff hours, and so on, can be used to measure efficiency of a metropolitan mass transit system.

Thus, if performance is the factor for the viability of a PE, effectiveness and efficiency of its operations

are the key indicators to be examined and improved. While it may be relatively easier to set up metrics and subject them to BM for generic support functions, it may not be so easy to set up metrics for mission-specific effectiveness and to benchmark them. When faced with internal or external targets and imprecise measures of effectiveness, measures of economic efficiency and cost reductions will dominate the BM performance metric.

Customer Satisfaction

It seems easy to select customer service as one area that may be common to PEs enterprises. Both should strive for better customer service. However, “even here there is a difference, because the definition of customer is different in the two cases.”^[6] Besides, the concept begs the question—who is the customer? The user of the service or the ones funding the PE (the taxpayer)? Further, even if the consumer of the service is accepted as the customer, it may be optimal to allow a certain amount of dissatisfaction as part of the mission.^[4] For example, it may be desirable to leave the recipients of public dole a little bit dissatisfied with the service to coax them away from systemic dependence on public welfare.

Stakeholders

In the case of business enterprises, business owners can easily be identified as the primary stakeholders. However, for PEs, it is difficult to identify one major group of stakeholders. By definition, PEs exist for the public good. Taxpayers and founders of the PEs, regulators, legislators, and the general public have a stake in successful mission accomplishment. The various stakeholders may not have a unified vision for the PE and its mission, thereby making the exercise of performance measurement and BM extremely fractious.

Open Information

In business enterprises, information is often protected due to the fear of losing the competitive advantage through the loss of intellectual property. Among PEs, generally, it is easier to secure cooperation and open communication. However, the information may still be guarded due to legal, security, and safety concerns.

Uniqueness

For a business enterprise, being unique can be a success factor. However, because competition is generally

not an issue for PEs, sameness and economies of scale are more important than being different from others.

Club v. Out-of-the-Box Approach

Because competition is usually not an issue for PEs, it is easier to form “innovation rings” or consortia of similar PEs to foster innovation and to learn from each other. The managers are more comfortable sharing the information with each other in such clubs. The knowledge transfer is easier due to the shared language and technologies.

However, such clubs can easily degenerate into mutual back-scratching clubs trying to establish excuses for their poor performance. A certain amount of “soldiering” can also take place. Nonetheless, such clubs can help to quickly establish in-class best practices.

Out-of-the box or “world-class” out-of-the-class BM processes can be helpful in establishing the best practices at a global level. This approach is especially useful when establishing best practices for generic business processes.

Information Technology

Information Technology is both the enabler and enforcer for BM. Advances in information technology (IT) and communication technology have lowered the costs for identifying and transferring best practices and monitor their implementation. The IT advances can assist in readily evaluating the impact on all aspects of an organization, both in detailed and aggregated levels. Knowledge management tools, such as Enterprise Resource Planning (ERP) systems, will continue to evolve and provide management even better controls over the service processes.

However, in the case of PEs, IT can become a stumbling block due to the lack of adequate investment in IT, inability to attract and retain the right kind of personnel, and the need to constantly upgrade and reconfigure. Often, IT can become a hinderance rather than a facilitator of BM. Sometimes, IT itself can become a key area for BM for the PEs.

GENERIC STEPS FOR BM IN PEs

Jarrar and Zairi list six major steps in BM for transferring the best practices—searching, evaluating, validating, implementation, review, and routinizing.^[5] Some of the major aspects that one should consider in BM practices include top management involvement, team, BM partners, networking, consultants, proactive approach, change champions, cost-benefit analysis, and continuous improvement.

The following are generic steps that can be used in the BM process in PEs:

1. *Choose the core service processes that need to be benchmarked.* The related question is “how do the selected processes fit into the overall mission of the organization from the stakeholders’ perspective?” Senior management support and understanding of the process is critical to BM. However, in the case of PEs, this may be problematic because the locus of real power may lie outside the organization.
2. *Select a team with well-defined scope and team leader.* The team should have adequate influence throughout the organization. Because the power structure in PEs is generally more centralized and most of the authority is concentrated at the top level, it is imperative that some members of the top management join this team to provide support and leadership to the project. This kind of highly visible team composition is also essential for post-BM activities. In short, the team has to become the change champion.
3. *The team scope should include a timeline for deliverables.* This issue becomes even more relevant in today’s fast-changing, quick-response environment. The team should be fully responsible for training, providing assistance and guidance, and transferring and sharing the knowledge within the organization. The team may also need adequate training and resources to carry out the process.
4. *Time and resources have been identified as primary challenges to BM; adequate budgeting is necessary.* Innovative funding techniques that involve different agencies and partnership with the private sector may be necessary. The organizational culture in PEs is different from the corporate sector—it is generally more bureaucratic and resistant to change. Therefore, the BM may not be able to deliver immediate short-term returns.
5. *Engage an outside consultant.* If there is not enough in-house expertise or understanding of the issues involved in BM then it may be necessary to engage an outside consultant who has a proven track record of BM in similar areas.
6. *Analyze and understand the core service process thoroughly before looking elsewhere.* This analysis itself will reveal significant insights into the current operations and help to improve the service process. Benchmarking essentially relates both inside and outside the organization.
7. *Adopt a strict protocol (Code of Conduct) for sharing information and proprietary knowledge with the BM partners.* This is critical not only for satisfying the concerns of your potential BM partners but also to avoid any potential antitrust, ethical, or legal issues. Well-established and tested protocols are readily available, and can be easily modified and adapted to meet the specific needs.
8. *Identify potential BM partners within or outside the industry.* It is not always necessary to work with the best-in-class performing organizations. One can also learn from less-than-the-best performing organizations.
9. *Establish an open communication link with your potential BM partners.* If the potential BM partners believe that a partner is trying to steal their ideas, then there will be no positive response. Networking through professional and trade organizations can be extremely valuable in this process.
10. *Identify core areas of service that need to be improved through BM.* Develop a few key performance measures and a suitable metric that can be used for the BM process. This metric will have to be developed in close cooperation with the BM partners.
11. *Identify the best practices in the key performance measures from BM partners.* Here, the focus should be on understanding the underlying processes that lead to the best practices. The team should distill the best practices for adaptation by the organization. Mere imitation or copying does not work.
12. *Perform a cost–benefit analysis to see if it is worthwhile to adapt such practices.*
13. *Share the results and the knowledge gained through this exercise throughout the organization on a formal basis.* Organizations involved in BM generally miss out on the opportunity to take full advantage of the potential gains, in learning and changing the culture, that can arise from sharing the knowledge gained in BM.
14. *Adapt, implement, and monitor the identified improvements in the service processes.*
15. *Perform periodic followup to evaluate the impact of improvements initiated by BM.* The evaluation should include achievement of financial and strategic objectives. Such a followup will help to further improve the process by internal BM. It will also help to leverage the BM process throughout the organization by promoting the learning culture—the fundamental focus of BM.

Benchmarking is not merely a process of measuring the best performance. The primary focus has to be on identifying business practices that lead to superior

performance, and determining how such practices can be adapted and emulated within the organization in a cost-effective manner. In other words, learn from examples set forth by others.

CONCLUSION

The research clearly indicates that BM has earned its place as a management tool and will continue to be used in the future.^[1,3,5] It is a fundamental process that promotes learning and change, thereby improving the overall efficiency and mission effectiveness of PEs.

Benchmarking is the most economical method to improve productivity and efficiency of any organization. However, it requires a sustained long-term continuous effort and cannot be a quick fix. The growth in the academic and practitioner literature in BM is indicative of its momentum and potential.

Application of BM in the PEs requires careful fine-tuning, specifically in the involvement of top management, identification of measures of mission effectiveness, and a commitment to a continuous and sustained effort toward service process improvement. The organizational culture, time, and resources can become major stumbling blocks. Certain PEs might even enjoy a limited amount of monopolistic powers due to their political right to exist and may not have any incentive to improve.

Benchmarking in PEs is a viable management practice with considerable potential for making a significant improvement in the performance in an economical manner without reinventing the wheel.

REFERENCES

1. American Productivity and Quality Center. *Benchmarking: Final Report*; APQC: Colorado Springs, CO, 1997.

2. Camp, R. *Benchmarking: The Search for Best Practices that Lead to Superior Performance*; ASQC Quality Press: Milwaukee, WI, 1989.
3. *Benchmarking: Leveraging Best-Practice Strategies*; American Productivity and Quality Center. <http://www.bettermanagement.com> (accessed April 2002).
4. Kouzmin, A.; Loffler, E.; Klages, H.; Korac-Kakabadse, N. Benchmarking and performance measurement in public sectors. *Int. J. Public Sect. Manag.* **1999**, *12* (2), 12–14.
5. Jarrar, Y.; Zairi, M. Internal transfer of best practice for performance excellence: a global survey. *Benchmarking: Int. J.* **2000**, *7* (4), 239–246.
6. Arveson, P. *Translating Performance Metrics From the Private Sector to the Public Sector*. <http://www.balancedscorecard.org/metrics/translating.html> (accessed April 2002).

BIBLIOGRAPHY

1. American Productivity and Quality Center: <http://www.apqc.org>.
2. *Benchmarking in Practice (Journal)*; American Productivity and Quality Center.
3. *Benchmarking: Int. J.*
4. Berman, E.M. *Productivity in Public and Non-profit Organizations*; Sage: London, 1998.
5. Bogan, C.E.; English, M.J. *Benchmarking for Best Practices*; McGraw Hill Inc.: New York, 1995.
6. Bruder, K.A.; Gray, E.M. Public-sector benchmarking: a practical approach. *Public Manag.* **1994**, *76*, S9–S16, September.
7. European Center for Total Quality Management. *Benchmarking: A Global Survey*; ECTQM Report Series: Bradford, 2000.
8. Loffler, E. A survey on public sector benchmarking concepts. In *Quality, Innovation and Measurement in the Public Sector*; Hill, H., Klages, H., Loffler, E., Eds.; Peter Lang: Frankfurt, 1996; 137–159.
9. Walsh, K. Quality and public services. *Public Adm.* **1991**, *69* (4), 503–514.

Biomedical Ethics and Public Policy Discourse

John Abbott Worthley

*Department of Public Administration, Seton Hall University,
South Orange, New Jersey, U.S.A.*

INTRODUCTION

Developments in medical technology during the latter part of the twentieth century revitalized the field of biomedical ethics. As the twenty-first century unfolds, broad availability and application of this powerful technology have brought the biomedical ethics dialog foursquare into the public policy-making forum. The issues being deliberated are both organizational and clinical. This article illuminates the questions being debated including political, economic and organizational issues. Specific attention is given to stem-cell research and end-of-life issues. Questions of conflict of interest and decision-making power are raised.

THE NATURE OF BIOMEDICAL ETHICS PUBLIC POLICY DISCOURSE

The clinically related issues for public policy fundamentally revolve around one of the oldest questions of bioethics: Should we do everything that we can do? Specific attention-getting issues include human cloning, stem-cell research, fertility technology, euthanasia, and abortion.^[1] Technology has largely set the agenda. As the 21st century began, a Worcester, Massachusetts, company announced that it had successfully cloned human embryos. This technological development immediately called the question: Now that we can do it, should the polity permit, encourage, proscribe, or limit the cloning of human cells? Recent scientific research has also found that human cells harvested from embryos offer significant pathways to healing spinal-cord injuries, dementia, and other medical conditions. Should we fund stem-cell research aimed at curing such disabilities and diseases? Or should issues of fetal life prevail? At the end of the twentieth century, a 63-year-old woman was successfully fertilized in vitro through frozen embryo transfer. She gave birth to a 6-lb girl. Should senior citizen women be legally denied fertility therapy? Or, should the issue not even enter the public policy forum? Largely based on the availability of sophisticated pharmaceuticals, Holland in Europe and Oregon in the United States have developed regulations for hastening

the death of terminally ill citizens through the use of such drugs. Should public policy ban “mercy killing” or continue to develop regulatory procedures for applying technology to painful end-of-life circumstances?

Advocates on all sides are lobbying to have these kinds of questions addressed in the public policy-making process. As a result, for example, in the fall of 2001, the U.S. House of Representatives passed legislation, supported by the sitting president, banning all human cloning. In January 2002, a Presidential Council on Bioethics was appointed to advise the government on cloning and other issues. Stem-cell developments, organ selling and transplantation, as well as issues of nutrition and hydration for the terminally ill are among other subjects being addressed. The ethics of abortion, of course, has had a significant place on the public policy agenda for some time, both in the courts and in our legislatures.

Organizational issues have also entered the forum prompted by social and economic developments.^[2] Socially, the professional role of medical professionals has changed, and societal attitudes toward them have followed. Physicians have become powerful lobbyists, corporate officers of health-care organizations, and more group based than individual oriented. Nurses have unionized, entered managerial positions, and become more powerful in the health-care enterprise. Not surprisingly, consumer perspectives on these professionals have evolved. Deference to medical paternalism has diminished. Physicians are more routinely subject to legal processes for malpractice. Patients have become more alert to self-serving inclinations of physicians and less tolerant of medical hubris. As a result, discussion of truth telling, confidentiality, conflict of interest, and informed consent has proliferated, and public policies on these matters have been developed.

Economically, the escalated cost of health care has impelled debate on issues of resource allocation, fiduciary competence, insurance configurations, and compliance with standards and regulations. Is it ethical within the reality of limited resources to duplicate health-care services in defined geographic areas? Is it ethical to close an emergency room or a maternity ward to realize economies of scale within local communities? How should health-care resources—needed by

all but sufficient for only some—be distributed? Should limited resources be expended on expensive technologies that benefit relatively few or on more generic treatments that benefit many? How should such decisions be made? Should insurers be required to offer certain coverages and prohibited from some denials of coverage? Should health-care providers be monitored for strict compliance with standards and penalized criminally, as well as financially, for non-compliance? Should these questions be addressed privately or within the political process? Should answers become a matter of public policy? These are some of the concerns that have come to the fore in shaping modern bioethics discourse within the public policy context.

FUNDAMENTAL QUESTIONS

The public policy debate over these and other issues revolves around a few fundamental, timeless, and profound questions. The clinically focused debate inevitably centers on questions about human life. When does human life begin? What, in fact, constitutes human life? When does human life end? Should society directly intervene in the beginning and ending of life, or should society leave it to more natural processes? The abortion public policy-making experience, for example, is grounded in the question of when human life begins. The conclusion that life begins at conception leads to a public policy position to ban all abortion. The conclusion that viable life begins at the third trimester of pregnancy leads to a public policy position to fund abortion within limits and regulations. The conclusion that human life begins at birth leads, of course, to less restrictive abortion policy proposals. Perspective on the nature of human life enters the debate when prenatal screening reveals anomalies such as genetic defects. When deformities and abnormalities are viewed as a diminution of human life, policy proposals permitting late-term abortion tend to emerge. When a child is born in great distress, perspective on the end of life spawns views on policy for intervention or nonintervention.^[3]

Euthanasia policy dialog is largely rooted in conceptions of the nature of human life. Those who see a certain level of brain function as essential tend to seek policies permissive of euthanasia actions in situations such as a persistent vegetative state. Those who view the “human soul” as the essential factor seek proscriptive policies. Those who view physical well-being, such as freedom from continuous and severe pain, as essential to human life tend to propose less restrictive public policies. The emergence of technology for human cloning, of course, is now bringing the

question of the nature of human life to another level of discernment.

In recent years, the case of Terri Schiavo in Florida highlighted both the debate on the nature of human life and the role of the public policy process in that discourse. In addition to the patient’s family, courts, the state legislature, and even the U.S. Congress entered the discourse of whether or not a feeding tube could be removed ethically. The discourse was further complicated and enriched by the bifurcated role of the church in the matter: Some officials and theologians argued that a feeding tube is extraordinary and artificial (and, therefore, can be removed); others maintained that a feeding tube is now ordinary and natural and should not be removed.^[4]

State governments have entered the discourse arena with an array of policies that range from permitting health-care providers to remove life support even against family wishes to forbidding such action. Among the more permissive is the state of Texas, whose “futile care law” permits hospitals to withdraw life support from patients whose conditions have been deemed hopeless by the hospital. If the family objects, the law provides for a 10-day period in which to find another willing facility. Failing that the hospital is legally able to end artificial support.^[5]

The organizationally focused debate centers on questions of conflicting interests. Should the health-care interest of each individual person be the chief concern of bioethics or should the interest of society as a whole be primary? If both interests are viewed as essential, what should the policy be when the interests conflict? How do we honor the dignity of every person and still maintain the viability of the community? Because the health-care resources of the polity are less than the full health-care needs of all the individuals who form the community, how should allocation of the limited resources be directed? Should the finite health-care resources be allocated on cost-benefit principles of economics, on power and influence principles of politics, or on philosophical principles of the greatest good with the least harm? Who should make the decisions—ethics committees organized within health-care enterprises, medical experts organized through their professional societies, politicians elected by the masses, jurists organized in legal systems, insurers structured capitalistically and regulated governmentally, or philosophers organized academically? These are basic questions today, and, increasingly, the public policy-making process is being used as an arbiter for framing answers. Implicit in the discussion is a disconcerting issue—in policy decisions involving bioethics, whose judgment should ultimately prevail—the patient’s, the health-care provider’s, the jurist’s, the ethicists, or the judgment of the polity through the public policy-making process?

AN EXAMPLE OF BIOETHICS AND PUBLIC POLICY

Current developments in human cloning and stem-cell research, because they are markedly bringing the fields of bioethics and policy making into the public arena, provide rich fodder for better understanding the expanding intersection of biomedical ethics with the public policy process. The technology now enables society to produce human life outside normal reproductive processes and to improve human health through stem cells. The fundamental question is whether society should now use this new, profound capability or proscribe, limit, and/or regulate its application. The debate is spawning new insights on the nature of human life, as well as refining understanding of life's beginnings and endings.^[6] For example, because cloning and stem-cell research entail destruction of some embryos, the issue of protecting innocent human life is a hallmark of the discourse. Scientists are, therefore, active in clarifying distinctions of biological reality. Some are noting that parthenogenetically activated eggs used in stem-cell research cannot result in human birth and, therefore, do not fit into the notion of "conception." Others, while accepting the scientific distinction, disagree with the conclusion. They argue that these embryos are potential human life and, therefore, are sacred and deserving of protection. Perspective on the issue of the beginning of human life is, in effect, expanded from the idea of conception to the notion of potentiality. The public policy of protecting life is thus challenged.

The issue of the ending of human life is similarly germane in this dialog around cloning and stem-cell research. Is human life unnaturally ended when embryos are destroyed in the process of cloning? For those who judge in the affirmative, public policy concerning homicide comes into play. Of course, implicit in this dialog is the question of just what constitutes human life. Some ethicists see a significant distinction between eggs produced artificially from parthenotes and those embryos fertilized naturally. From this sense of the nature of humanity, these ethicists are arguing for public policies that prohibit the sale of human gametes but do not prohibit parthenogenetic cloning per se. Similarly, other ethicists distinguish individuality from personhood. They observe scientifically that the early embryo (called a blastocyst) is not individualized until about 2 weeks into its developmental process. At that point, they note that the organism has the human genome but lacks personhood. Extraction of stem cells for medical or therapeutic purposes can then be distinguished from cloning for reproductive purposes. Based on this sense of human life, they argue for public policies that ban reproductive cloning but allow medical cloning.

This dialog on cloning also illustrates the tension between the interests of the individuals and that of the community. The interests of the polity in defining and defending human life can easily lead to policies that conflict with the interests of individuals, who want reproductive assistance. Moreover, cloning and stem-cell therapy are expensive. Should limited public health-care funds be spent in these dramatic areas that benefit relatively few individuals, or would they more ethically be devoted to preventive health-care needs that benefit the community as a whole? Public budgetary policies face this question.

Additional organizational issues arise when public policy in this area focuses on regulation and limits on research. Would it be managerially possible to develop, monitor, and enforce such restrictions within ethical standards of competence and compliance? As cloning technology continues to evolve, these kinds of questions confront public policy with increasing insistence.

State legislatures have been central in the evolving discourse. In 2004, California enacted a law explicitly supporting stem-cell research with public funds. In 2005, Connecticut, Illinois, New Jersey, and Virginia took similar action. In 2006, Maryland's General Assembly passed a bill after several years of failed attempts, and its experience is particularly illustrative of the nature of the discourse of bioethics in the public polity. Key participants in the Maryland dialog were biotechnological corporations and research universities that stressed the economic factors and benefits involved. Equally outspoken was the Catholic church of Maryland, which stressed concerns of respect for life. In the end, political linguistics was paramount: The discourse that began years before with a prevalence of highly charged and sensitive verbiage evolved into a dialog that consciously avoided trigger words. The language of bioethics in the public policy process continues to evolve, and this development appears to be the key as public policy responds.^[7]

CONCLUSIONS

The subject of biomedical ethics will accelerate its expansion on the public policy-making agenda as the first decade of the twenty-first century proceeds. Continuing and increasing resource limitations auger a higher level of intensity in the area of managerial and organizational policies. Who should get how much health care and when, as well as where should they get it? These will be focal questions for policy makers. Technological advances are also likely to soar, presenting fundamental questions in ever-challenging ways to bioethicists and public policy makers alike. Responses in the form of new public policies arguably will reshape

society in some fundamental aspects. Bioethical analysis will continue to be essential, and its linguistic presentation may be the critical determinant of the policies that ensue.

REFERENCES

1. Loewy, E.; Loewy, R. *Textbook of Healthcare Ethics*; Plenum Press: New York, 2004.
2. Worthley, J. *Organizational Ethics in the Compliance Context*; Health Administration Press: Ann Arbor, MI, 1999.
3. Arras, J.; Steinbock, B. *Ethical Issues in Modern Medicine*; Mayfield: Toronto, 1998.
4. Coleman, G. Take and eat: morality and medically assisted feeding. *America* **2004**, *188* (7), 16–20.
5. Witt, H. Who gets to make decision on end of life? *Chicago Tribune* **2006**, *September 17*, 4.
6. Shannon, T. Human cloning. *America* **2002**, *186* (5), 15–18.
7. Alt, P. The political linguistics of Maryland's stem cell research bill. *Ethics Today* **2006**, *8* (4), 6–13.

Bond Ratings

Robert J. Eger

Andrew Haas

*Public Administration and Urban Studies Department, Georgia State University,
Atlanta, Georgia, U.S.A.*

INTRODUCTION

This article provides a broad overview of bond ratings. It includes the definition of a bond rating, types of bonds, the bond rating process, the history of bond ratings, why government entities get a bond rating, and what information is used to determine a bond rating. Information charts detailing bond ratings by each rating agency are provided.

WHAT IS A BOND RATING?

A bond rating is the assignment of a grade by a private bond rating firm as to the creditworthiness (risk of default) associated with a particular bond issue. It is an evaluation of the probability that an issuer will repay the principal and interest of a bond issue. Bonds are rated by private companies (bond rating agencies) that estimate a bond issuer's fiscal health and assess its capacity to repay obligations in a timely manner.

TYPES OF BONDS

Bonds are financial instruments issued to finance long-term debt. They are issued in two categories: general obligation (GO) bonds and revenue bonds.

General obligation bonds are secured by a pledge of the full faith and credit of the issuing government entity with the tax base of the issuing entity acting as collateral for the bonds. General obligation bonds are considered low risk and do not require high administrative costs. Most government entities are required to hold a referendum before they can issue GO bonds. This can delay issue of a bond but imparts voter consent for the issue.

Revenue bonds, such as housing or transportation bonds, are principally secured by user fees or service charges paid by users of a particular government service. They can be issued by special districts (such as housing or transit authorities) without a formal referendum. This allows for financing of projects that

generate revenue exogenous of principal revenue sources. They are generally considered to be more complicated and higher risk and therefore carry a higher interest rate than GO bonds.

THE BOND RATING PROCESS

The bond issuing government entity hires a bond rating agency when the decision to issue debt is made. All relevant information is sent to the rating agency and is combined with internal data sources and objective research to determine a rating for the bond issue. Quantitative and qualitative data are used to determine the rating. Factors taken into account when determining a credit rating cover four primary areas: economy, debt, finances, and management strategies. There is some variation in the factors based on the rating agency and the type of debt issued.

HISTORY OF BOND RATINGS

Moody's Investor Services was founded in 1909 by John Moody when he began rating railroad bonds. By 1910, Moody broadened his ratings to utility and industrial bonds. Poor's Publishing Company began as a journal reporting financial and operating statistics for most of the American railroads and entered the bond rating business in 1916. In 1941, Poor's merged with Standard Statistics to become Standard and Poor's Corporation (S&P). Fitch ICBA entered the ratings industry in 1924 as Fitch Publishing Company and merged with ICBA Limited in 1997.

When bond ratings were first issued, ratings agencies purchased information from municipalities and resold their analysis to investors in the form of bond ratings. This model changed in the mid-1970s with high default rates in the debt market. Investors wanted reassurance and issuers wanted ratings to bolster confidence in their debt issues. It soon became standard practice to have a credit rating for every debt issue. Information is now free to investors, with payment coming from the issuing government entity.

WHY GET A BOND RATING?

Government entities have a strong incentive to keep their overall bond rating high. A direct inverse relationship exists between the bond rating and interest rates paid on a bond issue. Government entities with high bond ratings have a lower cost of capital and those with low bond ratings have a high cost of capital. These higher costs come in the form of higher interest rates on bond issues. A rating that is too low can remove the government entity from the debt market completely. Decision makers are induced to keep expenses low to help maintain a high bond rating. As an indicator of a government's financial health, bond ratings can have significant influence on government operations. A low or declining rating can indicate that a government is experiencing financial difficulty and might be an investment risk. This increases the government's cost of capital. High or increasing bond ratings mean lower borrowing costs and an enhanced fiscal reputation.

One example of the influence of bond rating agencies is the fiscal crisis of New York City in 1975. Beginning in the 1960s, New York City government increased expenditures for many social welfare and service programs. A general population decline coupled with urban deindustrialization eroded the tax base. Taxes were raised but were not sufficient to compensate for the increased expenditures. In an effort to prop up its finances and pay for the new programs, short-term notes were issued by the city. This cascaded into a series of note issuances to pay back previous short-term notes coming due. The city was downgraded from investment grade to speculative-grade status ("A" to "Caa") by Moody's in October 1975. The result of this downgrade removed the city from the debt market and forced a massive reorganization of the city's finances. The Municipal Assistance Corporation and the Emergency Financial Control

Board were given extraordinary power to remedy this fiscal disaster.

BOND RATING BY AGENCY

S&P uses a 10-symbol system that ranges from AAA, the highest investment quality, to D, the lowest credit rating. S&P considers bonds in the top four grades (AAA, AA, A, and BBB) to be of investment-grade quality. In categories ranging from AA to BB, a plus (+) or minus (−) sign is used to designate relative standing within the rating categories. In general, S&P tends to consider the issuer's economic environment as the most important element in its analysis. It also makes more of a distinction between state and local GO bonds due to certain legal powers the state has that municipalities do not have. S&P seems to view state GO bonds as significantly stronger than those of lower level jurisdictions.

Moody's uses a nine-symbol system that ranges from Aaa, the highest investment quality, to C, the lowest credit rating. Moody's considers bonds in the top four grades (Aaa, Aa, A, and Baa) to be investment-grade quality. Moody's rating focuses on the debt burden and budgetary operations of the issuer. In contrast to S&P, Moody's essentially believes that state and local bonds are not fundamentally different. It applies the same debt and budget concerns to state GO bonds as it does to municipalities.

Fitch uses a rating system from AAA to D. A plus (+) or minus (−) sign is used with each rating except the AAA category.

BOND GRADES

Bond ratings for long-term investment grade bonds are given in Table 1. Those for long-term bonds of speculative grade are given in Table 2.

Table 1 Bond ratings for long-term bonds of investment grade

Rating	Explanation
Moody's Aaa, S&P's AAA, and Fitch AAA	These ratings are the highest grade a bond can be assigned; triple-A bonds have a relatively small degree of risk because payment is secured by a stable revenue source
Moody's Aa, S&P's AA, and Fitch AA	These ratings are similar to that of triple A, but differ only in that the revenue sources for double-A rated bonds are slightly less secure than the revenue sources of triple-A bonds
Moody's A, S&P's A, and Fitch A	Considered upper-medium grade, but revenue sources are relatively susceptible to fluctuations in relevant economic conditions
Moody's Baa, S&P's BBB, and Fitch BBB	Medium-grade obligations that are adequately protected and secured, but nonetheless may be unreliable if relevant economic conditions have long-run adverse effects on revenue source

[From Public Finance Department, Moody's Investor Service. *An Issuer's Guide to the Rating Process*; New York, NY, 1993 (information pamphlet); Standard and Poor's Corporation. *Municipal Finance Criteria*; New York, NY, 1994 (information pamphlet); Fitch Investor Service. *Fitch Ratings*; New York, NY, 1994 (information pamphlet).]

Table 2 Bond ratings for long-term bonds below investment grade

Rating	Explanation
Moody's Ba, S&P's BB, and Fitch BB	Lower-medium-grade obligations that are presently adequately protected and secured, but represent long-term risk whether relevant economic conditions are favorable or not
Moody's B, S&P's B, and Fitch B	These bonds are presently adequately protected and secured and represent risk regardless of economic conditions. In addition, it is likely that future relevant economic conditions will be unfavorable, thus intensifying the probability of default
Moody's Caa, S&P's CCC, and Fitch CCC	Besides future risks typical of bonds in the previous category, these are presently not adequately protected and secured, as present relevant economic conditions pose a threat to revenue source
Moody's Ca, S&P's CC, and Fitch CC	High degree of present and future risk; greater chance of default by issuer; debt issued in same conditions as those that produced CCC rating of a prior issue; given CC rating because of additional insecurity of being issued after CCC bonds
Moody's C, S&P's C, and Fitch C	For S&P's and Fitch, debt issued at same conditions as those that produced a CCC rating in a previous issue is given a C rating because of additional insecurity of being issued after CCC bonds; for Moody's, these are "the lowest rated class of bonds, and issues so rated can be regarded as having extremely poor prospects of ever attaining any real investment standing"
S&P's CI	Reserved for income bonds on which no interest is being paid
S&P's D and Fitch DDD, DD, and D	Assigned these ratings when payment is due and issuer defaults
S&P's and Fitch plus (+) or minus (-)	Indicate relative standing within the major categories from AA to CCC
Moody's "1"	Used to distinguish best bonds in each of five categories, Aa, A, Baa, Ba, and B

[From Public Finance Department, Moody's Investor Service. *An Issuer's Guide to the Rating Process*; New York, NY, 1993 (information pamphlet); Standard and Poor's Corporation. *Municipal Finance Criteria*; New York, NY, 1994 (information pamphlet); Fitch Investor Service. *Fitch Ratings*; New York, NY, 1994 (information pamphlet).]

CONCLUSIONS

Bond ratings are a powerful tool that municipalities use to raise capital for day-to-day activities or specific projects that require long-term financing. Municipalities must be mindful of their bond rating when it comes to expenditures, since there is a direct inverse relationship between a bond rating and capital costs. A change in a municipal bond rating can be an important signal to concerned parties if it warns of information that was previously not available on the financial condition of a municipality. Bond rating agencies provide an important private sector "check" on the public sector to ensure that spending is closely monitored.

BIBLIOGRAPHY

- Cantor, R.; Packer, F. The credit rating industry. Fed. Reserve Bank N.Y. Q. Rev. **1994**, Summer/Fall, 1–26.
- Congressional Budget Office. The causes of New York City's fiscal crisis. Polit. Sci. Q. **1975–1976**, 90 (4), 659–674.
- Fabozzi, F.J.; Fabozzi, T.D.; Feldstein, S.G. *Municipal Bond Portfolio Management*; Irwin Professional Publishing: Burr Ridge, 1995; 179–189.
- Hackworth, J. Local autonomy, bond-rating agencies and neoliberal urbanism in the United States. Int. J. Urban Reg. Res. **2002**, 26 (4), 707–725.
- Kilger, D.; Sarig, O. The information value of bond ratings. J. Finance **2000**, 55 (6), 2879–2902.
- Lamb, R.; Rappaport, S.P. *Municipal Bonds*, 2nd Ed.; McGraw-Hill Book Company: New York, 1987; 36–50.
- Lipnick, L.H.; Rattner, Y.; Ebrahim, L. The determinants of municipal credit quality. Gov. Finance Rev. **1999**, December, 35–41.
- Loviscek, A.L.; Crowley, F.D. What is in a municipal bond rating? Financ. Rev. **1990**, 25 (1), 25–53.
- Municipal Bond Information Guide. <http://www.southwest.msus.edu/rdic/munibondtoc.html> (accessed November 1, 2004).
- Roden, P.F.; Bassler, J. Effect of underwriter prestige on the interest cost of municipal bond offerings. Financ. Rev. **1996**, 31 (3), 641–666.
- Sylla, R. A historical primer on the business of credit ratings. Prepared for Conference on "The Role of Credit Reporting Systems in International Economy," The World Bank: Washington, DC, March 1–2, 2001.
- White, L.J. The SEC's other problem. Regulation **2002/2003**, 25 (4), 38–42.

Bond Valuation

Richard L.B. LeCompte

*Department of Finance, Real Estate and Decision Sciences, Wichita State University,
Wichita, Kansas, U.S.A.*

INTRODUCTION

Bond valuation is the process of valuing long-term debt securities. Bonds are fixed-income financial claims issued to investors by corporations, governments, and government-sponsored enterprises. There are many different types of bonds with a vast array of features that vary widely. Bond types range from the “plain vanilla” described in most textbooks to exotic high-risk offerings, some that even have equity features and options embedded in them. The process of valuing each type of bond, however, has a similar starting point. The starting point in valuing bonds, as it is with other financial securities and investment projects, is a discounted cash flow (DCF) analysis. This approach takes forecasted cash flows occurring over the life of the instrument and discounts these projected values to account for the risk inherent in owning these instruments and the time value of money. Other factors may play a significant role in the pricing, but the key elements in bond valuation are the estimation of the cash flows and determining the appropriate discount rate to be used in calculating the present value or price of the bond in question.

BOND DEFINITION

A bond is a long-term contractual debt obligation that is issued by a corporation or governmental entity. In return for the purchase price, the bond’s present value, the bond issuer promises to pay the investor a specific series of payments on designated dates in the future. These provisions classify bonds as fixed-income securities. A note, which is a financial claim similar to a bond, differs in that it has shorter original maturity than a bond, less than 10 years, whereas a typical bond has an original term to maturity of 10 years or greater. Typical bonds have 20- and 30-year maturities. The most basic type of bond is a fairly simple debt instrument that provides the bondholder the promise of: 1) a series of equal semiannual coupon payments, an annuity, over the life of the bond; and 2) a much larger lump-sum payment on the bond’s maturity date. This final payment is the par value of the bond. It may also be referred to as the bond’s face or maturity value and

is analogous to the principal value of a loan. For standard bonds, par value is typically \$1000 and the annual coupon payment is a fixed percentage of that value, but the coupon payments are usually made on a semi-annual basis. Bonds are normally quoted on a “per 100” basis, where a bond value quoted as 96 3/4 sells for \$96.75 per \$100 of par value, meaning that the bond with the typical par value of \$1000 sells for \$967.50. Bonds are generally sold “with interest.” This means that the purchaser must pay the seller the quoted price plus the proportion of the next coupon interest payment that has accrued on the bond since its last coupon payment.

FEATURES OF A BOND AND ITS INDENTURE

The promises or contractual obligations of a corporate bond issuer and rights of investor are set forth in great detail in what is known as an indenture. The indenture normally specifies a trustee, whose function is to represent the rights of the bondholders. The trustee is also a party to the contract and is responsible for enforcing the contracts within the indenture. Specific covenants in the indenture detail features of the bond such as the coupon rate, security or collateral provided by the issuer, maturity date, call and sinking fund provisions, and equity features such as conversion rights or warrants that may be attached to the issue. The indenture also contains covenants to protect the investor’s rights relative to the issuer. Typical protective covenants include subordination clauses, dividend restrictions, and limitations on the values for financial ratios such as a firm’s current ratio or its debt ratio. Protective covenants may also be used to protect the bondholder in the event of the firm’s merger or a significant change in organizational structure or control, and provisions may also set limitations on the sale or disposition of assets. Violating the terms of an indenture is classified as a default that requires the trustee to undertake certain specified actions in order to protect the rights of the bondholder. Default also occurs when the issuer of the bond fails to fully pay either principal or interest when due. Bondholders’ financial claims have priority rights over preferred and common

shareholders with respect to the income and assets of the corporation. This makes bonds less risky investments compared to preferred and common shares, which are paid only after bondholder claims are satisfied.

VALUATION OF BONDS

To value a bond or any other financial asset, one must compute the present value of all remaining cash flows using a DCF analysis. In the case of a bond, the cash inflows consist of the scheduled coupon payments and the return of principal at maturity. In addition to these payments, contractual features contained in the bond's indenture may also influence the value of cash flows and when they are received. Using DCF analysis, the value of a bond is the present value of the coupon payments plus the present value of the par value of the bond discounted at the appropriate rate of interest. This interest rate is the bond's yield to maturity (YTM), and it is the discount rate that equates the present value of the cash inflows with the current price or value of the bond. The YTM used in bond valuation is analogous to the internal rate of return (IRR) utilized in the evaluation of the return on real capital projects.

Determining the value of a bond is similar to valuing any project or investment using DCF analysis. First, the cash flows are estimated, and then one must determine the appropriate discount rate to use in calculating the present value of the expected future cash flows. Even though a bond's coupon payments and principal repayment are specified, there is a probability of default, and this impacts the discount rate utilized by the investor. The discount rate is the investor's required rate of return and is a risk-adjusted rate in accordance with the risk of the bond, the investor's opportunity cost, and the investor's level of risk aversion. Most bond valuations typically utilize only one rate to discount all cash flows throughout the life of the bond, and it is assumed that this rate remains constant throughout. This assumption can be relaxed to produce a bond valuation framework where different required rates of return are used for different time periods. The discount rate used for any bond can also be viewed as the real required rate of return plus various risk premiums for items such as the expected rate of inflation, default risk, liquidity or lack of it, tax characteristics, callability, sovereign risk, industry risk, foreign exchange risk, etc.

The value of a bond is equal to the present value of an annuity (the coupon payments) plus the present value of a single final payment at maturity (the

par value). In general, the price of the bond can be computed using the following formula:

$$\text{Price} = \sum_{t=1}^N \frac{C_t}{(1+r)^t} + \frac{M}{(1+r)^N}$$

where N is the number of periods until maturity, t is the time of the periodic payment, C is the annual coupon payment in dollars, r is the required rate of return or YTM, and M is the par or maturity value.

Since most bonds pay coupon payments semiannually, the above equation becomes

$$\text{Price} = \sum_{t=1}^{2N} \frac{C_t/2}{(1+r/2)^t} + \frac{M}{(1+r/2)^{2N}}$$

For example, assume that a \$1000 par value bond pays a stated 10% annual coupon rate, resulting in an annual coupon of \$100. The bond's maturity date is five years from today. Dividing the annual payment by two leaves the investor with a \$50 semiannual coupon payment for 10 semiannual periods. Assume that the bond has a YTM of 8%, indicating a 4% semiannual rate, using the following equation; the value of the bond would be \$1081.11:

$$\begin{aligned} P &= \frac{50}{(1.04)^1} + \frac{50}{(1.04)^2} + \frac{50}{(1.04)^3} + \frac{50}{(1.04)^4} \\ &\quad + \cdots + \frac{50}{(1.04)^{10}} + \frac{1000}{(1.04)^{10}} \\ &= \$1081.11 \end{aligned}$$

where \$50 is the semiannual coupon payment in dollars, 4% is the semiannual interest rate, \$1000 is the par value, 10 is the maturity in semiannual periods, and 8% is the YTM, *annual rate*.

As can be seen in this example, the bond above sells for more than its par value, or the bond is selling at a *premium* to par. When the discount rate used is less (greater) than the coupon rate paid on the bond, then the bond sells for a price that is greater (less) than the par value. Bonds selling below their par value are called *discount* bonds.

There are also discount bonds called zero-coupon bonds. For a zero-coupon bond, there are no periodic coupon payments, thus the name zero-coupon bond or zero. Zeroes have only one payment, the par value of the bond repaid at maturity. To value a zero-coupon bond, one can use the valuation equation above and simplify it by dropping the first term and using the remainder. That is, the present value or price of a zero-coupon (single payment) bond is

just the present value of the payment that occurs at maturity:

$$\text{Price} = \frac{M}{(1 + r)^N}$$

where N is the number of periods until maturity, r is the required rate of return or YTM, and M is the par value.

The following valuation of a zero-coupon bond is based upon the cash flows previously discussed in regarding to the coupon-paying bond. In this case, the par value is \$1000. There are no coupon payments and there are 10 semiannual periods until maturity, with a semiannual discount rate of 4%:

$$P = \frac{\$1000}{(1.04)^{10}} = \$675.56$$

In the coupon-paying bond example, this is just the value of the last term in the equation.

Some zero-coupon bond are issued by the U.S. government and are known by their acronym STRIPS. This is short for Separate Trading of Registered Interest and Principal of Securities. These zero-coupon bonds are created by the stripping of U.S. government bonds and notes into separate securities for each of the individual coupon payments and the principal payment for sale to the public. Treasury bonds and notes with original maturities that are greater than 10 years are eligible for the STRIPS program, and these securities are direct obligations of the U.S. government. The creation of these securities clearly shows how a coupon-paying U.S. government bond can be viewed and valued as a portfolio of zero-coupon bonds (STRIPS). Other types of zero-coupon instruments include

U.S. government issued treasury bills (short-term debt obligations) and zero-coupon bonds issued by some corporations.

FUNDAMENTAL BOND VALUATION RELATIONSHIPS

There are several fundamental relationships that are important to understand regarding the valuation of bonds. Since the cash flows for a bond are fixed at the time of issue, any change in the required rate of interest requires that the price of a bond move in the opposite direction to the change in the interest rate associated with the security. Fig. 1 shows the relationship between the price of a bond and its YTM. The bond used in the diagram is a bond with an 8% YTM, \$1000 par value with a 40-year maturity, and an 8% annual coupon interest payment.

As can be seen in Fig. 1, at an 8% YTM, the bond's price is equal to its par value, \$1000. As the YTM rises, the price falls. Conversely, as the YTM on the bond falls, its price rises. The price-yield relationship illustrated in the diagram is specific to the bond in question. Different bonds with different characteristics have graphic profiles that differ from this 8% coupon, 40-year maturity bond. Longer-term bonds have a steeper slope in that they are more price sensitive than shorter bonds. For bonds alike in all respects except time to maturity, a given change in the market rate of interest causes the price of the longer maturity bond to change more in percentage terms than the price of the shorter maturity bond. Low coupon bonds also create a graphic profile with a steeper slope in that they are more price sensitive than high coupon bonds. That is, for bonds alike in all respects except coupon size, a

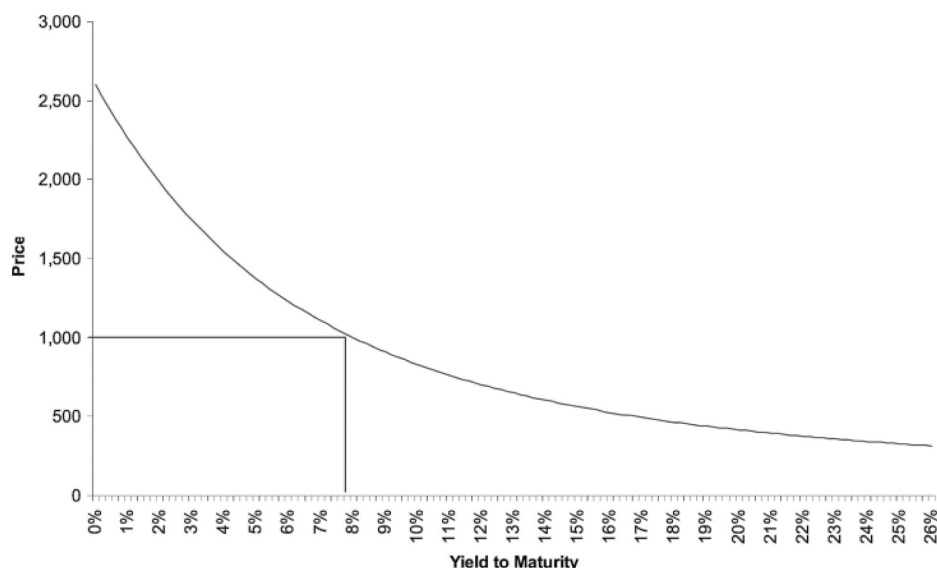


Fig. 1 Bond valuation—8% annual coupon rate, 40-year maturity. The inverse relationship between price and yield to maturity on a bond.

given change in the market rate of interest causes the price of the lower coupon bond to change more in percentage terms than the price of the higher coupon bond. The price risk of a bond is characterized by the changes in the bond's value for a given change in interest rates and is shown graphically by comparing a bond's price with the range of possible YTM's, as is done in Fig. 1. The two prime factors determining price risk for fixed-income securities are the coupon and maturity effects described above. The bond's price risk can also be measured by its duration, which in a single metric accounts for both of these effects.

Other factors that have an impact on the value of a bond are such issues as its marketability and liquidity in secondary markets, tax treatment, embedded options, credit risk, and features such as callability and convertibility. A call feature on a bond allows the issuer to retire the bond prior to its maturity. The call option is generally beneficial to the issuer and detrimental to the investor and, therefore, results in a lower bond price than that of a non-callable bond, *ceteris paribus*. In the indenture of a bond, the call provision specifies a price at which the bond could be called. This is typically the par value plus a call premium. The investor is generally protected from a call for a certain specified period, and the value of the call premium generally declines as the bond gets closer to its maturity. Conversion provisions are significant features for some bonds. The conversion option may enhance the bond's value in that it allows the bondholder, under certain conditions, to convert the bond into shares of the firm's common stock. Municipal bonds are also valued using the DCF methodology. The tax exemption provided for municipal bond interest payments enhances the value of the cash flows from these bonds, especially to those individuals in the highest marginal income tax brackets. The impact of the tax exemption reduces the required rate of return on municipal bonds relative to bonds with taxable interest payments, *ceteris paribus*. Another significant factor in the valuation of bonds is the credit rating assigned by various rating

agencies. These credit ratings exhibit their impact in the size of the default risk premium, which investors require to be added to the discount rate they apply to risky bonds.

CONCLUSIONS

This article has provided a presentation of the DCF methodology utilized to value bonds. Specifically, the value of a bond is equal to the present value of its expected cash flows, which have been discounted using an appropriate risk-adjusted discount rate based upon the bond's characteristics, investor preferences, and time value of money. Specific examples have been provided for valuing a "plain vanilla" coupon-paying bond and the valuation of a zero-coupon bond. In addition to the DCF approach to valuation, it is recognized that other factors and features included in the bond's indenture may alter value. Various pricing relationships for bonds have also been discussed in the context of the risk created by various characteristics of a specific bond. Also, the similarity between valuing bonds and the valuation of other assets has been discussed. Further information regarding bond valuation is available in the references provided.

BIBLIOGRAPHY

1. Bodie, Z.; Kane, A.; Marcus, A. Bond prices and yields. In *Investments*, 6th Ed.; McGraw-Hill Irwin: New York, 2005; 447–478.
2. Fabozzi, F. *Bond Markets, Analysis and Strategies*, 3rd Ed.; Prentice Hall: Upper Saddle River, NJ, 1996.
3. Livingston, M. Coupon-bearing bonds. In *Money and Capital Markets*, 3rd Ed.; Blackwell: Cambridge, MA, 1996; 187–223.
4. Van Horne, J. Prices and yields for bonds and money market instruments. In *Financial Markets: Rates and Flows*, 6th Ed.; Prentice Hall: Upper Saddle River, NJ, 2001; 48–58.

Budget Execution and Management Control

L. R. Jones

*Graduate School of Business and Public Policy, Naval Postgraduate School,
Monterey, California, U.S.A.*

Jerry McCaffery

*Department of Systems Management, Naval Postgraduate School,
Monterey, California, U.S.A.*

Fred Thompson

*Atkinson Graduate School of Management, Willamette University,
Salem, Oregon, U.S.A.*

INTRODUCTION

Once funds have been appropriated for spending in the sequence of budget formulation, negotiation, and enactment, budgets then must be executed as planned and authorized.^[1] However, budgets also must be responsive to changing circumstances during the period of spending execution. All organizational entities employ management control systems with imbedded incentives, sanctions, and a myriad of rules to control spending and the budget execution process. Agreed upon authority relationships, institutional roles and responsibilities, job performance criteria, rules and regulations, incentives and motivational structures, and intra-organizational arrangements comprise the management control systems to provide accountability in an effort to stimulate decision making that results in efficient and effective use of funds.^[2] Various levers of management control are used in the private sector to achieve organizational objectives.^[3] In the public sector, these levers of control are employed by elected officials and, more typically as a result of delegation, by budget control agents and agency managers in budget execution. Control systems and controls have important consequences for how managers behave—this is the explicit purpose of these systems and levers. In budget execution, rules may be designed to tightly control what money is spent on before it is spent (ex ante controls). Alternatively, controls may be applied after the fact, focused on measuring, rewarding, and sanctioning the results of managerial action after spending has occurred (ex post controls). Correctly designed control systems help to stimulate efficient use of resources and development and use of appropriate reporting systems. Conversely, inappropriate control systems and controls tend to create budget execution procedures and processes that are simultaneously over-controlled and out of control. Where budget execution controls are inappropriate for what

is demanded of them and do not report correct information in a timely manner, they will fail to stimulate efficient spending decisions and management behavior and, paradoxically, may create or exacerbate waste, fraud, and other forms of funding misuse.

DESIGNING THE SYSTEM

Critical choices must be made in designing the appropriate budget execution management control system. The successful design should fit the objectives of control and the nature of the entity to be controlled. For example, increasing costs and homogeneous outputs imply one kind of design, while decreasing costs and heterogeneous outputs imply another. The purpose of the analysis of budget execution and management control provided here is to improve understanding of control dynamics, incentives, disincentives, and the behavior of the various participants in the budget execution and management control process. Our analysis focuses most closely in this regard on the roles of central executive branch budget agency controllers, agency budget officials, and program managers.

The other purpose served by this analysis is to ask whether there is cause to change budget execution and management control processes, as they now operate in many, if not most, public-sector organizations. If so, what direction should change take? The analysis delineates alternative types of control applied in executing budgets and the rationale for employing these different methods. A distinction is drawn between budget execution control intended to influence the behavior and performance of managers of government programs from controls applied to affect independent private-sector firms that contract to deliver goods or services to the public on behalf of government. The central theme of this analysis is that budget execution and management control system design should fit the

objectives of control and the nature of the entity to be controlled.^[4]

Perhaps the most challenging task in budgeting is to execute well, so that the best program outcomes are achieved with some degree of efficiency and a genuine concern for the proper use of public funds. Budget execution skill is required to respond to inevitable contingencies that arise to complicate the implementation of programs in the manner planned and according to the commitments made in budget formulation. Accountability and management control must be maintained, and employees must be motivated to perform to high standards. At the same time, uncertainty must be accommodated, as budgets are executed to meet the demands of the real world—demands that often cannot be anticipated and budgeted accurately in advance. While the highly visible politics of budget formulation and enactment garner the attention of the media and of budget negotiators and critics, it is only in the phase of budget execution that performance, outputs, and outcomes are obtained. In our assessment of budgeting, we must not lose sight of the fact that it is consumption (use) of assets in attempt to achieve the objectives of organizations and governments that matters most if our goal is to learn to spend and manage better.

BUDGET EXECUTION AND MANAGEMENT CONTROL ISSUES

Budget execution typically is highly regulated to control what program managers may and may not do. Controllers are driven by the objective of insuring that budget appropriations in total and by legally segregated account are not overspent by programs.^[4] However, controllers also must be concerned with underexecution. Department and agency budget officers and program managers do not want to under-spend and thereby lose claim to resources in the following year. Central executive budget office controllers do not want programs to execute without good cause, so that money not used in the manner justified in budget formulation may be withdrawn from program managers. Executive controllers want to be able to withdraw funds from programs to protect the integrity of the appropriation process and to reallocate money to areas where it will be spent efficiently in response to client demand. For these reasons, budget execution typically is monitored and controlled carefully by agency budget staff and central executive budget controllers. Often, execution is also monitored closely by legislative oversight committees and their staffs out of a desire to insure that legislative will is implemented faithfully, and to make sure that benefits are distributed to the clients targeted in the appropriation process.^[4] Because the electoral fortunes of

legislators are tied to some degree to the public perception that they are solving the problems and meeting the demands of their constituents, legislators have considerable interest in budget execution control.

BUDGET EXECUTION TECHNIQUES

Among the techniques used to control budget execution, variance analysis is probably the most familiar. Controllers and budget officials in government program offices monitor the differences between projected and actual revenues and expenses in total and by account. They monitor revenue and expense rates against allotment controls by quarter, month, week, and day—temporal control generally is required by the allotment process. Other variables monitored are purpose of expense relative to budget proposal and appropriation rationale, and location of revenues and expenses by unit and, at times, by geographical location. Monitoring of actual revenue and expense rates as well as program output and demand, where measurable, is done to compare current spending to proposals made in the budget for the next year that is under negotiation at the same time as the fiscal year budget is expended. Comparisons are made to historical revenue and spending trends, in some instances, to better understand how current programs are performing. Budget execution monitoring and control are particularly important toward the end of the fiscal year for reasons stated above, to avoid over- and underexpenditure relative to appropriation.

BUDGET EXECUTION MANAGEMENT CONTROL CHOICES

Research in public finance has paid considerable attention to budget formulation but has tended to ignore budget execution.^[5–7] The reasons for this oversight are understandable. Government budgets are formulated in public, and the issues debated during this stage of the public spending process are dramatic and crucial. On the other hand, budgets are executed in private, and the issues raised in their execution are often mundane. Because of this selective attention, observers and participants in the public spending process understand program analysis far better than controllership. Consequently, the conduct of program analysis has come to be guided by a fairly coherent set of professional standards. There is agreement on what is good analysis and what is good accounting practice. Although the design and operation of control systems can profoundly influence governmental performance, budget execution controllership is not guided by a coherent set of professional standards.

Without appropriate performance standards, budget officers cannot be held accountable for performance of this function. Consequently, control systems are not designed to optimize the quality, quantity, and price of goods and services purchased with public money, but “to facilitate the controller’s [other] work.”^[4]

EX ANTE AND EX POST CONTROLS DEFINED

The choice of whom to subject to controls and when to execute those controls is not easy. The control system designer has at least four options. First, the subject may be an organization or an individual. Second, controls may be executed before or after the subject acts. The former may be identified as *ex ante* and the latter as *ex post* controls.^[8] *Ex ante* controls are intended to prevent subjects from doing wrong things or to compel them to perform well. Necessarily, they take the form of authoritative commands or rules that specify what the subject must do, may do, and must not do. Subjects are held responsible for complying with these commands, and the controller attempts to monitor and enforce compliance. In contrast, *ex post* controls are executed after the subject decides on and carries out a course of action and after some of the consequences of the subject’s decisions are known. Because bad decisions cannot be undone after they are carried out, *ex post* controls are intended to motivate subjects to make good decisions. Subjects are held responsible for the consequences of decisions, and the controller attempts to monitor consequences and rewards or sanctions accordingly. The control system designer may choose between four distinct design alternatives: individual responsibility, *ex ante* or *ex post*, and organizational responsibility, *ex ante* or *ex post*.

PURPOSES OF EX POST CONTROLS

Ex post financial controls are used to reveal demand. They are executed after operating decisions have been made, after asset acquisition and use decisions have been carried out, and after output levels have been monitored. Their subject may be a free-standing organization, such as a private contractor or a quasi-independent public entity, or an individual manager subordinate to the controller, such as a responsibility center or program manager within a government agency.

In a free-standing organization, the structure of authority and responsibility within the organization is assumed to be an internal matter. The controller establishes a price schedule and specifies minimum service quality standards or a process whereby these

standards are to be determined. This price or cost schedule may entail all sorts of complex arrangements, including rate, volume and mix adjustments, and default penalties. Where one organization can optimally supply the entire market, the controller may grant it a monopoly franchise, for example, in garbage collection for a small town or neighborhood. The significant characteristic of this approach is that a unit price cost schedule remains in effect for a specified time period.^[9–11] This means that the government’s financial liability will depend on the quantity of service provided and not on the costs incurred by the organizations supplying the service.

Where this budget control system design is employed, for example, where a municipality purchases gasoline at the spot-market price or where states commit to pay free-standing organizations such as a university a fixed price for performing a specific service such as enrolling students or treating heart attacks, or where the Air Force buys F-16s for a fixed price, the controller must rely upon interorganizational competition to provide sufficient incentives to service suppliers to produce efficiently and make wise asset acquisition and use decisions. If interorganizational competition is effective, those organizations that do not produce cost-effectively will not survive.

However, even where the declining marginal cost of the service in question makes monopoly appropriate, *ex post* controls can still be employed. This is done in businesses and business-like public-sector enterprise organizations by holding managers responsible for optimizing a single criterion value, subject to a set of constraints.^[4] A prevalent approach through which this control system design is employed at the federal level is the revolving fund.^[12,13] In this circumstance, the controller gives the manager the authority to make spending decisions and to manage cash flow for an extended period of time, e.g., for one year or more. For example, the manager is given the authority to make spending decisions to acquire and use people and assets, subject to output quality and quantity constraints determined by customers, and is held responsible for supplying agreed upon outputs while minimizing costs.

THE LOGIC OF EX POST CONTROL

Large private-sector firms produce comprehensive operating reports describing the performance of responsibility centers and programs, but their budgets seldom are detailed. The logic of *ex post* control is that the purpose of the budget is to establish performance targets that are high enough to elicit from the organization’s managers their best efforts. Such budgets might

contain only a single number for each responsibility center—an output quota, a unit-cost standard, a profit, or a return-on-investment target.

Under this approach to budget control, the structure of authority and responsibility within the organization is of interest to the financial controller. The effectiveness of this design depends on the elaboration of well-defined objectives, accurate and timely reporting of performance in terms of objectives, and careful matching of spending authority and responsibility. Its effectiveness also depends upon the clarity with which individual reward schedules are communicated to responsibility center managers and the degree of competition between alternative management teams. Finally, under this approach, the financial liability of government depends on the costs incurred in providing the service and not merely on the quantity or quality of the service provided.

PURPOSES OF EX ANTE CONTROLS

In contrast to ex post budget controls, ex ante controls are demand concealing. Their distinguishing attribute is that the controller retains the authority to make or exercise prior review of spending decisions. Ex ante financial controls are executed before public money is obligated or spent, and they govern the service supplier's acquisition and use of assets. Examples of ex ante financial controls include object-of-expenditure appropriations, apportionments, targets, position controls, and fund and account controls that regulate spending by account and the kind of assets that can be acquired by governmental departments and agencies. Such controls also govern the behavior of private contracting entities that supply services to government or to clients on behalf of governments.

Execution of ex ante controls requires assessment of the consequences of asset acquisition decisions. This consideration may be implicit, as it is in the execution of the traditional line-item budget and basic-research contracts, or explicit, as in the execution of performance and program budgets and systems development contracts. It is often influenced by information on current and past performance, but consideration of the consequences of spending decisions is always prospective in nature.

The logic of ex ante control is that constraining managerial discretion is the first purpose of budget execution. Because the degree of constraint will depend upon the detail of the spending plan, as well as the degree of compliance enforced by the controllers, these budgets need to be highly detailed. A department or agency budget must identify all asset acquisitions to be executed during the fiscal year and make it clear who is responsible for implementation.

Under ex ante budget control, service-supplying organizations must be guaranteed an allotment of funds in return for continuously providing a service for a specified period. The service provider will assume some responsibility for managing output levels or delivery schedules, service quality, or price to the government customer. Government is directly responsible for all legitimate costs incurred in the delivery of services, regardless of the actual quantity or quality of the services provided.

ENFORCING EFFICIENCY WITH EX ANTE CONTROLS

Where a manager seeking to increase a budget is subject to tight ex ante controls, the controller can enforce efficiency during the budget period by requiring affirmative answers to the following questions:

1. Will a proposed change permit the same activity to be carried out at a lower cost?
2. Will higher priority activities be carried out at the same cost?
3. Will proposed asset acquisitions or reallocations of savings support activities that have a higher priority than those presently carried out?

When operating managers are faced with these criteria, they respond appropriately. Controllers approve most changes in spending plans proposed by operating managers because only mutually advantageous changes will be proposed in most circumstances.

WHEN TO DELEGATE EX ANTE CONTROLS

When line-item or lump-sum appropriations have a comparative advantage, to say that ex ante controls are a necessary means of reinforcing the controllers' bargaining power should not imply that tight ex ante controls always must be administered by them. Under certain conditions, authority to spend money, transfer funds, fill positions, etc., may be delegated to subordinate managers. The threat of reimposition of ex ante controls will be sufficient to insure that the manager's behavior corresponds to the controllers' and elected officials' preferences. In order for such delegation to take place, the following conditions must be present:

1. Reimposition of controls must be a credible threat.
2. The gain to the manager from delegation must more than offset the associated sacrifice in bargaining power; the manager of an agency in the stable backwaters of public policy has little to

gain from relief from ex ante controls if the price of such relief is a change in business as usual.

3. Controllers must be confident that their monitoring procedures, including postaudit, will identify violations of “trust.”

Clearly, all long-term relationships with private contractors and government goods and service suppliers rely to some degree on ex ante controls. Even the operation of fixed-price contracts requires prior specification of product quality standards and delivery schedules. But, flexible-price, cost-plus type contracts and appropriated budgets require considerably higher levels of reliance on ex ante controls and also on monitoring and enforcing compliance. And, the cost of tightly held budget execution control is high.

ACCURATE RECORDS AND EFFICIENCY

At the least, adoption of the budget execution control systems described herein means that controllers must take steps to ensure that suppliers fairly and accurately recognize, record, and report their expenses. This, in turn, requires careful definition of costs and specification of appropriate account structures, accounting practices and internal controls, direct costing procedures, and the criteria to be used in allocating overheads. Still, accurate accounting does not guarantee efficiency. Even where the service supplier's financial and operational accounts completely and accurately present every relevant fact about the decisions made by its managers, they will not provide a basis for evaluating the soundness of those decisions. This is because cost accounts can show only what happened, not what might have happened. They do not show the range of asset acquisition choices and trade-offs the supplier considered, let alone those that should have been considered but were not.

APPROPRIATE DESIGN OF BUDGET CONTROL SYSTEMS

Under line-item or lump-sum budgets and flexible-price contracts, asset acquisition decisions must be made by the contractor, but the contractor cannot be trusted completely to make them efficiently. Consequently, the contractor must be denied some discretion to make managerial decisions. The fundamental question is, how much must be denied? To what extent should government officials or their controller agents regulate, duplicate, or replace the contractor's managerial efforts?

This question must be addressed, because oversight is costly in terms of monitoring and reporting costs,

and also because of the benefits sacrificed due to failure to exploit the contractor's managerial expertise. The controller and the government official will seldom be more competent to make asset acquisition decisions than the contractor. The answer to this dilemma is that controllers and officials should do the minimum necessary, given the incentives faced by and the motivations of the contractor. Still, at times, the minimum necessary is a great deal. This decision depends on circumstance and the controller's skill in exploiting the opportunities created by the contractor's response to institutional constraints. In other words, all long-term relationships between government officials and contractors must rely on incentives, even those governed by lump-sum budgets and flexible-price contracts. The difference is that when these control system designs are employed, the incentives are deeply embedded in the process of budget and contract execution.

Budget execution management control should be matched to circumstances: increasing costs and homogeneous outputs imply one kind of design, while decreasing costs and heterogeneous outputs imply another. However, what we observe in practice is that this match is not always achieved. Controllers tend to rely on monopoly supply and ex ante controls.^[14–17] This combination cannot be appropriate for every service to which it is applied. Evidence can be marshaled to show that a variety of services might be performed satisfactorily by competing organizations, including in air traffic controls,^[18] custodial services and building maintenance,^[19,20] day-care centers,^[19] electrical power generation,^[19] fire protection services,^[21,22] forest management,^[23] management of grazing lands,^[24] hospitals and health care services,^[25] housing,^[26] postal services,^[25] prisons and correctional facilities,^[25] property assessment,^[27] refuse collection,^[28,29] security services,^[25] ship and aircraft maintenance,^[19,30] urban transit,^[25] and wastewater treatment.^[25] Furthermore, even when controllers eschew monopoly supply, they frequently fail to fully exploit the benefits of competition. In New York City, for example, the municipal social services agency acquires child-care services for its clients from public and private day-care centers. Public centers are subject to the full panoply of ex ante controls associated with lump-sum appropriation budgets, and private centers are subject to those associated with flexible-price contracts.

DEPARTMENT OF DEFENSE PRACTICES

To cite another example, Department of Defense (DoD) policy often has restricted the use of flexible-price contracts to situations characterized by considerable procurement risk, e.g., in R&D projects. In other contracts, the degree of incentive is supposed to be

calibrated to project risk. The first ship in a multiship construction program is supposed to be constructed under a cost-plus-fixed-profit contract, while later editions are supposed to be built under fixed-price contracts, on the assumption that experience permits the contractor to manage to a narrower range of cost outcomes and to assume a greater share of the risk burden. It may be observed that while DoD generally makes the proper transition from cost-plus to award-fee contracts as it moves from design to prototype development, it may not make the transition to a fixed-price contract for downstream production.^[31] To be consistent with the logic advanced here, selection of defense system production suppliers should be reduced to a question of cost and price search. Acquisition should be based upon fixed-price contracts, awarded by competitive bidding. Where production volume is sufficient, DoD should try to maintain long-term contractual relations with two or more producers, as multisourcing permits price search at each negotiation of the relationship between DoD and its suppliers.^[32] Nevertheless, winning a contract to develop a weapons system continues to be tantamount to winning subsequent production contracts in a high proportion of cases. Finally, not only do controllers tend to subject government agencies and contractors to a wide array of ex ante controls, they often hold them to tight output, quality, and service delivery schedules. Performance targets that can be met all of the time are not very ambitious.

CAUSES OF BUDGET AND MANAGEMENT CONTROL MISMATCH

What accounts for mismatches between how budgets are controlled in practice and the approach advanced here, which relies on a more sophisticated set of controls dependent on analysis of cost structure (increasing or decreasing) and type of output (homogeneous or heterogeneous)? One explanation is ignorance of consequences on the part of the controllers and elected officials. Also, some of the empirical data required to employ the control criteria outlined here are often unavailable. The most critical gap in this knowledge is how costs vary with output. Definitions and measurements of service outputs and activities are often inadequate. Insufficient effort has been made to correct this situation in most public organizations. Of the two tasks, getting knowledge about the shape of cost functions is the more difficult. But if we first answer the question, “Cost to do what?,” this knowledge can be derived deductively in a manner similar to the methods used in cost accounting and conventional price theory. Cost and supply analysis can yield highly useful information about marginal and average costs. Finally, experimentation

with funding and output levels will increase our knowledge of service supply and cost functions.^[33–35]

The kind of information called for here requires a high level of analytical sophistication in budget execution and system design, a skill that staff responsible for executing budgets may lack. Indeed, even if controllers had good information on cost and service supply functions, some might not know how to use it. Their experience tends to orient them to the administration of the traditional line-item, object-of-expenditure budget. Effective administration of a lump-sum or line-item appropriation requires no more than a modicum of arithmetical ability combined with a substantial amount of horse sense and bargaining savvy. However, matching control systems design to circumstances requires a practical understanding of applied microeconomics and financial and managerial accounting. Controllers often fail to understand the ideas outlined here or how to implement alternatives to the line-item appropriations budget—where to exercise judgment and where to exercise specific decision rules. This is demonstrated by the persistent attempt of controllers to employ techniques devised for use within organizations, such as standard costs based on fully-distributed average historical costs, to establish per-unit prices for public organizations such as hospitals and universities.

OBSTACLES TO REFORM

Absence of knowledge of options and objectives is not a satisfactory explanation for controller decisions to resist reform. Such problems can be corrected, and incompetence may be weeded out. If a better match between control system design and circumstances would have a substantial payoff, why has this situation not been corrected? One answer regarding the implementation of reform is as follows.

A large part of the literature on budgeting in the United States is concerned with reform. The goals of the proposed reforms are couched in similar language—economy, efficiency, improvement, or just better budgeting. The President, the Congress and its committees, administrative agencies, even the citizenry are all to gain by some change. However, any effective change in budgetary relationships must necessarily alter the outcomes of the budgetary process. Otherwise, why bother? Far from being a neutral matter of “better budgeting,” proposed reforms inevitably contain important implications for the political system, that is, the “who gets what” of governmental decisions.^[36]

If controllers and elected officials empowered to determine the methods used in executing budgets are rational (we understand that politics often compromises rationality as we define it from a management

perspective), this quote implies that they have a strong interest in maintaining the status quo. To explain the persistent mismatch between budget execution control system designs and practice, it is necessary to determine who benefits from the status quo and, therefore, who will oppose the adoption of a more appropriate type of control.^[37] Members of Congress, state legislators, city council members, and any politician with a constituency worth cultivating would appear to lose as a result of reforms proposed. As the collective holders of the power of the purse, legislators clearly have the authority to order budgets to be executed in almost any way they like, including the power to delegate this authority to controllers. Efficiency implies an exclusive concern with the supply of goods and services to the citizenry, with some indifference as to the means used to supply the goods or even to the identity of the suppliers. However, legislators are frequently as concerned about where public money is spent and who gets it, as they are with what it buys.^[38–41] Line-item appropriations in general, and object-of-expenditure budgets in particular, are ideally suited to the satisfaction of legislative preferences with respect to how public money is spent, where it is spent, and who gets it.

REFERENCES

1. McCaffery, J.L.; Jones, L.R. *Budgeting and Financial Management in the Federal Government*; Information Age Press: Greenwich, CT, 2001.
2. Jones, L.R.; Thompson, F. *Public Management: Institutional Renewal for the Twenty-First Century*; JAI-Elsevier Science Press: Stamford, CT, 1999.
3. Simons, R. *Levers of Control*; Harvard Business School Press: Boston, MA, 1995.
4. Anthony, R.; Young, D. *Managerial Control in Non-Profit Organizations*; Irwin, 1984; 20, 21, 288; Chaps. 7–9.
5. Simon, H., et al. *Centralization vs. Decentralization in Organizing the Controller's Department*; The Controllership Foundation: New York, 1954.
6. Schick, A. Control patterns in state budget executions. *Public Adm. Rev.* **1964**, 24, 97–106.
7. Schick, A. Contemporary problems in financial control. In *Current Issues in Public Administration*, 2nd Ed.; Lane, F., Ed.; St. Martin's Press: New York, 1982; 361–371.
8. Demski, J.; Feltham, G. *Cost Determination*; Iowa State University Press: Ames, 1967.
9. Goldberg, V. Regulation and administered contracts. *Bell J. Econ.* **1976**, 7, 426–428.
10. Thompson, F. How to stay within the budget using per-unit prices. *J. Policy Anal. Manag.* **1984**, 4/1.
11. Thompson, F.; Fiske, G. One more solution to the problem of higher education finance. *Policy Anal.* **1978**, 3/4.
12. Bailey, M. Decentralization through internal prices. In *Defense Management*; Enke, S., Ed.; Prentice-Hall: Englewood Cliffs, NJ, 1967; 337–352.
13. Beckner, N.V. Government efficiency and the military: buyer–seller relationship. *J. Polit. Econ.* **1960**, 68.
14. Thompson, F.; Zumeta, W. Controls and controls: a reexamination of control patterns in budget execution. *Policy Sci.* **1981**, 13, 25–50.
15. Pitsvada, B.T. Flexibility in federal budget execution. *Public Budg. Finance* **1983**, 3/2.
16. Draper, F.; Pitsvada, T. Limitations in federal budget execution. *Gov. Account. J.* **1981**, 30, 3.
17. Fisher, L. *Presidential Spending Power*; Princeton University Press: Princeton, NJ, 1975.
18. Poole, R. Air traffic control: the private sector option. *Herit. Found. Backgrounds* **1982**, 216.
19. Bennett, J.; DiLorenzo, T. Public employee labor unions and the privatization of public services. *J. Labor Res.* **1983**, 4, 42, 43.
20. Blankart, C.B. Bureaucratic problems in public choice: why do public goods still remain public? In *Public Choice and Public Finance*; Roskamp, R., Ed.; Cufas: New York, 1979; 155–167.
21. Poole, R. Fighting fires for profit. *Reason* **1976**, May.
22. Smith, R.G. Feet to the fire. *Reason* **1983**, May, 23–29.
23. Hanke, S. The privatization debate. *Cato J.* **1982**, 656.
24. Hanke, S. Land policy. In *Agenda*; Howill, R., Ed.; Heritage Foundation: Washington, DC, 1983; Vol. 83, 65 pp.
25. Hanke, S. Privatization: theory, evidence, implementation. In *Control of Federal Spending*; Harris, R., Ed.; Academy of Political Science, 1985; 106–110.
26. Weicker, J. *Housing*; American Enterprise Institute: Washington, DC, 1980; 80 pp.
27. Poole, R. *Cutting Back City Hall*; University Books, 1980; 164.
28. Savas, E.S. Policy analysis for local government. *Policy Anal.* **1977**, 3, 49–77.
29. Bennett, J.; Johnson, M. Public v. private provision of collective goods and services. *Public Choice* **1979**, 4, 55–63.
30. Bennett, J.; Johnson, M. *Better Government at Half the Price: Private Production of Public Services*; Caroline House, 19XX.
31. Thompson, F.; Jones, L.R. *Reinventing the Pentagon*; Jossey-Bass: San Francisco, 1994.
32. Jones, L.R.; Bixler, G.C. *Mission Financing to Realign National Defense*; JAI Press: Greenwich, CT, 1992.
33. Wildavsky, A. *Budgeting: A Comparative Theory of the Budgetary Process*; Little Brown: Boston, 1975; 118–119.
34. Larkey, P. *Evaluating Public Programs: The Impact of General Revenue Sharing on Municipal Government*; Princeton University Press: Princeton, NJ, 1979.
35. Cothran, D. Program flexibility and budget growth. *West. Polit. Q.* **1981**, 34, 593–610.
36. Wildavsky, A. Political implications of budget reform. *Public Adm. Rev.* **1961**, 21, 183–190.
37. Zimmerman, J. The municipal accounting maze: an analysis of political incentives. *J. Account. Res.* **1977**, 21, 107–144.
38. Arnold, D. *Congress and the Bureaucracy*; Yale University Press: New Haven, CT, 1979.

39. Ferejohn, J. *Pork Barrel Politics*; Stanford University Press: Stanford, CA, 1974.
40. Fiorina, M. *Congress: Keystone of the Washington Establishment*; Yale University Press: New Haven, CT, 1977.
41. Shepsle, K.; Weingast, B. Political preferences for the pork barrel. *Am. J. Polit. Sci.* **1981**, 25.

BIBLIOGRAPHY

1. Barton, D. Regulating a monopolist with unknown costs. *Econometrica* **1982**, 50.
2. Breton, A.; Wintrobe, R. The equilibrium size of a budget maximizing bureau. *J. Polit. Econ.* **1975**, 83, 195–207.
3. Cheung, S.N.S. The contractual nature of the firm. *J. Law Econ.* **1983**, 25.
4. Coase, R. The nature of the firm. *Economica* **1937**, 4.
5. Fox, R. *Arming America: How the U.S. Buys Weapons*; Harvard University Press: Cambridge, MA, 1974.
6. Hofstad, G.H. *The Game of Budget Control*; Van Gorcum, 1967.
7. Holstrom, B. Moral hazard and observability. *Bell J. Econ.* **1979**, 10.
8. Meyerson, R.B. Incentives compatibility and the bargaining problem. *Econometrica* **1979**, 47.
9. Mirlees, J. The optimal structure of incentives and authority within an organization. *Bell J. Econ.* **1976**, 7.
10. Mitnick, B. The theory of agency: the policing “paradox” and regulatory behavior. *Public Choice* **1977**, 30.
11. Morgan, J. Bilateral monopoly and the competitive output. *Q. J. Econ.* **1949**, 63.
12. Peck, M.; Scherer, F. *The Weapons Acquisition Process: An Economic Analysis*; Harvard Business School: Cambridge, MA, 1962.
13. Scherer, F. *The Weapons Acquisition Process: Economic Incentives*; Harvard Business School: Cambridge, MA, 1964.
14. Stark, R. On cost analysis for engineered construction. In *Auctions, Bidding, and Contracting*; Englebrecht-Wiggins, R., Shubik, M., Stark, R., Eds.; New York University Press: New York, 1983.
15. Stark, R.; Varley, T. Bidding, estimating, and engineered construction contracting. In *Auctions, Bidding, and Contracting*; Englebrecht-Wiggins, R., Shubik, M., Stark, R., Eds.; New York University Press: New York, 1983; 121–135.
16. Thompson, F. Utility maximizing behavior in organized anarchies. *Public Choice* **1981**, 36.
17. Wildavsky, A.; Hammann, A. Comprehensive versus incremental budgeting in the department of agriculture. *Admin. Sci. Q.* **1956**, 10, 321–346.
18. Williamson, O. *The Economics of Discretionary Behavior*; Prentice Hall: Englewood Cliffs, NJ, 1964.
19. Williamson, O. *Markets and Hierarchies*; Free Press: New York, 1975.

Budget Reform

Jerry McCaffery

*Department of Systems Management, Naval Postgraduate School,
Monterey, California, U.S.A.*

INTRODUCTION

Budget reform involves restructuring the budget process to provide better information so that decision makers can make better resource allocation decisions. Reforms have focused on process improvements such as program budgeting, performance budgeting, and zero-based budgeting, and desired outcomes such as balanced budgets and mechanisms to produce them. After an examination of the meaning of budget reform and its historical context, this article discusses how the concepts of planning, management, and control form core elements of all budget systems, but are emphasized differently in different budget systems and reforms.

MEANING OF BUDGET REFORM

Budget reform has been hampered because government budgeting itself is not simple. In sending his proposal to create an executive budget system to Congress in 1912, President Taft said, “The Constitutional purpose of a budget is to make government responsive to public opinion and responsible for its acts” (Ref.^[1], p. 19). The proposal further noted that a budget served numerous purposes, from a document for Congressional action, to an instrument of control and management by the President, to a basis for the administration of departments and agencies. Since then, the multiple purposes for budgeting have been well recognized.^[2] As long as a multiform definition of budget is acceptable, defining budget is relatively easy. Considering budget reform is much more difficult. What is budget reform? Budget reform involves reshaping the budget and the budget process so that money is spent more efficiently, effectively, and wisely. Ultimately, budget reform might result in the more efficient application of public expenditures to public needs. In 1940, V.O. Key^[3,4] suggested that the critical question for budget decision makers was how to supply the answer to the fundamental question, “On what basis shall it be decided to allocate x dollars to activity A instead of activity B?” Key argued that this was fundamentally a political question, and how one answered that question was dependent on the one’s values.^[5] If the answer to Key’s question is no clearer

now than it was in the 1940s, it is also clear that budget practitioners before and after Key believed that improving the elements of the decision process was useful (For example, some researchers have found that budget type influences budget discussion, i.e. program budgets provoke more programmatic discussion. See Refs.^[6,7]).

EARLY AMERICAN PATTERNS: PREHISTORY

Budget reform has a rather extensive history. Early American budgetary patterns were both part of and separate from their predominantly English colonial heritage. They were part of that heritage in that the American colonies inherited English historical experience with gradual imposition of limits on the power of the monarch and the gradual expansion of legislative power. They were separate from this legacy in that the American revolutionaries carefully tried to avoid anything that smacked of giving power to an executive; having rid themselves of one king, they had no taste for another. In the turmoil during and after the American Revolution, two giants emerged on stage almost simultaneously. The reforms of William Pitt the Younger reshaped British budgetary practices (Wildavsky calls Pitt a pivotal figure in British budgetary history. See Ref.^[8]), while in the new United States Alexander Hamilton almost single-handedly created a budgetary system and practice for the new American nation.

As Chancellor of the English Exchequer from 1783 to 1801, Pitt faced a heavy debt load as a result of the American Revolution. Pitt consolidated a maze of customs and excise duties into one general fund from which all creditors would be paid, reduced fraud in revenue collection by introducing new auditing measures, and instituted double-entry book-keeping procedures. Moreover, Pitt established a sinking fund schedule for amortization of debt, requiring that all new loans made by government impose an additional 1% levy as a term of repayment. Pitt raised some taxes and lowered others to reduce the allure of smuggling. This model encompassed a royal executive with varying degrees of strength and a legislative body attempting to exert financial control over the Crown by requiring parliamentary approval of sources of revenue and expenditure. Approval was provided through

appropriations legislation and ultimately made administrative officials accountable to Parliament.

Hamilton's problems were different from Pitt's, but equally serious. After the Revolutionary War, the colonies departed from English tradition when they created fiscal systems that virtually excluded the executive, vesting power in various legislative arrangements. As a result of their distaste for central government, the powers of the first Congress established under the Articles of Confederation were weak. Their fear of creating a new monarch was evident in the manner in which powers were delegated to both the legislative and executive branches. The colonists were also averse to a system of national taxation. Taxation imposed a special hardship on the colonies because hard coinage was scarce and bills or letters of credit were used irregularly.

In 1789, as the first Secretary of Treasury, Alexander Hamilton strongly influenced development in this period (For an excellent biography of Hamilton, see Ref.^[9]). Initially, Hamilton set to establish the credit of the new government. His first two reports to Congress urged funding the national debt at full value, the assumption by the federal government of all debts incurred by the states during the Revolutionary War, and a system of taxation to pay for the debts.^a Strong opposition arose to these proposals, but Hamilton's position prevailed after he made a bargain with Thomas Jefferson, who delivered southern votes in return for Hamilton's support for locating the future nation's capital on the banks of the Potomac. Hamilton's third report to Congress proposed a national bank, modeled after the Bank of England. In a few short years, Hamilton submitted four major reports to Congress, gaining acceptance of three that funded the national debt at full value, established the nation's credit at home and abroad by creating a banking system and a stable currency, and developed a stable tax system based on excise taxes to fund steady recovery from the debt and to provide for future appropriations. Hamilton's role in establishing a system for debt management, securing the currency, and providing a stable revenue base make him the founding father of critical elements of the American budgeting system.

FIRST EFFORTS AT REFORM

As early as 1800, civilian agency budgets were presented in carefully detailed object-of-expenditure

form, whereas military expenditures tended to be appropriated as lump sums not unlike specific program categories. Early debate on budget development focused on enhancing flexibility and program accomplishments rather than on strict agency accountability in the arguments over the merits of lump sum versus line item appropriations, over the merits of allowing for executive discretion in spending money to execute programs as opposed to requiring that agencies follow explicit legislative directions constrained by line item appropriations. However, for most of the nineteenth century, government functions were few and simple, and were supported by equally simple budget structures and procedures. Tyler and Willand noted that by 1900, the United States was the "only great nation without a budget system" (Ref.^[10], quoting Ref.^[11]).

Debate about form and content of the federal budget accelerated as a result of local government reforms implemented in New York City and elsewhere from the mid-1870s through the beginning of the twentieth century, and culminated in Congressional enactment of the Budget and Accounting Act of 1921. The stimuli for passage of this act may be found in the deficits incurred in World War I and the various reform movements rising in local governments as a reaction to egregious examples of boss rule and political corruption in the late 1870s and after. It also had a national background. Irene Rubin^[12] suggested that some part of federal reform came from efforts to oversee the railroads at the end of the nineteenth century. Railroads were big and important to the public for carrying goods and passengers. With the expansion of railroads and the acquisition of one railroad by another, it was not always easy to tell if a railroad was making a profit or, indeed, if it were solvent. This meant it was also difficult to tell if the rates that were set for freight and passengers were fair. Thus, in the 1880s, the Interstate Commerce Commission was put in the position of regulating rates, but to do this it had to develop better accounting tools in consultation with the railroads. Once these procedures had been implemented for railroads—considered to be the largest, most complex, and most sophisticated of private corporations—government was then urged to make similar improvements for its own operations. Thus, reformers had rather clear ideas about what they were against and what they were for. Nevertheless, Rubin suggested that business leadership in budget reform has been overstated, and that it was government officials and academics who invented and modified public budgeting in the United States, while business methods have historically been given more credit than they deserve (Ref.^[12], p. 444).

In the 1900s, reform efforts were continued by Frederick Cleveland at the New York Bureau of

^aHamilton's program was outlined in four reports: Reports on the Public Credit of January 14, 1790 and December 13, 1790; The Report on a National Bank, December 14, 1790 and The Report on Manufactures, submitted to Congress on December 5, 1791.

Municipal Research and by President Taft's Commission on Economy and Efficiency in 1912. Long before the federal government passed the Budget and Accounting Act of 1921, many American cities had passed local budget reform acts that prevented city councils from appropriating money outside the confines of a budget. As early as 1899, the National Municipal League drafted a model municipal charter that incorporated a budget system under the direct supervision of the mayor (Ref.^[1], p. 13). Passage of the 1921 Act was an important point in the development of the executive budget power, giving the president the power to organize the executive budget process and to create a staff agency to compile, analyze, and make recommendations to Congress about agency budgets. The power of the executive over the budget process would grow steadily for the next 50 years as the functions of the federal government expanded under the impress of the Great Depression and Presidents Hoover and Roosevelt in the 1930s, World War II, and the social service revolution of the 1960s. These trends resulted in dramatic changes in the nature of the federal budget. From 1950 to 1990, federal government spending shifted from favoring discretionary expenditures with more than one-half of the annual budget spent on defense, to mandatory expenditures with more than one-half being spent on social security, health, and welfare programs. This made tools and processes related to discretionary expenditure less important than those aimed at analyzing, predicting, and controlling mandatory expenditures (see Refs.^[13],^[14] and ^[15]), and ensured that the search for better budget processes and tools would continue.

In the United States, twentieth-century budget reform spread from the local level to state and national levels. Although the influence of state and local reforms on national patterns has varied, state and local governments have remained a vital force for reform. Smaller and with a less complicated mechanism than the federal government, subnational governments faced less inertia and were more flexible in adopting reform elements and discarding them for something better. States and certain large cities must be seen as first adopters; the wise student of innovation will always find something of interest occurring there. Recently, states have been involved with target-based budgeting, performance budgeting, and performance measurement and accountability (see Refs.^[16-19]). Federal adoption of a particular state reform may be seen as mainstream acceptance of that reform; state adoption of a federal pattern often comes as an attempt to increase the synchronization between state and federal patterns in order for the state to extract more money from the federal budget.

Evaluation of the history of reform has been mixed. Irene Rubin argued that budget reforms are much

more widely adopted than is generally recognized. The benefits tend to be oversold and when exaggerated outcomes do not seem to appear, many participants lose interest and deem the reform a failure, regardless of whether it leaves real improvements in the budget process.^[20] Forrester and Adams cited the paucity of successful reforms and warned that budget reform requires attention to the organizational context, and that the objectives and execution of a reform must match the dynamic objectives and needs of the organization and be cognizant of the organizational culture.^[21] Notwithstanding that reform is neither linear nor certain, the last 100 years indicate reform efforts have a certain irrepressibility.

MODERN REFORMS OF PROCESS AND OUTCOME

Reforms have involved both process and outcome. Much of the history of budget reform involved finding ways to present budget information to decision makers in a more meaningful way so that better decisions can be made about how to allocate scarce resources. These reforms have included performance budgeting, program budgeting, zero-based budgeting, Planning-programming-budgeting (PPB) and various systems focused on target- or mission-based budgeting. Some reforms have focused on the budget process itself, in the belief that better staff or a more timely process would provide a better budget process. Paramount among these sorts of reforms are the Budget and Accounting Act of 1921, and the Congressional Budget Reform and Impoundment Act of 1974.

Dissatisfaction with budget outcomes, usually in respect to growing deficits, has resulted in a different type of budget reform basically attempting to mandate outcomes (e.g., deficit reduction or balanced budgets). Examples of this kind of budget reform at the federal level include in the Gramm-Rudman-Hollings (GRH) Acts of 1985 and 1987, and the Budget Enforcement Acts of 1990 and 1993, with their spending caps and pay-as-you-go requirements, and several attempts at a Constitutional Balanced Budget Amendment. These efforts were not without some adverse consequences. Attempts to meet the GRH targets often involved optimistic estimates of spending and revenues, and certain gimmicks such as shifted paydays, which probably did more to damage the budget process than reform it, especially because few if any of the GRH targets were met. With the Budget Enforcement Act of 1990, the United States shifted from a deficit control approach to a spending cap approach. This was a looser discipline, but it was used throughout the 1990s, being reenacted in the Balanced Budget Agreement of 1997 for the 5-year period out to 2002. For

discretionary accounts this meant tight discipline, often at or below a growth rate equal to that of inflation. Both Congress and the President seemed to move toward supplemental appropriations and emergency spending designations to escape some of this discipline.

In terms of process, the spending cap focus appeared to put more emphasis on the beginning of the budget process, and it is arguable that the 1990s saw the full fruition of the Congressional Budget Reform and Impoundment Control Act of 1974. The budget committees and the budget resolution became important forces in the budget process because they were used by both parties, but especially the Republicans, to seek new notions of the shape and size of the federal government. This new role was to be a diminished one, with power and programs being pushed toward more innovative and effective state and local governments, or even toward the private sector. New national security concerns occurring with the terrorist attacks of 2001 may lead to a reevaluation of this position. Providing for the common defense and responding to a great crisis has always been the role of the central government.

PLANNING, MANAGEMENT, AND CONTROL

In 1964, Allen Schick offered an analytic framework for budget reform that has withstood the test of time; its insights remain as penetrating and relevant now as they were then,^[22,23] and its explanations just as powerful. Schick said that every budget system comprised elements of planning, management, and control, and that different budget systems arranged and emphasized those elements differently. Planning involved the determination of objectives, the scrutiny of alternative courses of actions, and the selection of programs. Management involved programming of approved goals into specific goals and activities, designing programs to carry out those activities and providing staffing and resources to carry out those designs. Control involved the process of “binding” operating officials to the policies and plans set by their superiors, through the appropriation and budget execution process, and through budget control procedures such as position controls, limits on staffing, and transfer authority, purchase requisition, procurement, and travel procedures, capital outlay procedures, and pre- and postaudit processes.

Schick observed that planning, management, and control were not equally emphasized in budget systems because time was a scarce commodity in the budget process and participants had to choose which of these elements to maximize. In addition, Schick concluded that different budget systems inherently emphasize one of these elements over another. Consequently, the observer can examine different budget systems

for different characteristics, but these characteristics cluster around concepts of control, management, and planning. For example, control-oriented budget systems focus on line item budget systems, value the skills of accountants, focus on inputs (e.g., dollars, personnel), and are focused most heavily on budget execution. In contrast, management-oriented budget systems focus on the preparation stage of budgeting, are dominated by managers rather than accountants, focus on the units of work to be done (e.g., cases to be cleared, miles of highway to be built), and seek efficiency and effectiveness, rather than legally correct expenditure patterns. Budget systems that focus on planning were the most broad-gauge and focus on the prepreparation stage of budgeting (e.g., What is good policy? What should government do to bring about good policy?), emphasize benefit–cost analysis and the skills of economists and systems analysis, rather than managers or accountants, to work out streams of future benefits and costs. A planning-focused system in its insistence on supporting policy with dollars was the most generalized and difficult system to operate in Schick’s paradigm.

As a practical matter, Schick was concerned with explaining the differences between the early version of the PPB system then in use in the Department of Defense and the powerful analysis of non-defense spending offered by Aaron Wildavsky under the rubric of incrementalism, where the future is controlled by the past and departs modestly from it as budgets change by increments from year to year. Under PPB, Schick suggested that change would occur from the top down, rather from the bottom up. PPB systems would implement change as a result of comprehensive policy-making procedures, rather than an aggregation of individual decisions made year by year. PPB would be rational. Incrementalism would be experiential, and rely heavily on feedback and course correction procedures as a counterbalance to the lack of rationality and foresight. Although incremental budgeting would result in information and budgets building up from the bottom, PPB systems would result in a downward flow of budget decisions resulting from choices about policy.

After a brief governmentwide trial in the 1960s PPB was again limited to defense, not because defense is well suited to certain analytical skills or benefit–cost analysis logic, but because the existence of a threat drives everything. In defense, the critical budgetary questions each year are “what is the threat?”, “how has it changed from last year?”, and “what will we need to meet this threat 5 years from now.” PPB allows for these questions to be answered centrally in a unifying scenario and for subsequent decisions to flow down through the organization in response to that scenario. In most years, the defense budget is an

incremental system, in the same sense that federal non-defense budgets are incremental. However, occasionally, the threat situation changes dramatically. When that happens, the planning and budgeting system must reflect those changes. After the fall of the Berlin Wall and the collapse of the USSR, the threat situation for the United States changed dramatically. It was obvious that U.S. defense spending could now be lower, but efforts to articulate a new threat situation and a response to it took almost a decade and were not really resolved by the time of the terrorist attack in September 2001, although the defense establishment was downsized by about one-third during that period. Nonetheless, PPB remains alive and well in the Department of Defense because the defense budget must attempt to ascertain long-term threats, as well as to perfect the mix of current and future assets that will meet that threat. PPB is especially well suited to plan to meet threats to the United States by designing force structure (e.g., number of troops, tanks) and managing a stream of resources to fund each year what is necessary to meet the threat this year and in the future. In most years, PPB outcomes appear and may be incremental, reflecting an incremental change in the threat situation. No reason exists why PPB could not be used in the health or aging areas, except that it is an expensive system to use and it does tend to surface the total cost of packages that political decision makers would perhaps prefer to ignore until they need to be resolved in a climate of emergency.

Philip Joyce suggested that an understanding of the history of budget reform can be organized around Schick's concepts; thus, he divides U.S. budget history into four periods.^[24,25] The early period (1873–1937) is characterized by weak executive power, little central control, and idiosyncratic processes. Reformers sought to emphasize central control and to bring sound business practices to government, while focusing on increasing economy and efficiency. Joyce asserted that perhaps the single greatest outcome of this period was the Budget and Accounting Act of 1921, with its creation of the Bureau of the Budget and the establishment of the General Accounting Office to audit the results of federal spending and report to Congress. Joyce suggested the second period lasted from 1937 to 1960; this period emphasized management, thanks to the various challenges faced by the federal government with the Great Depression of the 1930s and World War II and its aftermath. Early efforts at performance budgeting were a key element of this period.

Joyce suggested that 1960 to 1985 was the third period of budget reform. Here, the emphasis was on planning for outcomes with program budgeting and PPB systems. During this period, reforms sought to move beyond efficiency and management to comprehensive program decisions based on notions of comprehensive

and rational decision making. Zero-based budgeting, another variant of comprehensive budgeting, also enjoyed a brief popularity. These systems failed to displace incremental patterns based on line item budgeting,^[26,27] although they did leave a legacy biased toward more analysis in the budget process. In 1974, the Congressional Budget Reform Act gave Congress its equivalent to the Office of Management and Budget in the Congressional Budget Office, and provided a highly capable source for budget estimation and appropriation scorekeeping. The budget resolution and the budget committees also helped Congress assume some comprehensive planning functions for the budget at the start of its budget process. Joyce suggested that these comprehensive reforms were only partially successful, but that they did leave a legacy favoring greater analysis, and zero-based budgeting, with its concentration on levels of effort around the budget base, has led to target-based budget systems at state and local levels. Joyce noted that the period after 1985 was mostly occupied with deficit reduction measures with GRH I and II and the various Budget Enforcement or Balanced Budget Acts, as well as a resurgence of performance-based budgeting focused on evaluating the results of government spending with the Government Performance and Results Act of 1993. Rubin^[28] argued that two additional emphases should be added to Schick's framework to represent trends from 1970 through the 1990s—prioritization and accountability. This is a useful clarification, but if prioritization is understood to be a primary emphasis in planning and accountability in control, then Schick's framework may stand.

OBSERVATIONS

Schick's powerful analytical construct helps us to think about current budget reform efforts. Clearly, the Chief Financial Officers Act of 1990 is a control mechanism. It is focused on budget execution, fiduciary concerns, and correct and appropriate spending of budgetary funds. With its emphasis on auditable financial statements for the federal government and individual agencies, it is concerned that dollars of input be spent for correct and appropriate purposes; that fraud, waste, and misuse funds be avoided; and that the public be able to trust their government to faithfully administer budgets within the parameters in which they have been passed. The Government Performance and Results Act of 1993, however, clearly emphasizes the management aspects of budgeting, with its emphasis on outputs and outcomes, and their costs. It pushes beyond the propriety of control aspects to efficiency and effectiveness, and even beyond with an emphasis on outcomes—for example, measurably more justice or measurably

less poverty. With linkage to strategic planning efforts, it also moves toward a planning orientation and one that has rather widespread input to it.

Schick's analysis is also a caution about the difficulty of budget reform. For example, reforms that emphasize one element usually impair another; increased control usually means decreased management. Thus, balancing these concepts is important. Finally, while no magic formula for a good budget process has been discovered, despite strenuous efforts, modern American budgeting at the federal level is remarkably accessible to its citizens. In a democracy, much should be made of this. Moreover, budget reformers remain remarkably persistent about ensuring that government should act as a wise fiduciary agent, that decisions reflect attempts to get the most value from the dollars spent, and that government target important and achievable ends. There is much to praise in these efforts.

REFERENCES

- Burkhead, J. *Government Budgeting*; John Wiley & Sons: New York, 1956.
- Wildavsky, A. *The Politics of the Budgetary Process*; Little, Brown: Boston, 1964. Revised editions in 1974, 1979 and 1984. Also, Caiden, N. and Wildavsky, A. *The New Politics of the Budgetary Process*. Little, Brown: Boston, 1988 and 1992.
- Key, V.O., Jr. The lack of budgetary theory. *Am. Polit. Sci. Rev.* **1940**, 34, 1137–1144.
- Wildavsky, A. Political implications of budgetary reform: a retrospective. *Public Adm. Rev.* **1992**, 52 (6), 594–603.
- Jones, L.R. Changing how we budget: Aaron Wildavsky's perspective. *J. Pub. Budg. Account. Financ. Manag.* **1997**, 9 (1), 46–71.
- Grizzle, G. Does budget format govern actions of budgetmakers? *Public Budg. Finance* **1986**, 6, 60–70.
- Pettijohn, C.; Grizzle, G. Structural budget reform: does it affect budget deliberations? *J. Pub. Budg. Account. Financ. Manag.* **1997**, 9 (1), 26–45.
- Wildavsky, A. *A Comparative Theory of Budgetary Processes*; Little, Brown: Boston, 1975; 272. The best recognized biography on Pitt is J. Holland Rose, William Pitt and National Revival, London: G. Bell and Sons, 1911. See also J. H. Rose, William Pitt and the Great War, London: G. Bell and Sons, 1911; and J. H. Rose, Pitt and Napoleon, London: G. Bell and Sons, 1912.
- Miller, J.C.; Hamilton, A. *Portrait in Paradox*; Harper: New York, 1959.
- Tyler, C.; Willand, J. Public budgeting in America: a twentieth century retrospective. *J. Pub. Budg. Account. Financ. Manag.* **1997**, 9 (2), 182–219.
- Buck, A.E. *Public Budgeting: A Discussion of Budgetary Practice in the National, State and Local Governments of the United States*; Harper: New York, 1929; 10 pp.
- Rubin, I. Who invented budgeting in the United States?. *Public Administration Review*. September/October **1993**, 53 (5), 438–444.
- Schick, A. *The Capacity to Budget*; Urban Institute Press: Washington, DC, 1990; 86 pp.
- White, J. Entitlement budgeting vs. bureau budgeting. *Public Adm. Rev.* **1998**, 58 (6), 113–134.
- Schick, A. From the old politics of budgeting to the new. *Public Budg. Finance* **1994**, 14 (1), 135–144.
- Douglas, J.W. Redirection in Georgia: a new type of budget reform. *Am. Rev. Public Adm.* **1999**, 29 (3), 269–289.
- Lee, R.L., Jr.; Burns, R.C. Performance measurement in state budgeting: advancement and backsliding from 1990 to 1995. *Public Budg. Finance* **2000**, 20 (3), 38–54.
- Willoughby, K.G.; Melkers, J.E. Implementing PBB: conflicting views of success. *Public Budg. Finance* **2000**, 20 (1), 105–120.
- Melkers, J.E. The state of the states: performance-based budgeting requirements in 47 out of 50. *Public Adm. Rev.* **1998**, 58 (1), 66–74.
- Rubin, I. Budget theory and budget practice: how good the fit? *Public Adm. Rev.* **1990**, 50, 179–189. See also Ref.[10].
- Forrester, J.; Adams, G. Budgetary reform through organizational learning: toward an organizational theory of budgeting. *Adm. Soc.* **1997**, 28 (4), 466–488.
- Schick, A. The road to PBB: the stages of budget reform. *Public Adm. Rev.* **1966**, 26 (4), 243–258.
- Schick, A. The road from ZBB. *Public Adm. Rev.* **1978**, 38 (2), 177–180.
- Joyce, P. Budget reform. In *International Encyclopedia of Public Policy and Administration*; Shafritz, J., Ed.; Westview Press: Boulder, CO, 1998; 276–281.
- Grafton, C.; Permaloff, A. Budgeting reform in perspective. In *Handbook on Public Budgeting and Financial Management*; Rabin, J., Lynch, T., Eds.; Marcel Dekker: New York, 1983; 89–124.
- Wildavsky, A. A budget for all seasons? Why the traditional budget lasts. *Public Adm. Rev.* **1978**, 38 (6), 501–509.
- White, J. (Almost) Nothing new under the sun: why the work of budgeting remains incremental. *Public Budg. Finance* **1994**, 14 (1), 113–134.
- Rubin, I. Budgeting for accountability: municipal budgeting in the 1990's. *Public Budg. Finance* **1996**, 16 (3), 112–132. Summer.

Budget Stabilization Fund

Yilin Hou

*Department of Public Administration and Policy, University of Georgia,
Athens, Georgia, U.S.A.*

INTRODUCTION

The budget stabilization fund (BSF), popularly called the “rainy-day fund” (RDF), is a fiscal device used by subnational governments to store extra revenues during economic booms for use in economic downturns to supplement inadequate resources for meeting outlay demands. Since the early 1980s, this countercyclical device has attracted increasingly more academic attention.

ORIGIN

The concept and practice of countercyclical reserves date back to Biblical times: Joseph saved Egypt from a great famine by storing up food in years of harvest.^[1] In the American government system, the BSF also enjoys a relatively long history, tracing back to the second half of the 19th century.^[2] With ineffective controls and too frequent transfers, many cities then had difficulty keeping departmental spending within budgeted levels; so the cities created contingency funds for emergencies.^[3] Early contingency funds were very small as a percentage of the budget or in absolute amounts^[3] (for example, Boston reserved only a quarter to half a percent of appropriations for emergencies and Cincinnati reserved \$50,000 for each half-year period). Rules regarding the use of the funds were developed involving consensus between the executive and legislative branches^[3] (in Boston, the city auditor could transfer from this fund only when he was directed to do so by the mayor, with approval of the council’s committee on finance. Council approval often had to be majority. In Cincinnati, mayoral direction had to be accompanied by consent of two-thirds of the council, while Minneapolis demanded supermajority approval—20 out of 26 council votes).

The earliest use of the name “rainy-day fund” was by New York City in the 1930s. The NYC rainy-day fund had inadequate restrictions on approval procedures and an insufficient mandate on replenishment; its balance was very “modest” by the early 1960s. “When NYC borrowed for operating expenses, it also decided to make use of its rainy day fund”^[4] (this is a comment by Mr. Ted Weiss who, in the early 1980s,

was chairman of the House Subcommittee on Intergovernmental Relations and Human Resources. Mr. Weiss was on the New York City Council in the early 1960s). The fund soon went into disuse and was forgotten. The lesson is apparent: a BSF cannot exist long without restrictive approval procedures. A consensus mechanism between the executive and legislative branches is a necessary feature.

At the state level, New York was the first to adopt a formal BSF (1946)—the tax stabilization reserve fund—“as a safeguard against possible future declines in revenues.”^[5] It should be noted that states give their BSFs different names. Ohio first used the name “budget stabilization fund” (1981); Michigan adopted a more descriptive name—“countercyclical budget and economic stabilization fund” (1977).

EVOLUTION

From Contingency Funds to General Fund Balances

Early contingency funds were created out of operational necessity, but maintaining the funds, especially when the financial condition was strong, often encountered intense political pressure. Clow^[2] mentioned that although they saw the strong management and operational rationale behind the fund, most finance officers involved in setting up the reserves were unwilling to stand the related outside pressures. They even preferred not to have a reserve fund.^[2]

As a consequence, finance officials often resorted to “unofficial” ways to end up with general fund balances (GFB)—money that they could use for contingencies. One way was to intentionally underestimate revenue or overestimate expenditures. Another way was to designate GFB for some purpose they knew would not occur or on which the entire designated amount would not be spent. This represents a tacit transformation of a contingency fund into GFB through management gimmicks.^[2] [Clow records that such dilemma had led to the purposeful complication of the contingency account by financial officers (e.g., treasurers), so that politicians could not easily make clear how much balances were available].

Countercyclical Role of General Fund Balances

The development of GFB as a countercyclical fiscal tool is closely related to balanced budget requirements (BBR) in the states. The requirements, in one form or another, are all in stock instead of in flow by nature.^[6–8] Thus instead of rigidly requiring each state to balance their budgets in each fiscal year (or each budget cycle for those biennial budget states), the BBRs allow states to balance over the years, creating the need and chance for the establishment of countercyclical reserves. A good illustration is that while most states require their governor to submit and sign and their legislature to pass a balanced budget, allowing deficit carryover into the next budget cycle is a common article in state BBRs. Gradually, states developed the practice of building up their GFB during boom years to help mitigate revenue shortages in downturns. Historical data confirm this proposition. From 1957 to 1984, over the cycles of six national recessions, GFB of the state and local sector in aggregation grew substantially during boom years (Table 1, Panel B) and the balances depleted very quickly in lean years (Table 1, Panel A). Professional organizations such as the National Association of State Budget Officers (NASBO) and Government Finance Officers Association (GFOA) have also advocated maintaining a certain amount of GFB as one of the “best practices” (for details, see Refs.^[9,10]).

General Fund Balances to Budget Stabilization Fund

General fund balances, however, became easy targets of attack by voters in more recent times, as in the case of Proposition 13 in California. Taxpayers in many states through referendum voted into effect tax and expenditure limitations to restrain state and local governments from keeping large GFB. The vulnerability of GFB made officials increasingly aware of the advantages of a properly structured countercyclical fiscal tool as a more reliable and easier-to-operate vehicle that is protected by law from spending pressure.^[4,11,12] This device is the budget stabilization fund.

DEFINING BUDGET STABILIZATION FUNDS

The fact that the BSF is a useful instrument to save assets during booms for use during recessions has been recognized by state as well as federal legislators,^[4,13,14] professional organizations^[10] rating agencies,^[15] and scholars.^[12] While there is a consensus regarding the utility of BSF, there does not seem to be consensus

yet on what a BSF is or what key structural features a BSF must possess to qualify as a real BSF. GFOA and NASBO, for example, list different features as essential for a BSF. The lack of consensus has led to confusion and misuse of the concept, which in turn causes inaccuracy in data collection.

Synthesizing past literature and related legislative history and examining the finance laws of the states identify three overarching features of the BSF. First, a BSF must have an enabling legislation as its legal basis. Second, a BSF is a countercyclical reserve fund across fiscal years or budget cycles. Finally, a BSF must be a government-wide reserve for general purposes.

Enabling Legislation

Budget stabilization funds are established with enabling legislation which describes the details of a BSF, such as its purpose(s), funding source(s), allowable balance level, procedure for use approval, replenishment after use, and so on. This legislation is the legal basis of the fund. It is very significant in several fundamental ways. First, the legislation serves as a firewall between the fund and elected officials. Executive officials cannot easily use the money at their discretion because the procedure for use approval is fairly strict in most cases. Legislators cannot readily engage in pork barrel spending with this fund either because the money is available only for predetermined purposes.

Second, the legal language creating BSFs provides a defense for elected officials against popular pressure for tax returns/refunds or overspending during boom years, especially in the era of tax revolts and expenditure limitation movements. Third, the law makes it compulsory to save when the economy is strong and state revenue goes above the expenditure needs and to replenish the fund after use, so the state can stay better prepared for revenue shortfalls.

Finally, the existence of a legal basis for a BSF serves as a dividing line between a countercyclical fiscal reserve fund and GFB, which is maintained more out of practice than as a legal requirement in response to daily operational needs. The fact that a few states do have a legally required fiscal year-end balance of the general fund does not diminish this distinction: such required balances do not go across fiscal year/budget cycles. Findings from a previous study indicate that when downturns hit, general fund balances are always the first to be depleted, followed by budget stabilization funds and other useable funds.^[12] When these resources are almost gone, tax increases, expenditure cuts, and accounting gimmicks come into use.

Table 1 State-local fiscal behavior, 1957–1984—average quarterly growth rates of aggregate general fund balances

Panel A: Behavior during recessions				
Contraction				
Peak	Trough	Revenue growth (%)	Expenditure growth (%)	Surplus change (bil \$)
1957 III	1958 I	1.7	2.9	−0.55
1960 I	1960 IV	1.9	2.1	−0.1
1969 III	1970 IV	2.8	3.2	−0.46
1973 IV	1975 I	2.6	3.3	−1.32
1980 I	1980 II	0.5	1.9	−4.7
1981 III	1982 IV	1.3	1.8	−1.66
Panel B: Behavior during expansions				
Expansion				
Trough	Peak	Revenue growth (%)	Expenditure growth (%)	Surplus change (bil \$)
1958 I	1960 I	2.4	1.5	0.34
1960 IV	1969 III	2.5	2.4	0.08
1970 IV	1973 IV	2.9	2.5	0.8
1975 I	1980 I	2.5	2.1	6.15
1980 II	1981 III	1.8	1.3	3.06
1982 IV	1984 IV	2.1	1.8	4.95

Sources: 1957–1977 data, Advisory Commission on Intergovernmental Relations (ACIR), *State-Local Finances in Recession and Inflation, Report-70*. Washington, DC, 1979; 1978–1984 data, Ronald Fisher, “Statement before the Intergovernmental Relations and Human Resources Subcommittee of the Committee on Government Operations,” in U.S. Congress, *Federal and State Roles in Economic Stabilization—hearings before a subcommittee of the Committee on Government Operations House of Representatives*, Nov. and Dec. 1984. U.S. Government Printing Office, 1985; 107.

Countercyclical Reserves Fund Across Fiscal Years/Budget Cycles

The BSF is designed as a countercyclical mechanism. As a reserve to be accumulated in boom years for use in lean years, it is necessarily a mechanism that operates across fiscal years (for those states that adopt biennial budgets, it operates across biennial budget cycles). The focal point of a BSF is to “balance,” or smooth out the peaks and troughs of, revenues and expenditures through the business cycle, instead of merely the fiscal year or budget cycle.

This feature separates the BSF from working capital funds and legally required year-end general fund balances. A working capital fund serving cash flow purposes operates mainly within the fiscal year so that the government does not have to resort to the debt market for current spending before tax revenues stream in. In some cases, the working capital fund does go beyond the fiscal year, but even in those instances, its focus remains on filling in the revenue gap between two budget cycles.

In a few states (Colorado, Kansas, New Mexico, and Wisconsin), their finance laws require a balance

(surplus) of the general fund at the end of fiscal years, ranging in size between 2% and 5% of the current year general fund revenue. Given the spending pressure from politicians and taxpayers, these required balances are not guaranteed to accumulate across fiscal years or budget cycles. As such, these required year-end balances cannot serve as a BSF.

Government-Wide Reserve for General Purposes

The U.S. General Accounting Office (GAO) classifies state reserves into three categories: 1) government-wide reserves for general purposes; 2) government-wide reserves for specific purposes; and 3) agency-specific reserves for specific purposes.^[16] A BSF should be a government-wide reserve for general purposes. Any reserves for single or special purposes are not in this category. Reserve funds that are set up for use by only certain agencies in the government are not in this category either. Contingency funds as currently in use by many state governments are across-fiscal-year reserves, and they can be government-wide

Table 2 Recessions and creation of BSF

Recession	State	Creation	Fund name	First balance
1945	New York	1946	Tax Stabilization Reserve Fund	1946
1957–1958	Florida	1959/65	Working Capital Fund/Budget Stabilization Fund	1965
1969–1970	Tennessee	1972	Reserve for Revenue Fluctuations	1972
1973–1975	California	1976	Special Fund for Economic Uncertainties	1977
	Georgia	1976	Revenue Shortfall Reserve	1976
	Delaware	1977	Budget Reserve Account	1979
	Michigan	1977	Countercyclical Budget and Economic Stabilization Fund	1978
	South Carolina	1978	General Reserve Fund	1978
	New Mexico	1978	General Fund Tax Stabilization Reserve	No balance
	Connecticut	1979	Budget Reserve Fund	1981
1980	Ohio	1981	Budget Stabilization Fund	1985
	Washington	1981	Emergency Reserve Fund	1989
1982	Indiana	1982	Countercyclical Revenue and Economic Stabilization Fund	1985
	Mississippi	1982	Working Cash-Stabilization Reserve Fund	1983
	Wyoming	1982	Budget Reserve Account	1983
	Kentucky	1983	Budget Reserve Trust Fund Account	1987
	Nebraska	1983	Cash Reserve Fund	1984
	Idaho	1984	Budget Stabilization Fund	1984
	Minnesota	1984/1996	Cash Flow Account/Budget Reserve Account	1984
	Maine	1985	Maine Rainy Day Fund	1985
	Maryland	1985	Revenue Stabilization Fund	1987
	Massachusetts	1985	Commonwealth Stabilization Fund	1987
	Oklahoma	1985	Constitutional Reserve Fund	1988
	Pennsylvania	1985	Tax Stabilization Reserve Fund	1986
	Rhode Island	1985	Budget Reserve and Cash Stabilization Account	1985
	Wisconsin	1985	Budget Stabilization Fund	No balance
	Utah	1986	Budget Reserve Account	1987
	New Hampshire	1987	Revenue Stabilization Reserve Account	1987
	North Dakota	1987	Budget Stabilization Fund	1990
	Vermont	1987	General Fund Budget Stabilization Reserve	1987
	Texas	1988	Economic Stabilization Fund	1990
	Alaska	1990	Constitutional Budget Reserve Fund	1992
	Louisiana	1990	Revenue Stabilization and Mineral Trust Fund	1999
	New Jersey	1990	Surplus Revenue Fund	1993
1990–1991	Arizona	1991	Budget Stabilization Fund	1994
	Nevada	1991	Fund to Stabilize Operation of State Government	1994
	North Carolina	1991	Savings Reserve Account	1991
	South Dakota	1991	Budget Reserve Fund	1992
	Iowa	1992	Economic Emergency Fund/Cash Reserve Fund	1993
	Missouri	1992	Budget Stabilization Fund	1992
	Virginia	1992	Revenue Stabilization Fund	1993
	West Virginia	1994	Revenue Shortfall Reserve Fund	1995

or agency-specific; but on the whole, these funds are reserves *for specific purposes*, e.g., natural disasters, litigation settlements, self-insurance, and so on. Therefore contingency funds are not the same as BSF.

Strict Definition of the Budget Stabilization Fund

Thus two distinct definitions of BSFs emerge, one relatively strict and the other loose. Adopting the strict or the loose definition of BSFs will end up with quite different data sets and thereby results of empirical analysis. The strict definition is clear and straightforward

with regards to the accumulation of the fund (during good times) and purpose (to cope with revenue shortfall during recessions). It is:

A budget stabilization fund is money set aside while good economic conditions prevail and then drawn down during poor economic times to lessen the extent to which state governments will have to reduce expenditures or raise taxes to cope with a recession.^[4,9,12]

The loose definition purports that the fund is not only for temporary shortfalls, but also for any non-recurring expenditures. The latter gives this definition a much broader scope. Consequently, the loose

definition is broad enough to count a *contingency fund* as a BSF. The loose definition of the BSF runs:

A budget stabilization fund is money set aside... against the chances of 'reducing service levels or raising taxes and fees because of temporary revenue shortfalls or unpredicted one-time expenditures.'^[10]

ADOPTION OF BUDGET STABILIZATION FUND BY THE STATES

The creation of budget stabilization funds by the states has been slow. But recessions and tax revolts have highlighted the usefulness of BSF as an effective fiscal tool. Gold^[6] noticed that the chronology of BSF adoption shows close correlation between establishment of BSF and recessions (Table 2).^[17]

Every one of the earliest state adopters of the BSF—New York, Florida, Tennessee, and Michigan—came within 3 years of a previous recession. The recessions in the early 1980s made the correlation more apparent: fiscal stress that led to spending cutbacks and increased taxes as an effort to sustain funding for public services prompted 16 states (from 1981 to 1985) to establish their BSFs.^[18] Once initiated, the adoption of BSFs spread broadly among the states. Then the 1990–1991 recession further increased the number of states with BSFs.

BUDGET STABILIZATION FUND DATA

Up to now, the *Fiscal Survey of the States* series compiled and published twice a year by the National Association of State Budget Officers (since 1977) is the only readily available source that provides BSF data of the states. Many empirical studies have based their analysis on this data source. A drawback with this series, however, is that the BSF figures provided by the states and reported in the series often do not match with those in the *Comprehensive Annual Financial Reports* (CAFR) of the states. The discrepancy may be a result of various reasons. However, because the CADRs are audited financial documents, they carry weight that should not be ignored. Therefore researchers and practitioners should exercise caution in their choice of data sources.

The usefulness of the budget stabilization fund as a subnational countercyclical fiscal device is now widely recognized—over 40 states have adopted the fund and researchers have begun to pay increasing attention to its study. However, comprehensive and conclusive examinations of this instrument are still lacking.

REFERENCES

1. The Old Testament, Genesis: 41.
2. Clow, F.R. *A Comparative Study of the Administration of City Finances in the United States, with Special Reference to the Budget*; MacMillan: New York, 1901. Sponsored by the American Economic Association.
3. Rubin, I. *Class, Tax & Power—Municipal Budgeting in the United States*; Chatham House Publishers, Inc.: Chatham, NJ, 1998; 35–36.
4. United States Congress. *Federal and State Roles in Economic Stabilization—Hearings Before a Subcommittee of the Committee on Government Operations House of Representatives, Nov. and Dec. 1984*; U.S. Government Printing Office: Washington, DC, 1985.
5. *Annual Report of the Comptroller, State of New York, for the Fiscal Year Ended March 31, 1946*; Comptroller's Office: Albany, NY, 1946; 5 pp.
6. Gold, S.D. *State and Local Fiscal Relations in the Early 1980s*; The Urban Institute Press: Washington, DC, 1983.
7. National Association of State Budget Officers (NASBO). *State Balanced Budget Requirements: Provisions and Practice*; 1992. Washington, DC.
8. Briffault, R. *The Balanced Budget Debate: Balancing Acts—The Reality Behind State Balanced Budget Requirements*; The Twentieth Century Fund Press: New York, 1996.
9. National Association of State Budget Officers (NASBO); National Governors' Association (NGA); National Conference of State Legislators (NCSL). *The Fiscal Survey of the States*; February, 1985; 6 pp. Washington, DC.
10. Government Finance Officers Association (GFOA). Develop policy on stabilization funds. In *Recommended Budget Practices*; 1999; 17 pp.
11. *Michigan Legislation 1977*. Michigan Compiled Laws, Chapter 18: Dept. of Management and Budget, Sec. 351–359.
12. Sobel, R.S.; Holcombe, R.G. The impact of state rainy day funds in easing state fiscal crises during the 1990–1991 recession. *Public Budg. Finance* **1996**, 16 (3), 28–48.
13. State of Michigan Act 76 of 1977 (repealed).
14. State of Michigan Act 431 of 1984.
15. Larkin, R.P. Impact of management practices on municipal credit. In *Special Report in Public Finance*; FITCH IBCA, May 4, 2000.
16. United States General Accounting Office. *Budgeting For Emergencies—State Practices and Federal Implications*; Government Printing Office: Washington, DC, September, 1999.
17. National Conference of State Legislatures. *Preparing for the Next Recession: Rainy Day Funds and Other Tools for the States*; Legislative Finance Paper No. 41, 1983.
18. Douglas, J.W.; Gaddie, R.K. State rainy day funds and fiscal crises: rainy day funds and the 1990–1990 recession revisited. *Public Budg. Finance* **2002**, 22 (1), 19–30.

Budgetary Analysis and Economics

John R. Bartle

*School of Public Administration, University of Nebraska at Omaha,
Omaha, Nebraska, U.S.A., and Center for Public Administration,
Sun Yat-Sen University, Guangzhou, Guangdong, P.R. China*

INTRODUCTION

The discipline of economics offers the potential to apply a systematic and rational approach to public budgeting. In particular, the concept of economic efficiency can be used to identify the goods and services for which a private market does a good job of distributing and also those goods and services for which it does not perform well. This analysis can help to answer one of the primary questions in budgeting—"What should government do?"—and provides a clear justification for the answer. Further, economic methods such as cost-benefit analysis (CBA) can provide answers to more specific budgetary questions, such as "Which program should be funded?," "How beneficial is the program?," and "When will the benefits and costs of the program be realized?" At the same time, there are some limitations with the application of economics to budgeting.

ECONOMIC EFFICIENCY AND GOVERNMENT

Markets are economically efficient if firms are producing output at the minimum cost and if those consumers who value the goods most get them. When this is not the case, a market failure exists. Market failures can exist for several reasons, specifically public goods, externalities, monopoly, and information problems in the market. When a market failure exists, the private market will not be efficient. From an economic viewpoint, this provides the central justification for government supply of certain goods or services, as appropriate government action can correct the market inefficiency. Musgrave referred to this rationale as the allocative role of government.^[1] For example, pollution is an undesirable by-product of the production of many desired goods, such as steel. People near the factory would be willing to pay to have the pollution reduced but find it difficult to organize to do so. The government can correct the market failure by taxing polluting firms for the cost to society of their emissions. This tax induces the firm to internalize the external costs into their production decisions. They may continue to pollute but, typically, will do so at a lower level. It is argued that without some government

intervention into this market, the externality will continue and the market will remain inefficient. In this manner, the government can make the economy more efficient. The same is true of the other types of market failures and is one of the justifications for programs such as national defense and food safety (public goods), zoning and anti-nuisance ordinances (externalities), antitrust litigation and price regulation (monopoly), and public pensions and health care regulation (information problems).

In 1960, Coase pointed out that the presence of a market failure should not necessarily lead to the conclusion that government should intervene.^[2] In the pollution example, it might be possible for the residents of the community to organize themselves in some other fashion to negotiate with the polluting firm. How possible this is depends on the transaction cost of their mobilization and collective action. Coase's point provided one of the intellectual justifications for the subsequent privatization movement. More generally, just as markets can fail to be efficient, so can governments. Therefore, the appropriate comparison in the face of any inefficiency is between the two institutions of the market and the government, and further between the various service provision arrangements possible. Some possible arrangements are direct government provision, intergovernmental agreements, contracts, franchises, vouchers, and voluntary action. This perspective ties these economic concepts to the public administration literature on policy implementation and service provision in a way that can help the student and the practitioner of public administration to conceptualize the rationale for action and the range of alternatives in providing services. In turn, this shapes the "big picture" of the budget, as well as the options a budget official may consider.

Musgrave's other two roles for government are stabilization and distribution.^[1] Stabilization refers to macroeconomic concerns about issues such as unemployment, inflation, international trade, monetary policy, and interest rates. The distribution role refers to concerns about the distribution of income and other resources among members of society. Of course, one might add to these roles. But Musgrave made it clear that other concepts besides economic efficiency were relevant in determining government programs and budgets.

Okun advanced the argument that in many policy decisions, there is a trade-off between efficiency and equity.^[3] For example, in health care, a pro-efficiency viewpoint would argue for reliance on the private market, with little or no interference by the government. Those who valued the services highly and could afford them would get them with little interference, and producers would have incentives to respond to consumers through the market. However, those who could not afford treatment might have to go without. A pro-equity argument might de-emphasize the private market as a basis for health service provision in favor of public provision or arrangement of health services, ensuring access and coverage to all. However, this strategy would likely require higher taxes, limit provider incentives, and create non-price forms of rationing that might be troubling. Okun advocated balancing these two social goals of efficiency and equity. Although he realized that this balance was difficult to strike in practice, many economists see it as a useful construct for policy decisions.

FISCAL FEDERALISM

The concept of economic efficiency can also help to sort out which level of government (national, state/provincial, or local) should provide which services. Oates pointed out that some goods are local public goods (e.g., public safety), whereas others are more appropriately supplied at the state, regional, or national level.^[4] Local public goods are characterized by a smaller scope of impact. For example, fire safety in Boston is important to residents of that city, and perhaps in other nearby cities, but not to Chicago. The efficiency criterion would dictate that citizens of Boston should decide what level of fire protection they want and the amount they are willing to pay for it. Because Bostonians' decisions have little or no cost and benefit on people outside that area, the correspondence principle would dictate that the geographic area of the benefit and the taxes to support the service should coincide. If tastes for public services and tax levels vary among communities, then this implies that levels of service will not be and should not be uniform. Further, because different goods have different spatial distributions of costs and benefits, a federal system has the potential to be more efficient than a centralized system, as it better allows for local variation.

POSITIVE AND NORMATIVE APPLICATIONS OF ECONOMICS TO BUDGETING

A positive theory attempts to explain "what is," whereas a normative theory attempts to explain "what

should be." Economic theories can be either positive or normative. A common criticism of the economic perspective of government action is that it is not a good positive model; many of the decisions about programs and public spending are political and do not follow this or any other prescribed theory about what government should do. For example, despite the fact that it does not address any market failure, education is provided publicly in many nations. Indeed, education is currently the largest functional area of spending among all governments in the United States.^[5] Further, the structure of American government often makes it difficult to enact efficient policies. The institutions of constitutional separation of powers, bicameralism, and federalism require that any policy proposal receive majority approval among many different political bodies. This tends to produce policies that have a broad distribution of benefits and a vague distribution of costs. Efficient policies, however, tend to have explicit and visible linkages between costs and benefits. Thus, as a general rule, the policies emanating from U.S. governments are not particularly likely to meet the strict standard of efficiency. This may be less true of unified governments, such as parliamentary systems.

Economists have produced a number of positive models relevant to government budgeting. The best known is the median voter model, which assumes that in each polity a single individual is the decisive voter. The preferences, income, and prices faced by this individual then help to predict their vote and, in turn, the government's budget. A great number and variety of articles have been published on this model, with mixed conclusions (for a review, see Ref.^[6]). Despite some problems with this model, it does demonstrate the attractive traits of positive economic models. They are testable and relatively simple and can be used to describe, explain, and predict government behavior.

Normatively, it is often argued that the economic perspective is flawed. There are many different angles that this argument takes, but a common one is that it is undemocratic. If a majority of the citizens want to tax themselves to provide a program that they want, then this alone is a sufficient justification, whether or not there is a market failure. Certainly, this criticism has merit; yet, it does seem that the efficiency rationale serves as one good reason—among many—for justifying government programs. The efficiency rationale is grounded in the utilitarian justification of achieving the greatest satisfaction possible for all individuals given social resources available. Although utilitarianism has its limits as a guiding social principle, most economists are comfortable with it, because it maintains the focus of policy on the individual and allows individuals to make their own decisions. Philosophies that put the government first and the individual second

tend toward collectivism. As with most normative debates, there are compelling arguments on either side.

COST-BENEFIT ANALYSIS

Most programs entail gains for some and losses for others. A vexing problem is how to resolve these distributional consequences, yet still be assured that the program is in the social interest. Economists have contributed three decision rules to attempt, which answer this question (for a detailed discussion, see Ref.^[7]). The strictest test is the “Pareto rule,” which states that a program should be undertaken if it makes at least one person better off and no one worse off.^[8] Conceptually, this is appealing, but practically it is of limited help because few programs leave no one worse off. A second test is the “Kaldor-Hicks rule,” which states that a program should be undertaken if those benefiting from it could compensate those hurt by it and still be better off. The winners need not actually compensate the losers; this test is conceptual rather than actual.^[9,10] The third, and least strict, test is the fundamental rule of CBA—do all projects with a positive net present value of benefits minus costs. (Present value is a method for adjusting economic values over time, so that dollar values of costs or benefits realized at different times may be compared.) These three tests guide decision makers in selecting programs with distributional consequences. The strictness of the standards of these tests is important because it affects which programs will be selected, with the fewest programs passing the Pareto rule and the most passing the CBA rule.

Once a market failure is identified and service provision arrangement is chosen, several other operational decisions need to be made about the design and scope of the program. CBA has been applied to budgeting to guide decisions about the advisability and design of programs, as well as a means to justify expenditures. In deciding about whether to adopt a program, CBA would recommend its adoption if the net value is positive and recommend against it if it is negative. If there are several competing programs and one were able to pick only one of them, CBA would recommend selecting the program with the highest net benefit. If one were trying to determine the total amount that should be spent on a collection of programs, CBA would recommend doing all those that had a positive net benefit and none of those that did not. Although the decision rule is simple and the application flexible, CBA does require care in identifying what values should be included; how they should be measured; how risk, timing, and uncertainty are dealt with; and the discount rate used to recognize the time value of money.

Although it is becoming more popular, CBA is not routinely used in most agencies. It requires advanced technical skills, and the details of the analysis can be difficult to explain. Further, many non-economic considerations come into play in decisions about public spending. Despite this, the growing demand for accountability in the public sector and the business-like “bottom line” that a CBA can deliver has recently made this approach more attractive to administrators, policy makers, and the public.

AN APPLICATION TO THE BUDGETARY PROCESS

In a classic piece, Lewis argued that the economic approach of marginal analysis could be productively applied to public budgeting.^[11] He suggested a budgetary process, where, in addition to their appropriation requests, each agency would also detail how their request would change if the appropriation were, for example, 10–20% lower or higher. This would present decision makers with packages of programmatic decisions meant to facilitate, if not force, a comparison both among programs and between public and private spending. The goal of such an approach is to compare the relative value of programs, in accomplishing social goals and making the opportunity cost of budgetary decisions readily apparent. This recognizes the basic economic insight that it is the marginal or additional unit that determines the value of a good or service.

The approach Lewis described foreshadowed many budget reforms that have been implemented (e.g., target-based budgeting, zero-based budgeting, and program budgeting), as well as institutional arrangements (e.g., the U.S. Congressional Budget Act of 1974, which in part attempted to present Congress with explicit decisions of this type). Studying these reforms illustrates the strengths and weaknesses of the application of basic economic principles to public budgeting. Some governments incorporate these principles into their process of budgeting, although clearly political, and other considerations are relevant and often more important.

It is, therefore, apparent that any thoughtful practitioner of budgeting should be aware of these principles of economic analysis, as both a theoretical foundation and an analytical tool that can be used. Certainly, other valuable perspectives differ in their fundamental premises and normative approach to public budgeting (see, e.g., Ref.^[12]), and so the practitioner may not embrace the economic approach. However, it will continue to be relevant to this central function of government.

REFERENCES

1. Musgrave, R.A. *The Theory of Public Finance: A Study in Public Economy*; McGraw-Hill: New York, 1959; 3–27.
2. Coase, R.H. The problem of social cost. *J. Law Econ.* **1960**, *3*, 1–44.
3. Okun, A.M. *Equality and Efficiency: The Big Trade-off*; The Brookings Institution: Washington, 1975.
4. Oates, W.E. *Fiscal Federalism*; Harcourt Brace Javonovich: New York, 1972; 3–20.
5. U.S. Census Bureau. *Statistical Abstract of the United States: 2007*; Economics and Statistics Administration, U.S. Department of Commerce: Washington, DC, 2007; pp. 271, 309.
6. Kearns, P.S.; Bartle, J.R. The median voter model in public budgeting research. In *Evolving Theories of Public Budgeting*; Bartle, J.R., Ed.; Elsevier Science Ltd.: Oxford, U.K., 2001; 83–100.
7. Gramlich, E.M. *A Guide to Benefit–Cost Analysis*; Prentice-Hall: Englewood Cliffs, NJ, 1990; 30–33.
8. Pareto, V. *Manual of Political Economy*; 1971 Translation of 1927 Ed.; Augustus M. Kelley: New York, 1906; 261 pp.
9. Kaldor, N. Welfare propositions of economics and interpersonal comparisons of utility. *Econ. J.* **1939**, *49*, 549–552.
10. Hicks, J.R. The valuation of the social income. *Economica* **1940**, *7*, 105–124.
11. Lewis, V.B. Toward a theory of budgeting. *Public Admin. Rev.* **1952**, *12*, 43–54.
12. Wildavsky, A. Political implications of budgetary reform. *Public Admin. Rev.* **1961**, *21*, 183–190.

Budgetary Balance and Deficit Controls

William R. Voorhees

School of Public Affairs, Arizona State University, Tempe, Arizona, U.S.A.

INTRODUCTION

Budgets are generally considered to be in balance when an organization's inflow of economic resources is equal to its outflow of economic resources. When an organization allocates resources beyond those currently available, a deficit occurs and additional resources must be collected in future periods reducing future opportunities. In its simplest form, budgetary balance is expressed by the budgetary equation:

Revenues = Expenditures

In a neat and tidy world, this equation would serve us well; however, in the real world, the equation is seldom as simple as it appears. Economic stability, political ideology, fiscal management skills, and fiscal policy will all influence whether a budget is in balance. Even when the budgetary equation appears to be in balance, it may be so only after an application of "smoke-and-mirror gimmickry," such as accelerating revenues from or delaying expenditures to subsequent periods. When the budgetary equation is not in balance, surpluses or deficits are the result.

There are two basic categories of budget deficits: cyclical and structural. Cyclical deficits are the result of downturns in the economy. When downturns occur, expenditures are prone to increase while revenues decrease, often resulting in a cyclical deficit. Structural deficits, however, are the result of permanent deficits resulting from the expenditure or revenue structures making up the budget.

CYCLICAL DEFICITS

When considering the budgetary equation and balance, it is necessary to understand how forecast accuracy affects each side of the budgetary equation. On the revenue side, forecasts are generally accurate in the United States, with states normally having less than a 2% error in their revenue forecasts. The problem arises with the inability to forecast turning points in the economy in time for inclusion in the budget. Turning points represent the peaks and valleys of a cyclical economy and because most revenues are conditional on the state of the economy, the accuracy of the revenue forecast is

affect when an economy turns. Because estimating turning points is extremely difficult, governments are often faced with revenue shortfalls when the economy turns downward. Likewise, when the economy begins a growth period after an economic downturn, revenues will often exceed forecasts resulting in excess revenues.

The behavior of expenditures during cyclical changes presents officials trying to maintain a balanced budget with equally challenging problems. During periods of economic downturns, expenditures will generally increase due to mandatory spending. Entitlements such as food stamps, unemployment relief, and welfare will increase during economic downturns as more of the population find themselves unemployed or working for lower wages. These additional expenditures create additional fiscal stress at exactly the time when revenues are falling. As workers lose jobs and wages fall, revenues derived from income and consumption taxes are affected correspondingly.

Recissions and Supplemental Budgets

The primary tools that governments have to fight cyclical imbalance are recissions and the supplemental budget. Recissions are cuts to the budget after the budget has been passed. Because governmental expenditures are in reality, planned activities of the government, a recission can present severe disruptions to government operations. Often, because expenditure commitments have already been made, recissions may not come from programs that either politicians or citizens want cut.

Recissions result in budgetary cutbacks that can come in several forms, including line item reductions, across-the-board departmental reductions, or programmatic reductions. Line item reductions may result in hiring and salary freezes, early retirement, unpaid furloughs, layoffs, reduced maintenance on equipment and fixed assets, and elimination of new equipment purchases.^[1] Departmental cuts may be across the board, with each department required to contribute a fixed percentage. Across-the-board cuts, where all departments bear an equal burden, are often contrary to optimal budget allocation and result in both high- and low-priority programs being cut equally. However, programmatic cuts fall on specific programs, especially those that are allocated large amounts of funds and

that are not federally funded, mandated, or supported by earmarked funds.^[2]

In the case of an upturn in the economy and excess revenues, supplemental budgets can be appropriated after the original budget has been passed. Supplemental budgets allow lawmakers to add additional dollars to the budget through additional expenditures or tax expenditures. Supplemental appropriations may fail to provide optimal resource allocations, and they often have limited citizen and legislative participation in the allocation of the additional resources.^[3]

Stabilization Funds

To mitigate the problems of forecasting, many states have adopted reserve funds to maintain budgetary balance. Typically, these funds go by several names, including rainy day funds, contingency funds, and stabilization funds. Although any of these funds can be used to smooth out revenue shortfalls and higher expenditures due to economic fluctuations, only stabilization funds are used specifically for that purpose.

Three features—fund-sized, funding mechanisms, and fund expenditure—identify the structure of a stabilization fund. The most common suggestion for fund size at the state level is a 5% reserve.^[4] In reality, some states seem not to have set target amounts for their funds, whereas other states have statutes controlling and limiting fund size. In 1998, most states with stabilization funds had balances between 2% and 14%.^[5]

The second feature, funding mechanisms, can vary from predefined constitutional requirements to no requirements at all. For example, South Carolina requires a 5% balance in the fund and, if the fund is less than the 5% amount, a minimum of 1% must be added each year until the 5% minimum is achieved.^[6]

Finally, fund withdrawal varies considerably among states with some states using a formula approach that is triggered in times of fiscal distress; other states withdraw funds after either a majority or supermajority vote of the legislative; and, in still other states, the executive need only declare a need to use stabilization funds.^[7]

At the municipal level, few cities have opted to implement stabilization funds for economic smoothing. Interestingly, those cities most sensitive to economic cycles are no more likely to establish stabilization funds than the least sensitive. In fact, those municipalities least dependent on cyclical revenue are most likely to institute such funds.^[8]

STRUCTURAL DEFICITS

Structural deficits are the result of permanent deficits resulting from the expenditure and/or revenue

structures of a government's budget. Whereas cyclical deficits disappear during economic growth cycles, structural deficits can only be corrected by permanently reducing expenditures or increasing revenues. Because of balanced budget requirements and debt limitations, state and local governments generally have fewer problems with structural deficits than the federal government. Fig. 1 shows the federal government deficits as a percentage of the gross domestic product (GDP). It is useful to consider the deficit in respect to the GDP to offset economic growth trends. In looking at the figure, a structural deficit is clearly apparent from 1975 to 1995, although some might argue that a structural deficit existed even prior to 1975.

Process Controls at the State Level

A primary means of controlling budgetary balance is through controls on the budgetary process. Such process controls may require specific reports or budgetary formats, allow for executive veto of budgetary items, or consist of laws that mandate a balanced budget. Budgetary formats may or may not place emphasis on the budgetary balance aspect of the appropriations process. Formats such as zero-based budgeting place a stronger emphasis on budgetary balance with their built-in mechanism for identifying potential budgetary cuts. In addition to traditional budgets, some states are now requiring tax expenditure budgets that aid in identifying revenue reductions that may lead to budgetary imbalance if other actions are not taken to increase other revenues or decrease expenditures.

The executive veto has often been suggested as a primary tool of the executive in maintaining budgetary balance at the state level. Even when the executive fails to exercise a veto, the threat of the veto may prevent the legislature from adding appropriations they know the executive has threatened to veto.

Several forms of the veto exist, including the popular line item veto. The traditional line item veto permits the executive to delete one or more budget lines. In some states, the item veto is extended to allow the executive to change the wording of the legislation. Currently, only the governor of one state has no veto power, whereas 42 states allow a line item veto, 40 allow an item veto of appropriations, 14 allow veto of selected words, and 3 allow the governor to veto and change the wording of an appropriations bill.^[9] It should be noted that veto power was given to the U.S. president in 1995 with passage of the Line Item Veto Act, but was rescinded in 1998 when the Supreme Court ruled it unconstitutional. The effect of the line item veto power was considered to be miniscule during the short time the president had the authority.^[10]

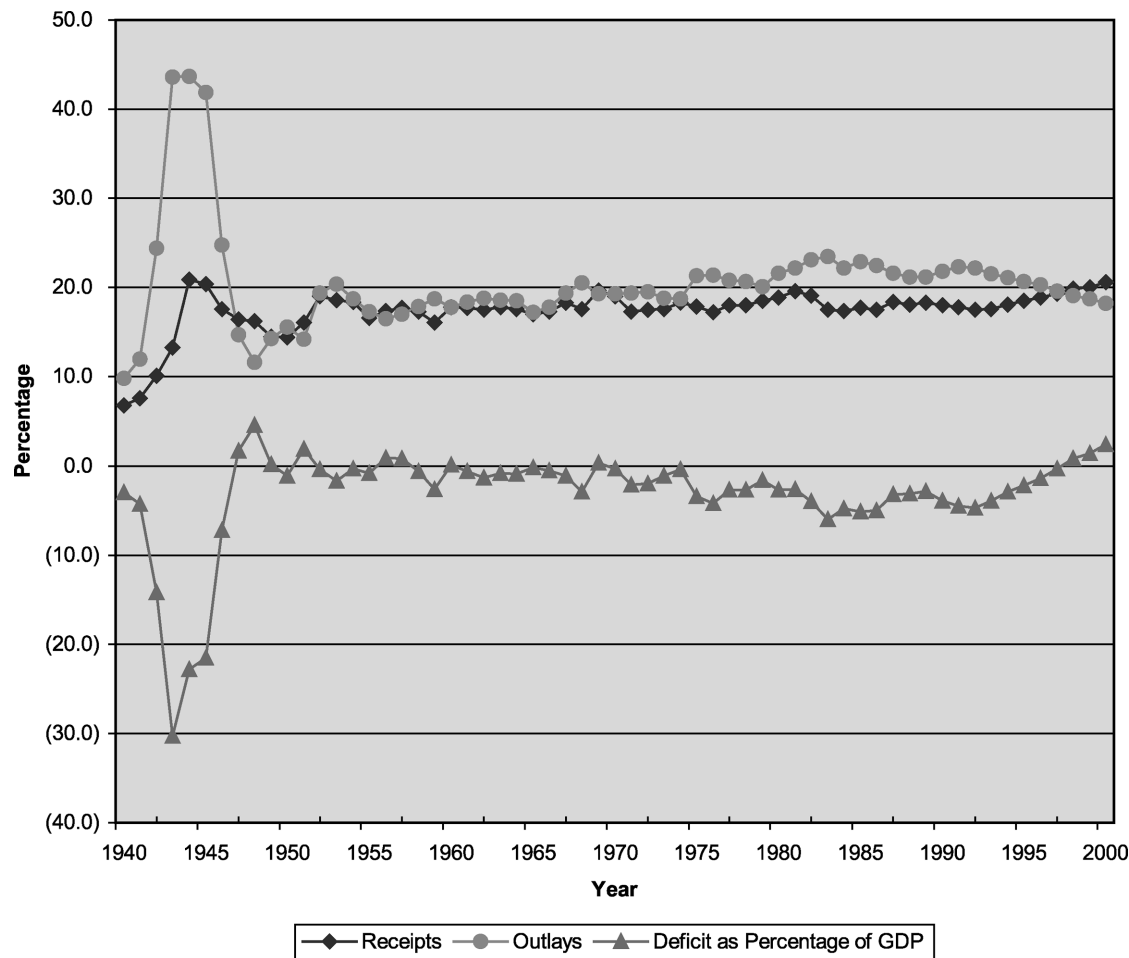


Fig. 1 U.S. receipts, outlays, and deficits/surpluses as percentage of GDP. (From the United States FY 2002 Budget.)

Other controls may take the form of laws that require states to have budgetary balance. These laws may set requirements at different points in the budgetary process such as submission of a balanced budget by the executive, passage of a balanced budget by the legislature, or a requirement for the executive to sign a balanced budget. Often, these requirements are written into the state's constitution, which generally requires a supermajority to modify. Statutory requirements may also be used by states to implement balanced budget requirements; however, in times of fiscal stress, statutory requirements are more easily circumvented.

Some states also have laws that require the state to achieve balance at year end (Table 1). In a survey conducted by the General Accounting Office (GAO), it was found that 39 out of 49 states required budgetary balance at year end. At the same time, it was found that 21 of those 49 states were allowed to carry over deficits to the following years. The same GAO report suggests that year-end budgetary balance is often "more of a perception than a legal requirement."^[11] For instance, in the state of Massachusetts, the law is written such that the budget may be balanced through the use of loans.^[12]

Table 1 Balance budget requirements in the states

	States with balance requirements ^a	Constitutional	Statutory
Governor must submit balanced budget	44	32	26
Legislature must pass balanced budget	40	33	18
Governor must sign balanced budget	35	30	17

^aSeveral states have both statutory and constitutional requirements. States are counted only once.
Source: NASBO, Budget Processes in the United States, 1999.

For states that do have budgetary balance requirements, the funds covered by the requirements may not always specify which funds fall under the requirement. The focus of balance regulations is usually directed at governmental operations and the general fund. Unlike the federal government, capital funds at the state level are generally excluded from budgetary balance requirements. Less clear is the manner in which fiduciary and enterprise funds are affected under balance requirements. In the case of enterprise funds, bond covenants may necessitate budgetary balance.

Process Controls at the Federal Level

At the federal level, the U.S. Congress has passed several pieces of legislation to control deficits. The first major effort by Congress to control deficit spending was the Anti-Deficiency Act of 1870, which prevented agencies from obligating more funds than appropriated and then submitting “coercive deficiency” requests to Congress to pay their obligations.

The Budget and Accounting Act of 1921 was the second major effort by Congress to control budgetary expenditures. This act centralized fiscal processes by codifying the president’s budget and creating both the Bureau of the Budget and the GAO. The result of the Budget and Accounting Act of 1921 was to shift substantial budgetary power and responsibility from the Congress to the executive branch of government.

In 1974, Congress passed the Congressional Budget and Impoundment Control Act, establishing the Budget Committees in both the House and the Senate. It also modified the budgetary process in Congress by requiring a Budget Resolution at two different points in the budgetary process, a tax expenditure analysis, a current services budget presentation by the president, and the creation of the Congressional Budget Office. It was the intent of Congress to create a budgetary process that both enhanced their control over the budget, and integrated revenue and appropriation decisions.

Persistent deficits of the late 1970s and early 1980s caused by increased budget complexity due to the growth of entitlements and multiyear spending, a structural imbalance due to the lost revenues of the Economic Recovery Tax Act of 1981, and increased budget vulnerability with respect to incorrect estimates and forecasts prompted Congress to once again address the budgetary process with attempts to limit expenditures.^[13] In 1985 and again with modifications in 1987, Congress enacted the Balanced Budget and Emergency Deficit Control Act (Gramm–Rudman–Hollings) with the intent of lowering the deficit and ultimately balancing the budget. This legislation

attempted to bring more fiscal discipline to the budgetary process by establishing new deadlines and deficit targets. If the deficit target was exceeded by more than \$10 billion, then automatic cuts or sequestration would be implemented. Sequestration required that the president make across-the-board spending cuts that were evenly divided between both domestic and defense programs to achieve the deficit targets. Because debt interest and most entitlement programs were either totally or partially exempt from budgetary reductions, the bulk of the reductions under sequestration came from discretionary appropriations. It was believed that the sequestration process would be so distasteful that Congress would attempt to avoid it at all cost. What was not anticipated was that Congress would devise means through the use of delayed appropriations and other “smoke-and-mirror gimmickry” to avoid the consequences of sequestration.

The deficits continued, forcing Congress to revisit the problem of budgetary balance in 1990 by passing the Budget Enforcement Act (BEA). The BEA created three categories of spending—domestic, defense, and international aid—with each category having separate deficit targets. The breach of a spending target in any single category would result in sequestration for programs in that category only and, if appropriations for one category fell below the target, the difference between the target and the appropriation could only be used for deficit reduction. This was to create a unique problem with defense appropriations in 1992. Due to the dissolution of the Soviet Union, defense spending realized a peace dividend, which spurred many legislators to proposed using the defense reductions for spending in the other two categories; however, by law, these categories were walled off preventing movement of funds from one category to another. After a protracted battle in Congress, the walls held and the reductions in defense spending were applied to deficit reduction.^[14]

In addition, BEA required that increases in direct spending or revenue reduction be financed on a pay-as-you-go basis (PAYGO) and prohibited any net change in the deficit. In effect, this requirement tied new spending to either tax increases or cuts in other expenditures, making new expenditures unlikely.

The Omnibus Reconciliation Act of 1993 revised and extended spending caps and significant reductions were realized in the budget between 1992 and 1997, with the budget deficit decreasing from \$290.4 billion to \$20 billion in 1997. The Balanced Budget Act was passed in 1990, requiring a balanced budget in the year 2002. This act once again revised and extended discretionary spending limits and PAYGO requirements, and made permanent a 5-year horizon for budget resolutions.^[15] With the Balanced Budget Act of 1997 in place along with strong revenues from a thriving

economy, the budget was balanced the following year, well ahead of the 2002 deadline.

CONCLUSION

The budgetary process is faced with two forms of deficits: cyclical deficits that are a result of the economic cycle, and structural deficits that are a result of the permanent imbalance between the institutional structures designed to raise resources and those design to use resources. Cyclical deficits are often solved through recissions and cutback management techniques. Structural deficits have been found to be more difficult to solve because they tend to pit various groups of resource consumers against each other over the long run. Governments have implemented various budgetary processes in attempts to control deficits, sometimes successfully and often to no avail. Often, the legislation was focused on integrating the revenue and expenditure decision-making processes. In the end, however, budgetary balance is probably best achieved through a combination of not only a strong economy and reasonable process controls, but also the political will of the groups competing for resources.

REFERENCES

1. Lee, R.; Johnson, R. *Public Budgeting Systems*, 6th Ed.; Aspen: Gaithersburg, MD, 1998; 140–144.
2. Lauth, T. Reductions in the FY 1992 Georgia Budget: responses to a shortfall. In *Case Studies in Public Budgeting and Financial Management*; Khan, A., Hildreth, B., Eds.; Kendall Hunt: Dubuque, IA, 1996.
3. Forrester, J.; Mullins, D. Rebudgeting: the serial nature of the municipal budgetary process. *Public Adm. Rev.* **1992**, *52* (5), 467–473.
4. National Conference of State Legislatures. *State Budget Actions in 1984*; National Council of State Legislatures: Denver, CO, 1984.
5. National Association of State Budget Officers. *The Fiscal Survey of States*; National Association of Budget Officers: Washington, DC, 1998.
6. National Association of State Budget Officers. *Budgeting Amid Fiscal Uncertainty: Stabilization Funds and Other Strategies*; National Association of Budget Officers: Washington, DC, 1985.
7. National Association of State Budget Officers. *Budget Stability: A Policy Framework for States*; National Association of Budget Officers: Washington, DC, 1995.
8. Wolkoff, M. An evaluation of municipal rainy day funds. *Public Budg. Finance* **1987**, *7* (2), 52–63.
9. National Association of State Budget Officers. *Budget Processes in the United States*; National Association of Budget Officers: Washington, DC, 1999.
10. Joyce, P. The federal line item veto experiment: after the Supreme Court ruling, what's next? *Public Budg. Finance* **1998**, *18* (4), 3–21.
11. United States General Accounting Office. *Balanced Budget Requirements: State Experiences and Implications for the Federal Government*; United States General Accounting Office, 1993; 16 pp.
12. Mikesell, J.L. *Fiscal Administration: Analysis and Applications for the Public Sector*, 5th Ed.; Harcourt Brace College Publishers: New York, 1999; 124 pp.
13. Leloup, L.; Graham, B.; Barwick, S. Deficit politics and constitutional government: the impact of Gramm–Rudman–Hollings. *Public Budg. Finance* **1987**, *7* (1), 83–103.
14. Doyle, R.; McCaffery, J. The Budget Enforcement Act in 1992: necessary but not sufficient. *Public Budg. Finance* **1993**, *13* (2), 20–37.
15. Leloup, L.; Long, C.; Giordano, J. President Clinton's fiscal 1998 budget: political and constitutional paths to balance. *Public Budg. Finance* **1998**, *18* (1), 3–32.

Budgetary Reforms in Southeast Asia

David Seth Jones

*Department of Public Policy, University of Brunei Darussalam,
Gadong, Brunei*

INTRODUCTION

In recent years, the states of Southeast Asia have implemented a range of budgetary reforms as part of a wider agenda to raise their levels of governance. The changes introduced have varied in their nature and extent according to the culture, political system, economic development, and administrative capacity of each country. In some cases, they have built upon an already well-developed system of budgetary management. In others, they represent the first serious steps to creating a viable budgetary system as in Laos, Vietnam, and Cambodia.^[1,2] This study examines the reforms implemented in the various countries concerning the different aspects of budgetary management, and highlights in the concluding section the challenges for further reform.

CASH MANAGEMENT IN BUDGET EXECUTION

An important area of budgetary reform in Southeast Asia is controls during disbursement to ensure that cash funds are available to meet current commitments. Too often in the past, cash flows and disbursements have been poorly regulated. This has meant that cash spending has not been well distributed during the financial year. Funds have dried up if the budget allocation has been disbursed too early. Alternatively, a spending spree may occur at the end if it is realized that there remains a surplus of cash that has not been used. In addition, poor cash management controls have allowed government agencies to undertake commitments and issue payment orders for which cash funds are not available. The upshot is that expenditure commitments are either delayed or shelved and the institutions that control the release of funds can exert arbitrary power over budget execution, which allows substantial deviations from the budget. This practice has been particularly noticeable over many years in Indonesia and was prevalent until recently in the Philippines.^[3,4] A further contribution to poor control of disbursements is the lack of proper authorization, verification, and accurate documentation of all payment transactions, which has undermined budget implementation management in Indonesia.^[4]

Consequently, several governments in the region have introduced measures to enable better management of cash funds by spending ministries and agencies, including tracking of cash transactions from month to month and assessing the status of their account balances. The Cambodian government is one which has taken seriously the need to do this. Monthly reports are now drawn up that provide details on payment transactions of government departments over the previous month and the status of their account balances. In addition, payment orders issued by government departments are strictly tied to cash availability for the forthcoming month. What is more, when a budget allocation has been finalized for the forthcoming financial year, ministries and agencies are then required to project their spending requirements for each month of the year. As the year progresses, they can identify the variance between how much was spent in each month and in the year to date, and how much was projected.^[5] In the Philippines, under a recent measure, the availability of cash funds is ensured through the automatic release of 80% of a budget appropriation for a spending institution at the beginning of the financial year.^[3]

DECENTRALIZATION

Common to several countries in Southeast Asia, which have introduced democratic reforms, has been a subsequent decentralization of public administration through the creation of elected local authorities with responsibility to provide public services and to manage and implement their own budgets based mainly on central government grants. The decentralization reforms have been recommended and guided by international organizations, such as the World Bank, Asian Development Bank, UNDP, and the IMF, which have argued that local institutions are better able than central administration to identify and target the most pressing needs within their locality.

One example has been Indonesia, where a newly decentralized system of public administration was created in 2001. The decentralization included delegating to elected regional parliaments discretion in apportioning central government block grants in the light of regional priorities. The central government is obliged

to transfer a minimum of 25% of its revenue to the regional authorities, which also are given limited powers to levy local taxes.^[6-9] Cambodia is the most recent example of decentralization. In 2002, elected local government structures were created known as commune councils, following the first local elections. Each commune council is given a block grant from the Commune Fund managed by the central government. This grant can be spent by the commune council in accordance with local needs, especially on poverty alleviation and employment creation schemes.^[10-12]

TRANSPARENCY

Another important aspect of budgetary reform is the establishment of more transparent systems of state financial management, which is vital to ensuring control and accountability in the budget process. At the beginning of the financial year, transparency necessitates the publication of a wide-ranging budget documents submitted to the legislature covering current and capital expenditures, government loans and grants, sources of revenue, and state assets and liabilities. Expenditure should be broken down according to ministry, department, program, activity, and object of expenditure (staffing, supplies, etc.).^[13]

At the other end of the budgetary process, transparency hinges on proper structuring and reporting of the public accounts at the end of the financial year in line with international accounting standards. The accounts should show the expenditures incurred, revenues collected, variance from the estimates and projections in the budget, extent and type of debt issued, as well as a detailed balance sheet showing assets (including capital reserves and equity holdings) and liabilities (including contingent liabilities) of the government. The accounts should be drawn up based on double-entry book keeping.^[13]

In most of the countries of the region, steps have been taken to improve budgetary and public accounts disclosures, especially in tandem with the strengthening of accountability institutions, such as the supreme audit institution and the legislature. The Philippine government has taken seriously its duty to disclose budget information in a detailed and comprehensive format following the introduction of democratic politics in 1986. The budget document submitted to the legislature, and authorized by the passing of the General Appropriation Bill, lays out for each spending institution, expected expenditures for its programs and objects of expenditure, including transfers and loans. It also spells out in detail assets and liabilities of the state.^[3]

An example of improvements in the disclosure of public accounts is provided by the Government of

Vietnam, which has sought to bring its accounting standards in line with International Accounting Standards (IAS) as part of its long-term effort to create a modern market and trading economy.^[12,14] The Government of the Philippines has followed a similar path in adopting IAS as the basis of its budget accounting.^[15] Likewise in Indonesia, new accounting standards for public finance were implemented in 2002 to include double-entry bookkeeping and disclosure of the overall balance sheet of the government, which had hitherto not been done.^[16,17]

A continuing practice that erodes transparency is off-budget expenditure. All too often ministries and other public agencies engage in spending outside and unrelated to any legal appropriation within the budget. Off-budget expenditures may escape disclosure and scrutiny, and thereby erode accountability. They can also seriously distort the overall soundness of a state's public finances. The opportunity for off-budget expenditure arises particularly when there are many small bank accounts in which budget funds are deposited, as opposed to a single consolidated account.^[13]

To tackle off-budget expenditure, the Indonesian government has initiated measures to create a single consolidated account for all government revenue, out of which all public spending will be funded and cash disbursements undertaken. It is hoped this will reduce the loopholes that had hitherto existed for off-budget spending through the myriad bank accounts for public funds, with one or more for each spending agency.^[16-19]

ACCOUNTABILITY

Another aspect of budgetary reform has been the strengthening of watchdog functions to ensure that: a) spending and revenue raising comply with laws and regulations governing financial procedures; b) public monies are disbursed in accordance with what has been authorized in the budget; c) and mismanagement, waste, and corruption are reduced. Achieving a more accountable system of financial management of public resources has necessitated improvements in auditing, legislative scrutiny, and anticorruption programs.^[13]

Central to creating effective oversight of public spending is the strengthening of the external audit of government finances through the supreme audit institution (SAI). In most Southeast Asian countries, reforms have been implemented in recent years that have either created a SAI or enhanced its powers and independence so as to ensure that public spending and revenue raising are in accordance with budget provision, financial regulations, and overall sound financial practices. The aim, of course, is to bring to light

non-compliance, waste, corruption and other irregularities.^[13] Linked to this has been the promotion by international organizations, such as the IMF, World Bank, and Asian Development Bank (ADB), of higher standards of government auditing within Southeast Asian countries, so that they are on par with internationally acceptable standards.

The creation or strengthening of SAIs has been one of the consequences of the democratic reforms in certain Southeast Asian states. For example, the 1987 Constitution in the Philippines, under Section 2(2), gave the SAI, the Commission on Audit, a full remit to ensure 'the prevention and disallowance of irregular, unnecessary, excessive, extravagant, and unconscionable expenditures or use of government funds.'^[3] In Cambodia, a new SAI has been created under the *Law on Audit* passed in 2000, known as the National Audit Authority. It reports to both houses of the legislature, viz., the National Assembly and the Senate, within 9 months after the financial year, and its role is to expose any irregularity in budget implementation and in the management and accounting of public moneys and assets. The Authority began work in auditing the implementation of the national budget in 2002.^[5,11,12] According to the ADB, the *Law on Audit* and the creation of the National Audit Authority improved the situation immeasurably with regard to the government audit function in Cambodia.^[10]

Even in countries where democratic reform has not taken place, improvements in auditing of public accounts can be observed. Although at the apex of a one-party state and without the necessity to be democratically accountable, the Government of Laos (with the help of the ADB) has initiated reforms in the independent auditing of public accounts, especially by the setting up of a SAI, the National Audit Office, to conduct audits of line ministries, provincial authorities, and capital projects.^[2,12,20]

In addition, in four countries (the Philippines, Indonesia, Cambodia, and Thailand), the establishment of a more open and democratic form of government has enabled the legislature to exercise greater independence in the scrutiny of the budget estimates and end-of-year results.^[9,21,22] This has been done through more extensive debates on budget appropriations and tax legislation and through the work of relevant committees in the legislature. In the Philippine legislature, independent reviews of budget implementation and state financial management are conducted by the Congress Appropriations Committee and the Senate Finance Committee on the basis of the findings of the Commission on Audit. Such independent scrutiny was hardly possible prior to the introduction of democracy in 1986.^[3]

Accountability in budget management also depends on measures to combat corruption. In most countries

of Southeast Asia (Singapore being a notable exception), bribery and misappropriation have long undermined the ability of governments to achieve their budget goals and spend money for the purposes intended.^[23] However, steps have been taken in some countries to set up at least the legal and institutional framework to tackle corruption. For example, in Thailand, under the 1997 Constitution, the National Counter Corruption Commission was established. Its powers and ethical standards are spelt out in the National Counter Corruption Commission Act of 1999. The reach of the Commission includes the power to investigate the authenticity of disclosures of personal and family assets and liabilities, which senior public officials are required to make. It can also arrest officials suspected of corruption, and if it is proven to impeach them and confiscate their assets. Another part of its mandate is to identify and prevent conflicts of interest at senior levels of government.^[12,24] However, the Commission still lacks the resources and manpower to conduct detailed enquiries and bring to book those who have abused their office. In so far as it is a branch of the Prime Minister's Office, it may also lack the independence to pursue corrupt officials, where exposure could be politically damaging.^[12]

It may be argued that combating corruption also depends on more transparent systems of information disclosure, and the exposure of financial irregularities by the SAI and committees of the legislature. This is true to a point. However, in the Philippines, despite the transparency and accountability of its budgetary system, high levels of corruption continue. This perhaps attests to the necessity in tackling corruption for a highly specialized anticorruption agency with far-reaching powers and investigative expertise such as that which exists in Singapore but which is lacking in the Philippines.

NEW PUBLIC MANAGEMENT

Budgetary reform in Southeast Asian states, with well-developed budget systems, have in the last few years increasingly reflected the influence of so-called New Public Management (NPM), by which public administration is modeled on business practices.^[25] In the Governments of the Philippines and Malaysia, the influence of NPM is reflected in the shift to accrual accounting and in the use of target setting and performance indicators for public services.^[3,15,26]

However, it is in Singapore, well noted for the effectiveness of its budgetary management, that the imprint of NPM has been most noted. Included in the range of initiatives that the Singapore government have introduced in the last 10 years are linking budgets to performance and results, management autonomy for service

providers, contestability in service provision, accrual accounting, interdepartment charging, point of delivery charging based on cost recovery, and revenue retention by ministries and agencies that levy such charges.^[27,28]

The absence of NPM in its fullest sense in other Southeast Asian countries is not altogether surprising, given the added weight of responsibility, it places upon public organizations in terms of data gathering and analysis, performance monitoring, and professional competencies of middle and junior managers. Only when major advances are made in these areas will it be possible to apply NPM to the budgetary process in these countries.

CONCLUSIONS

In summary, significant budgetary reforms have been introduced in most countries of Southeast Asia in the last 10–15 years. They have focused on improving fund disbursement, decentralizing budget decisions, creating transparency, strengthening accountability, and applying NPM in a few cases. For these purposes, a wide range of laws and regulations have been introduced governing budgetary and financial procedures. Progress in the effective implementation of the reforms has, however, been more variable, and in certain countries impeded by poor organization, lack of training, concerns over the political implications of greater transparency and accountability, and most seriously the continuing prevalence of corruption. If these impediments can be resolved, then budgetary reform will have an even greater impact.

REFERENCES

1. Painter, M. Public administration reform in Vietnam. In *Governance and Public Sector Reform in Asia: Paradigm Shifts or Business as Usual?* Cheung, A., Scott, I., Eds.; Routledge Curzon: New York, 2003; 208–227.
2. Bourdet, Y. *The Economics of Transition in Laos: From Socialism to ASEAN Integration*; Edward Elgar: Northampton, MA, 2000.
3. International Monetary Fund. In *Philippines: Report on the Observance of Standards and Codes—Fiscal Transparency*; International Monetary Fund: Washington, DC, 2002.
4. World Bank. In *Combating Corruption in Indonesia: Enhancing Accountability for Development*; World Bank: Washington, DC, 2003.
5. Royal Government of Cambodia. In *Letter of Intent to IMF, Supplementary Memorandum of Economic and Financial Policies and Technical Memorandum of Understanding*; Royal Government of Cambodia: Phnom Penh, 2002.
6. Jemal-ud-din, K.; Steer, A.; Homi, K.; Hoffman, B. *Decentralizing Indonesia: A Regional Public Expenditure Review: Overview 7*; World Bank: Washington, DC, 2003.
7. Fanany, I. The first year of local autonomy. In *Autonomy and Disintegration in Indonesia*; Kingsbury, D., Aveling, H., Eds.; Routledge Curzon: New York, 2003; 189–200.
8. Ray, D.; Goodpaster, G. Indonesian decentralization. In *Autonomy and Disintegration in Indonesia*; Kingsbury, D., Aveling, H., Eds.; Routledge Curzon: New York, 2003; 75–96.
9. Rosser, A. What paradigm shift? Public sector reform in Indonesia since the Asian crisis. In *Governance and Public Sector Reform in Asia*; Cheung, A., Scott, I., Eds.; Routledge Curzon: New York, 2003; 227–247.
10. Asian Development Bank. In *Financial Management and Governance Issues in Cambodia*; Asian Development Bank: Manila, 2000.
11. International Monetary Fund. In *Cambodia: Selected Issues and Statistical Appendix*; International Monetary Fund: Washington, DC, 2003.
12. Westcott, C. *Key Governance Issues in Cambodia, Lao PDR, Thailand, and Viet Nam*; Asian Development Bank: Manila, 2002.
13. Shivaio-Campo, S.; Tommasi, D. *Managing Government Expenditure*; Asian Development Bank: Manila, 1999.
14. Asian Development Bank. In *Financial Management and Governance Issues in Viet Nam*; Asian Development Bank: Manila, 2000; Vol. 1.
15. Narasimham, R.; Reid, B. *Diagnostic Study of Accounting and Auditing Practices in the Philippines*; Manila: Asian Development Bank, 2002.
16. Financial Management Reform Committee, Government of Indonesia. In *Reform of Public Financial Management System in Indonesia: Principles and Strategy*; Ministry of Finance, Government of Indonesia: Jakarta, 2001.
17. Rosser, A. Governance, markets and power: the political economy of accounting reform in Indonesia. In *International Financial Governance under Stress: Global Structures Versus National Imperatives*; Underhill, G., Zhang, X., Eds.; Cambridge University Press: New York, 2003; 263–282.
18. Asian Development Bank. In *Indonesia: Country Strategy and Program, 2003–2005*; Asian Development Bank: Manila, 2002.
19. World Bank. In *Combating Corruption in Indonesia: Enhancing Accountability for Development*; East Asia Poverty Reduction and Economic Management Unit, World Bank: Washington, DC, 2003.
20. International Monetary Fund. In *Lao Peoples's Democratic Republic: Selected Issues and Statistical Appendix*; International Monetary Fund: Washington, DC, 2002.
21. Hayllar, M. The Philippines: paradigm lost or paradigm retained? In *Governance and Public Sector Reform in Asia*; Cheung, A., Scott, I., Eds.; Routledge Curzon: New York, 2003; 248–270.

22. Roberts, D. *Political Transition in Cambodia 1991–1999: Power, Elitism and Democracy*; Curzon Press: Richmond, U.K., 2001.
23. Quah, J. *Curbing Corruption in Asia: A Comparative Study of Six Countries*; Eastern Universities Press: Singapore, 2003.
24. Brewer, B. Thailand: building the foundations for structural and systemic transformation. In *Governance and Public Sector Reform in Asia*; Cheung, A., Scott, I., Eds.; Routledge Curzon: New York, 2003; 186–207.
25. Turner, M. Choosing items from the menu: new public management in Southeast Asia. *Int. J. Public Adm.* **2002**, 25 (12), 1493–1512.
26. Common, R. *Public Management and Policy Transfer in Southeast Asia*; Ashgate: Aldershot, U.K., 2001.
27. Jones, D. Performance measurement and budgetary management in the Singapore civil service. *J. Public Budgeting Accounting Financial Manag.* **2001**, 13 (4), 485–511.
28. Jones, D. Recent changes in the financial management of the Singapore civil service: the adoption of resource accounting, internal charging and revenue retention. In *Public Service Reform in East Asia: Reform Issues and Challenges in Japan, Korea, Singapore, and Hong Kong*; Cheung, A.B.L., Ed.; Chinese University Press: Hong Kong, 2005; 105–124.

Budgeting and Public Choice

John R. Bartle

*School of Public Administration, University of Nebraska at Omaha,
Omaha, Nebraska, U.S.A., and Center for Public Administration,
Sun Yat-Sen University, Guangzhou, Guangdong, P.R. China*

INTRODUCTION

Public choice theory applies the logic of microeconomics to politics and public affairs and generally assumes rationality and self-interested behavior among citizens and public officials, in the choices they make by and for the public. The persons most closely associated with public choice theory are Buchanan and Tullock.^[1] Buchanan described the genesis of the theory as being a specification of the constitutional and other rules for making choices by and for the public, as well as an investigation of their consequences, in the tradition of the original designers of the U.S. Constitution. Buchanan explained:

Tullock and I considered ourselves to be simply taking the tools of economics, looking at something like the structure of American politics in the way James Madison had envisioned it. That is, it was clearly not a majoritarian democracy, which would be the parliamentary model (which was the ideal, at that time especially, of all the political scientists), rather it was a sort of a constitutional structure. We were the first to start analyzing the Constitution from an economic point of view. There were other people who analyzed particular voting rules, like majority voting, but we put that in a constitutional structure and provided an argument for choices among voting rules. We concentrated on that. So, in a sense, I considered us to be simply writing out in modern economic terms more or less Madison's framework of what he wanted to do, as opposed to anything new and different. It turned out that nobody had looked at it in that way.^[2]

Public choice theory has had an important influence on governments across the world, as several advisors to conservative governments were scholars who helped to develop this perspective or were influenced by it. As a result, it has been a controversial perspective. Public choice theory is relevant to many other areas of public affairs and political science (e.g., voting, international relations, and macroeconomic policy); however, this article focuses on the influence of this perspective on public budgeting theory and practice.

RELEVANCE TO PUBLIC BUDGETING

Public choice theory deals with many topics relevant to public budgeting but, unlike most of the study of government budgeting, takes a deductive theoretical approach to these topics. In public choice theory, budgetary actors are usually assumed to be acting in their self-interest rather than in the interests of their agency or the public. Rather than focusing on “how to” topics that are the main substance of public budgeting texts, public choice looks at the incentive structure within institutions and attempts either to predict outcomes on a solely theoretical level or to compare these predictions with empirical findings. Public choice theory usually does not take the descriptive approach that most budgeting literature does but, rather, seeks to theorize, test, and refine its theories.

Public choice denies the notion that government is a “benevolent despot” that will act to correct market failures and make distributional judgments in an even-handed way. From this perspective, one “ceases to see government as standing apart from the economy adjusting, correcting, and fine-tuning the economy. Rather, one sees it as being thoroughly enmeshed in the economy.”^[3] It is argued that this is a better way to understand why certain government policies exist, such as price controls, price supports, tax preferences, government-sponsored monopoly, and transfers to business and the non-poor.

Despite many differences, public choice and the study of budgeting have much in common. Forrester pointed out that both perspectives have similar interests in the choices that governments make. That is, both focus on the actions of budgetary actors, both have similar definitions of bureaucratic success (the rate of budgetary increase), both accept the value of economical government, and both accept the argument that administration and politics are not separate.^[4]

Early writings in public choice tended to come from the conservative side of the political spectrum and often clashed with mainstream public administration. Many public servants and their teachers had a visceral reaction against public choice writings about “greedy bureaucrats,” the public choice rejection of Wilsonian principles of public administration, and

the rejection of the notion of the “public interest.”^[5] As a result, there was initially a deep rift between the perspectives. However, since then, many public choice theorists have emerged, who are less ideological, and many public administration scholars have come to see some value in public choice models. In particular, because public choice has launched many empirical studies that have carefully tested propositions from public choice theory, as well as public administration, there have been some productive attempts to bridge the gap and to see how the two perspectives can strengthen each other.

GOVERNMENT SUPPLY AND AGENDA MANIPULATION

Perhaps, the best-known public choice model in public administration is Niskanen's model of public good supply.^[6] Niskanen hypothesized that agency heads will seek to maximize their budget. In doing so, they are constrained by the fact that they must deliver a certain level of services to their political sponsor. A conflict arises where the sponsor, unlike the agency head, is unable to observe the “production process,” whereby inputs are turned into outputs. All sponsors know is that for x dollars, y output is produced. A self-interested agency head, Niskanen argued, would exploit this information advantage and supply more output than demanded. This, in turn, increases the agency's budget, which is in the interests of the agency head, but not those of the sponsor (nor, presumably, the public). This is wasteful because the level of output provided is higher than the efficient level, so that marginal social costs are greater than marginal social benefits. The agency head in a sense is empire building, by controlling the information the political sponsor receives. In a later piece, Niskanen modified his model to assume that agency heads maximize their utility, rather than their budget, and that utility is based on the discretionary budget and the output of their bureau.^[7] This adds another possible outcome that the output level is efficient but that it is done inefficiently.

Niskanen's model generated a vast array of work, either testing his hypotheses or questioning the logic of his model. For the more practical purposes of public budgeting, it has had three important contributions. First, it was the first explicit model of government supply, which enabled a more complete modeling of public goods demand and production. Second, it demonstrated the importance of agenda control. For instance, in some cases, budgets have to be approved by referendum rather than by a vote of the elective body. Theories such as Niskanen's can be developed and tested to see if budgetary outcomes are different when the process of approval is different. Third, it was one contribution that helped to provide a theoretical

justification for the move to limit government size. In the political arena, this contributed to the success of many of the tax and spending restrictions in U.S. states, and also in the central governments of the United States during the Reagan Administration and the United Kingdom during the Thatcher administration. Indeed, Niskanen himself was on Reagan's Council of Economic Advisors.

CONSTITUTIONALISM

Along with Niskanen, the work of Buchanan and his coauthors provided the intellectual foundation of the movement to increase the constitutional restrictions on government fiscal practice. This work is of fundamental importance to understanding movements, such as Proposition 13 in California, and similar limitations on democratic governments. Brennan and Buchanan made the standard economic assumptions about humans—that they are rational and self-interested—and asked what constitutional arrangement an actor will agree to before the establishment of the government.^[8]

Brennan and Buchanan argued that people will fear that any majority will act as a tyrant and use the coercive powers of the state to fund inefficient projects to the benefit of the majority or, simply, to use the tax and transfer system to effectively steal from the minority. Although the actor might be able to be part of the majority and therefore benefit from this redistribution, they might not, and, if they are risk averse, they will choose a safer arrangement. Because of this concern, they argued that the rational actor choosing the fiscal rules of the constitution will limit both the constitutional procedures for considering policy and the allowable range of fiscal outcomes to limit the revenue-raising potential of the state. These limitations form the basis for a fiscal constitution that may narrowly constrain the actions of the government. The constitutional procedures they suggest are supermajority votes, multiple levels of approval, clearly defined property rights with application of due process in the protection of property, uniformity in taxation, and constitutional restrictions to balance the budget. The fiscal outcome restrictions suggested are limitations on tax rates, constraints on the breadth of tax bases, earmarking of revenues for specific expenditures, and restrictions on debt issuance and money creation.

Some of these constraints have been in place since the Magna Carta or the founding of the U.S. republic, whereas others are hotly debated topics in seats of government throughout the world. If nothing else, this public choice model forces us to rethink the fiscal rules of the game for any government. Not only is Buchanan's perspective relevant in these important

current debates, but it also challenges the fundamental approach of contemporary public finance analysis and many of its most basic policy recommendations. For example, it is typically recommended that, in order to serve the goals of pro-efficiency and pro-equity, tax bases should be broad, with few “loopholes.” This public choice perspective argues the opposite, as broad tax bases only make it easier for government to increase its revenue. In the words of Stevens, rather than assuming that the state is a “good fairy,” this approach sees the state as the “wicked witch,” seeking to increase its power, with little regard for the welfare of the citizens.^[9] Although this approach rubs many public administration scholars the wrong way, it is congruent with the views of some American citizens, and probably a roughly accurate description of the practices of some despotic governments, both in the history and in the current day.

SIZE AND GROWTH OF GOVERNMENT

Since the early 1900s, government has grown in most nations, and sometimes at a high rate. What explains this? Several theories have been examined. In 1977, Buchanan looked at two views of government—“government by the people” and “government against the people”—and summarized the findings at that time.^[10] The first model, a benign theory of the state, hypothesized that this growth was attributable to factors such as population growth, urbanization, the growth of the service economy, the income elasticity of demand for public goods, and Baumol’s “productivity paradox.” He wrote that one-half to two-thirds of the growth in government could be attributed to these factors. The unexplained gap suggested that a more pernicious model of government may explain the trends. His “government against the people” view suggested that other factors might explain part of this growth: the voting of bureaucrats for larger government; the profit-seeking activities of elected officials; and “fiscal illusion,” the phenomenon that the full cost of government services is hidden or not understood by voters. Buchanan concluded that the motivational structure of government is probably the cause of this unexplained growth in government.

Later, Mueller reviewed several research articles speaking to this question. The range of factors is suggestive of the breadth of scholarship in this area. They include both good and bad reasons for government growth, specifically urbanization, income growth, the price elasticity of demand of public goods, the expansion of suffrage over time, changes in the distribution of income, voting of public employees and welfare recipients, interest groups, bureaucracy, and fiscal illusion. While it is difficult to generalize, Mueller

concludes, “all of the institutions of democracy—electoral rules, institutions of direct democracy, and federalist institutions—seem to be important in determining the size of government. These institutions differ across countries, and their efficacy changes over time as the economic and political environment of a country change.”^[11]

Some public choice theorists have concluded from this research that there is a good case for fiscal limitations of the sort that Buchanan suggested. These limitations may be constitutional, procedural, or statutory limits on government growth or tax rates. Recent reforms in the United States such as a proposed constitutional amendment to require a balanced federal budget; the statutory budget rules passed by Congress in the 1980s and 1990s; state tax, expenditure, and debt limits on local governments; and state restrictions on local revenue sources, in part, speak to these concerns. Although the connection between these institutional rules and the size and growth of government is not always as clear and direct as proponents may argue, these institutional changes have been an important part of recent conservative political reforms.

RENT SEEKING

Economic rents are surpluses that cannot be competed away, for example, the return from an asset that is unique and in high demand, such as a parcel of property with a favorable location. Rent seeking is “attempting to use the political process to allow a firm or group of firms to earn economic rents in excess of their opportunity costs,” or one firm or a group of firms acting to obtain special treatment by the government at the expense of others.^[12] Although firms might compete in the economic market for sales and profits, they may also compete in the political market for restrictions that enable them to receive rents. They might do so in legal ways, such as lobbying or campaign contributions, or illegal ways, such as bribery, intimidation, or kickbacks. There are numerous examples of market restrictions that might lead to rents: regulatory protection, patents, zoning, exclusive licenses, occupational licensing, and academic tenure. Not only might these uncompetitive practices lead to economic inefficiencies, but also the competition among businesses and individuals for these rights would be wasteful. A casual look at lobbying practices in most governments suggests that this phenomenon is an issue. Evidence from research on campaign contributions is supportive of this assertion.^[13]

Government budgets are susceptible to rent-seeking behavior, on the part of officials and interest groups. Constitutional economists would urge that the scope of action allowed to the state should be restricted,

so there will be little opportunity to extract rent by use of the apparatus of government. This might be done through requiring supermajorities for passage of such legislation or, simply, by abolition of regulatory bodies that might be prone to this behavior. Others who may not be true believers in public choice theory might agree for the need for structural defenses of the public treasury against rent seekers. Many public administration scholars suggest that a strong code of ethics can accomplish the same result. It is hard to generalize about which solution is best, but it is worth noting that both the public choice and the public administration approaches to the study of government recognize this problem.

CONCLUSION

Public choice theory has had an important influence on both practice and thinking about government budgets. Its validity in explaining behavior within agencies and patterns of change in the public sector is open to question; however, public choice advocates are energetic examiners of their theoretical speculations, so inquiry in this area continues to advance the ideas. Public choice has generally been seen in opposition to orthodox public administration, although, as Forrester pointed out, this need not be. In fact, the public choice view of government sees numerous actors with varying motives, in competition with each other, some scrupulous and some not. This is more congruent with what we know about politics than the "benevolent despot" view of the state. Both for polemic and for scholarly reasons, practitioners and scholars in public administration need to understand public choice theory and follow its development.

REFERENCES

1. Buchanan, J.M.; Tullock, G. *The Calculus of Consent: Logical Foundations of Constitutional Democracy*; The University of Michigan Press: Ann Arbor, MI, 1962.
2. Woodrow "James Buchanan" (Interview). *Fedgazette (Federal Reserve Bank of Minneapolis)* (1995), <http://woodrow.mpls.frb.fed.us/pubs/region/int959.html> (accessed January 16, 2002).
3. Levy, J.M. *Essential Microeconomics for Public Policy Analysis*; Praeger Publishers: Westport, CT, 1995; 95 pp.
4. Forrester, J.P. Public choice theory and public budgeting: implications for the greedy bureaucrat. In *Evolving Theories of Public Budgeting*; Bartle, J.R., Ed.; Elsevier Science Ltd.: Oxford, UK, 2001; 101–124.
5. Golembiewski, R.T. *Public Administration as a Developing Discipline*; Marcel Dekker Inc.: New York, 1977; 157–201.
6. Niskanen, W.A. *Bureaucracy and Representative Government*; Aldine-Atherton: Chicago, 1971.
7. Niskanen, W.A. Bureaucrats and politicians. *J. Law Econ.* **1975**, *18*, 617–643.
8. Brennan, G.; Buchanan, J.M. *The Power to Tax: Analytical Foundations of a Fiscal Constitution*; Cambridge University Press: Cambridge, 1980.
9. Stevens, J.B. *The Economics of Collective Choice*; Westview Press: Boulder, CO, 1993; 174 pp.
10. Buchanan, J.M. Why does government grow. In *Budgets and Bureaucrats: The Sources of Government Growth*; Borcharding, T.E., Ed.; Duke University Press: Durham, NC, 1977; 3–18.
11. Mueller, D.C. *Public Choice III*; Cambridge University Press: Cambridge, 2003; 532–533.
12. Stevens, J.B. *The Economics of Collective Choice*; Westview Press: Boulder, CO, 1993; 187 pp.
13. Mueller, D.C. *Public Choice III*; Cambridge University Press: Cambridge, 2003; 354–355.

Budgeting, Line-Item and Object-of-Expenditure Controls

Suzanne Leland

*Department of Political Science, University of North Carolina at Charlotte,
Charlotte, North Carolina, U.S.A.*

INTRODUCTION

A line item budget is a traditional budget format that uses object classifications by broad categories to present proposed expenditures. Primarily used for the purpose of financial control, the categories in a line item budget consist of lists of expenditures that are logical, and contain items such as personnel, travel, equipment, or supplies. A line item budget provides simple, but limited information beyond the cost of specific categories of expenditures or appropriations in order to ensure accuracy and control. In a household budget, an example of a line item would be the amount of household income spent on clothing, but the list of expenditures would not detail the particular items purchased for each family member.

LINE ITEM BUDGET FORMATS

When governments solely rely on the line item format, the budgetary process is said to be best characterized as “incremental” because it typically calls for the last completed fiscal year, the current year, and the forthcoming fiscal year, including funds, organizational units, and objects of expenditure.^[1] Although budget battles appear to be controversial matters, the outcome is seldom a little more than the prior year’s budget using this process. This often leads to the gradual increase in expenditures, and can lead to mounting debt and deficits overtime. Despite criticisms of the incremental nature of budgeting via the line item process, this method of budgeting does provide a sense of stability from year to year. The stability of the line item budgetary process is valued by lawmakers and administrators because they often rely on budgetary projections for political decision making.^[2]

According to Wildavsky’s theory of incrementalism, budgetary decision making is driven by department heads.^[1] Therefore, each year only the increments from an already existing and well-defined base are examined by policy makers. From this point of view, governmental budgeting is considered to be an annual repeated process that results in the bargaining between individuals who know one another through budgetary activities. This theory most adequately describes what

happens when a government uses a line item budget, but does not accurately describe other methods of budgeting used in government. For example, the theory of incrementalism could not take into account or explain entitlement programs such as social security, which over time squeezed out other categories of expenditures in the federal budget. Nor could Wildavsky’s theory of incrementalism account for top-down budgeting processes when bureaus and departments were bypassed or when cutbacks took place.^[3]

OBJECT OF EXPENDITURES

Technically, a budget following a line item format is a detailed financial plan developed according to governmental accounting based on object of expenditure classifications. Objects of expenditures (or objects of expense) are detailed categories of goods and services that are purchased by an organization, such as supplies, travel, equipment, or personnel services. Each object-of-expenditure classification is a form of accounting that attempts to limit expenditures by the organization. Normally, objects of expenditures are divided by departments or divisions within the organization that serve as appropriation guides to personnel.

USE OF LINE ITEM BUDGETS

The line item budget format is especially useful for purposes of financial control and is the simplest budget format to link to governmental accounting systems. The format of a line item budget is straightforward, readable to citizens, and can easily be converted to accounting systems. For these reasons, this form of budgeting is now common across public and non-profit organizations in the United States.

The first widespread use line item budget in the United States came about as part of the reform movement that was aimed at cleaning up municipal corruption in the early part of the twentieth century and the move to make government more efficient like business. The focus of the reform movement was on gaining financial control, and the budget reform preferred was the line item format.^[4]

For local governments, line item budgeting is closely associated with the introduction of city manager forms of government. The reformers at this point in time were keen on introducing “efficiency” measures into government administration, and accountability to the public for how and why taxes were spent.^[4]

Due to the advantages of the line item format, approximately 51% of U.S. local governments still use some form of line item budgets.^[5] Governments can incorporate the line item budgeting format in one of two ways. First, a government may use the line item format primarily to present their budgets to elected officials and the public. Second, and more common, the local government uses this format as a supporting document or to supplement to another format. Typically, in the case of the latter, the more comprehensive the budget format, the more varied and numerous the formats and systems for classifying data.

At the federal level, line item budgeting is associated with the first wave of budget reforms in the United States. The move to centralize the federal budgeting process was culminated in the passage of the Budget and Accounting Act of 1921. This law established the Bureau of the Budget (BOB) as the central agency in charge of the federal budgetary process. Along with the BOB, the act created the congressional Governmental Accounting Office to scrutinize federal expenditures. Gradually, after the passage of this act through the 1930s, the federal budgetary process became more detailed. Categories were narrowly defined so that expenditures were attached to line items to link expenditures to specific programs.

It is important to note that while line item budgets do not preclude planning or the aim of achieving objectives, they are primarily concerned with controlling and confirming the budgetary process to set aggregate spending limits. A well-designed budget format should provide a general statement of financial status, inform stakeholders of the cost of achieving agency goals, simplify operational procedures with well-classified expenditures, and serve as a means to evaluate objectives. The traditional line item format alone does not adequately meet these requirements and works best when used by governments in a supplemental manner to other budgetary formats such as performance-based budgeting.

REFERENCES

1. Wildavsky, A. *The Politics of the Budgetary Process*; Little Brown: Boston, 1964.
2. Pitsvada, B.; Draper, F. Making sense of the federal budget the old fashioned way—Incrementally. *Public Adm. Rev.* **1984**, *44* (5), 401–407.
3. Rubin, I. Aaron Wildavsky and the demise of incrementalism. *Public Adm. Rev.* **1989**, *49*, 78–81 (January/February).
4. Rubin, I.; Stein, L. Budget reform in St. Louis: Why does budgeting change? *Public Adm. Rev.* **1990**, *50*, 420–427 (July/August).
5. Lee, R.; Johnson, R. *Public Budgeting Systems*, 6th Ed.; Aspen Publications: Gaithersburg, MD, 1998; 120 pp.

Budgeting, Performance-Based

Julia E. Melkers

*College of Urban Planning and Public Affairs, University of Illinois at Chicago,
Chicago, Illinois, U.S.A.*

INTRODUCTION

Using performance data in the budgetary process means integrating information about outcomes and impacts in decisions about the allocation of funds where the goal is to be able to allocate financial resources using performance information to make a more informed decision. Performance measures are generally established at the agency level, oftentimes with input from executive and legislative budget offices. Ideally, performance measures are developed for both internal use by agency staff as well as use for external reporting, including use in the budget process. Improved accountability is the primary rationale for the push to performance-based budgeting at all levels of government. An important thrust of current budget reform efforts is to develop structures that improve communication both within and between government and citizens. This activity is referred to as performance-based budgeting (PBB).

OVERVIEW

Simply put, PBB is a recently revived phenomena in budget reform that involves the integration of performance measures and data representing agency activities in the budget decision-making process. Although it seems innate that aspects of performance will be included in decisions about the allocation of resources, this is often not the case. In recent years, however, governments and non-profit organizations are moving away from a line item approach and adopting performance budgets, where a link is established between allocated funds and what an organization will achieve with those funds. Line item budgets encourage decision-making processes that focus on individual items in a short-term framework with little or no explanatory information. Performance budgeting efforts take a broader approach, encouraging a long-term perspective with an emphasis on what will be achieved in the future.^[1]

TERMINOLOGY

With the trend toward the use of performance measures in budgetary processes, PBB has become increasingly

popular in this era of budget reform and government reinvention. However, what may be described as PBB, is actually a more diverse set of requirements, blending various aspects of current public management trends, including outcome measurement, performance measurement systems, strategic planning, and benchmarking. Although PBB may be the most common term to refer to this process, the actual terminology varies a great deal as organizations and governments design the process and “make it their own.” In fact, these variations are reflected in the performance-based requirements used in different governments and organizations. Some state and local governments refer explicitly to PBB, “performance-budgeting,” “results-based budgeting,” or “outcome-based budgeting system.” Others, however, are more descriptive, indicating that certain performance-related elements be included in the budgetary process. Here, they may use terms such as “program performance reports,” “performance measures,” “outcome measures,” and “performance standards and measures.” These differences, however, are simply a matter of language. The essence of these different requirements does not vary—each indicates that performance measures be actively incorporated in the budgetary process. Specifically, these performance budgeting initiatives involve:

- Identification of broad performance measures for selected organizations (programs/departments/agencies), with an emphasis on outcomes.
- Linking of outcome measures to organizational activities and their respective budgets.
- Targeting performance levels and budget levels together.

LEGISLATIVE AND ADMINISTRATIVE REQUIREMENTS

The use of performance data in budget processes and budget decision making is increasingly becoming the norm in public and non-profit organizations. Governments are not only *encouraging*, but *requiring* the integration of performance data in the budgetary process. In 1993, the Government Performance and Results Act (GPRA) was passed, requiring the establishment and use of performance measures in federal agencies.

One of the stated purposes of the act is to “improve congressional decision-making by providing more objective information on achieving statutory objectives, and on the relative effectiveness and efficiency of Federal programs and spending.” More specifically, as noted in Section 1115 of the act, the role of performance measures in the budget process are stated as follows:

In carrying out the provisions of section 1105(a)(29), the Director of the Office of Management and Budget shall require each agency to prepare an annual performance plan covering each program activity set forth in the budget of such agency. Such plan shall—

1. establish performance goals to define the level of performance to be achieved by a program activity;
2. express such goals in an objective, quantifiable, and measurable form unless authorized to be in an alternative form under subsection (b);
3. briefly describe the operational processes, skills and technology, and the human, capital, information, or other resources required to meet the performance goals;
4. establish performance indicators to be used in measuring or assessing the relevant outputs, service levels, and outcomes of each program activity;
5. provide a basis for comparing actual program results with the established performance goals; and
6. describe the means to be used to verify and validate measured values.

The tone set by the GPRA is reflected in many similar state and local PBB requirements. At the state level, all but three states (*Arkansas, Massachusetts, and New York*) have legislative or administrative requirements for the integration of performance data in the budget process, with most of these requirements established in the 1990s (Melkers and Willoughby, 1998).^[5] One state (Missouri) established PBB by executive order. States that are best known for the development and use of performance measures in the budget process include Texas, Oregon, Minnesota, and Florida. A similar trend toward the use of performance measurement information in the budget process is taking place in local governments, with many municipal and county governments requiring performance measurement with an intent to use it in the budgetary process.

In earlier budget reforms, integrating performance information in the budget process was viewed as a linear process. Research shows, however, that decision makers consider a range of factors in budgetary decisions.^[2,3] Today's reform, however, acknowledges the complexity of not only public programs, but also

of the budget process itself.^[4,5] For example, in the state of Washington, the budget guidelines are clear that performance data inform, but do not drive, budgetary decision making. Instead, the actual budget or policy decision involves performance measures, but is also affected by a variety of other factors, such as financial realities, stakeholder requirements, competing priorities, public sentiment, organizational capacity, and public sentiment.

SELECTING APPROPRIATE PERFORMANCE MEASURES FOR THE BUDGET PROCESS

The Government Finance Officers Association (GFOA) recommends a useful list of characteristics for performance measures that are used in budgetary processes. They suggest that financial, service, and program performance measures be developed and used as an important component of decision making and incorporated into governmental budgeting. Specifically, they state that performance measures should:^[6]

- Be based on program goals and objectives that tie to a statement of program mission or purpose.
- Measure program results or accomplishments.
- Provide for comparisons over time.
- Measure efficiency and effectiveness.
- Be reliable, verifiable, and understandable.
- Be reported internally and externally.
- Be monitored and used in decision-making processes.
- Be limited to a number and degree of complexity that can provide an efficient and meaningful way to assess the effectiveness and efficiency of key programs.

In an ideal performance budgeting process, output measures are included to show the level of work accomplished, but efficiency and effectiveness measures are most relevant and may be most informative in the budgeting process where decision makers are looking for links to actual funds. Taking output, outcome, and efficiency measures together gives policy makers a full view of activities completed, the cost and value of the outputs and outcomes, and what has actually been accomplished with the actual expenditures. In the current state of government reform, helping government to become more efficient is an important focus. In this same vein, as performance measures are introduced in the budgetary process, they are important for identifying the relative efficiency of services. Efficiency measures are important in the budget process because they help to answer the question, “How much output is there for a given input?” This type of information is especially important in making

fiscal decisions because they provide information on the cost of providing services and when linked to output measures, provide information on the cost of obtaining given outputs.

The rationale behind performance budgeting processes is that better resource allocation decisions will occur with the use of information that indicates how well an entity is performing. For many governments, this performance is linked to agency or program goals and objectives. Therefore, performance budgeting should indicate how well an agency or program is meeting those goals and objectives. It is imperative that performance measures in the budgeting processes are not limited to measures of outputs, but that include actual outcomes.

Although it seems straightforward to select certain categories of measures most useful for budgetary decision making, it is not a uncomplicated process. Along with challenges relating to the identification and selection of performance measures, there are three important challenges in the selection of measures for budget processes:

- Nature of the budget cycle.
- Perspective and attention span of policy makers.
- Pressure to demonstrate macroresults.

Typically, most budget cycles are on an annual basis, with some governments adopting a biennial budget. Because of this relatively short time frame, integrating performance measures in budget process means selecting measures that can show some change within that timeframe. Measures cannot be so global to not be meaningful within the budgetary calendar. One problem with measuring outcomes within the budgeting cycle is that the outcomes of most government and non-profit programs and services occur in a much longer timeframe. This is perhaps the greatest challenge in integrating performance measures in the budgetary process. Because of this challenge, it is important to identify interim or intermediate outcomes that show progress toward or a contribution to the ultimate projected outcome. In addition, performance targets or goals may also be shown to provide a context for the performance information.

Another important challenge is that policy makers and others involved in the budget process typically have to review budgets and measures for numerous programs and departments. Further, these individuals will have varying levels of familiarity with the individual budgetary entities. For this reason, it is important to include only key measures in the budget process. More detailed measures are more appropriate for the departmental level. For example, in Florida, the guidelines for their performance budgeting initiative, PB² (performance-based program budgeting), note

that measures of broad program results appear in the budget documents, while detailed output and outcome measures are maintained and used at the agency level. Similarly, in Washington State, the guidelines stress that key, summary measures appear in the budget documents, but must also link to the agency strategic plan. Further, they note that the focus should be on developing a core set of useful and verifiable indicators, rather than on a large quantity of indicators.

In addition, there may be pressure from policy makers to identify measures that reflect macrolevel policy goals in a government, such as “lower infant mortality rate” or “healthier society.” However, it is important that performance measures in the budget should be tied to activities and results that the agency can affect. It is important then that interim measures be selected that show progress toward those larger policy goals where performance may be linked to agency strategy and goals.

As performance measures are integrated in the budget process, guidelines for their selection, reporting format, and use are important to ensure consistency across government, but also to inform stakeholders about the changes and related requirements associated with the new process. In some cases, such as in the state of Florida, there are elaborate review procedures of the measures themselves to ensure consistency and quality of measures across agencies. In most cases where performance measures are required in the budget process, general guidelines are provided indicating the types of measures that are to be included.

IMPLEMENTING PERFORMANCE BUDGETING

Implementing performance budgeting means changing the way that agencies submit budget requests as well as changing the way that policy makers make decisions.^[7] The formalization of PBB requirements provides a framework or process so that performance data are considered in the budgetary process. For many systems, this requires a complex reworking of existing processes.

Because most governments have not been collecting performance data as a regular activity, implementing performance budgeting means not only refining the budget process, but also developing a parallel or concurrent process where performance measures are identified, and data collected and reported to fit with the existing budget cycle. Thus, implementing performance budgeting often means initiating performance measurement activities in the first place. Many governments that have adopted performance budgeting processes have done so in a multistage process, where different programs or agencies are selected to pilot the use of performance measures in the budget process

before the process is implemented governmentwide.^[5] Adopting a performance budgeting system in this way is important because it allows for learning, adaptation, and integration with existing systems. For example, the state of Florida implemented its PB² initiative over several years, adding new departments and programs each budget cycle. An important advantage of a pilot process is that, as new agencies or departments or programs begin performance budgeting, they are able to look at and communicate with other entities within their government as examples.

An important fiscal issue arises in this implementation process. If performance measures are meant to be used in budgetary decision making, what happens when programs or agencies do not meet their targets? For state governments, a handful provide explicit guidelines for actual agency attainment, or non-attainment of goals and objectives as part of their legislated performance budgeting requirements. These are guidelines that offer rewards to public managers to strive for and attain the performance goals that they have identified in their strategic planning or other process. Research has summarized and compiled the different variations that guidelines have taken in the states.^[8,9]

CONCLUSION

Overall, there is little evidence to show that performance data are affecting actual budget appropriations. This is not surprising or disturbing because the implementation of performance budgeting is still in the early phases for many governments. However, integrating performance data in the budget process can help decision makers to make better, more informed decisions along the way.^[10] Specifically, performance data are useful for helping decision makers in a number of areas:

- Understanding the activities and objectives of funded programs by viewing summary measures of performance.

- Understanding changes in performance over time compared with budgetary changes.
- Having more meaningful dialogs with public managers about agency activities, goals, and performance.
- Identifying poorly performing and high-performing programs and departments.
- Justifying fiscal decisions using evidence rather than anecdotes or impressions.

REFERENCES

1. Miller, G.J., Hildeth, W.B., Rabin, J., Eds.; *Performance Based Budgeting*; ASPA Classics Volume, Westview Press: Boulder, CO, 2001.
2. Grizzle, G. Does budget format really govern the actions of budget makers? *Public Budg. Finance* **1986**, 6 (1), 60–70.
3. Stanford, K.A. State budget deliberations: do legislators have a strategy? *Public Administration Review* **1992**, 52 (1), 16–26.
4. Joyce, P.G. Using performance measures for budgeting: a new beat, or is it the same old tune? In *New Directions for Evaluation*; Jossey-Bass Publishers: San Francisco, 1997; Vol. 75.
5. Melkers, J.; Willoughby, K. The state of the states: performance-based budgeting in 47 out of 50. *Public Administration Review* **1998**, 58 (1), 66–73.
6. Tigue, P.; Strachota, D. *The Use of Performance Measures in City and County Budgets*; Government Finance Officers Association (GFOA): Chicago, IL, 1994.
7. Forrester, J.P.; Adams, G.B. Budgetary reform through organizational learning: toward an organizational theory of budgeting. *Adm. Soc.* **1997**, 28 (4), 466–488.
8. Melkers, J.; Willoughby, K. Budgeters views of state performance budgeting systems: distinctions across branches. *Public Administration Review* **2001**, 61 (1), 52–62.
9. Willoughby, K.; Melkers, J. Implementing PBB: conflicting views of success. *Public Budg. Finance* **2000**, 20 (1), 105–120.
10. Broom, C.A.; McGuire, L.A. Performance-based government models: building a track record. *Public Budg. Finance* **1995**, 15 (4), 3–17.

Budgeting, Target-Based

B. J. Reed

*College of Public Affairs and Community Service, University of Nebraska,
Omaha, Nebraska, U.S.A.*

INTRODUCTION

Target-based budgeting (TBB) has become an increasingly common approach to the budgeting, particularly for local governments. It is a method of budgeting that identifies overall expenditures for an agency or unit based on the estimated revenue available to government. It allows control to be imposed at the executive or legislative levels while allowing decentralization and flexibility at lower levels of the organization. TBB is particularly useful when state and local governments have needs to control spending but also face turbulent and unpredictable demands for services.

WHAT IS TBB?

In most cases, the expenditure “target” for each unit of the organization under TBB is determined by the budgeting or finance office under the direction of the chief administrative officer and the policy board for that organization. This target is based on the strategic priorities or direction of the agency as well as the estimated availability of revenue.^[1] These expenditure ceilings are set well in advance of departments preparing their budgets and are not negotiable. Usually, a percentage of the total budget is set aside for contingencies as well as for discretionary spending. The discretionary spending allows for organizationwide priorities and may include departmental requests beyond those funded within the targets set for each unit.

Targets can be set a few different ways. One is to simply divide proportionately historical expenditures by department. For example, if public safety historically has taken 35% of a city’s expenditures, their target would be 35% of the revenue available after contingencies, discretionary allocations, and one-time expenditures have been determined. Rubin referred to this as the “constant services or maintenance of effort budget.”^[2] Another approach would be to combine historical spending patterns with strategic priorities of the government. If the elected officials determined that neighborhood development was to become a key focus for the city, targets for departments involved in neighborhood development could be adjusted upward.

However, in most cases, such strategic emphases would be included within the discretionary portion of the budget rather than in adjusting the targets for each agency.

Target budgeting is designed to centralize overall expenditure patterns among departments, while maximizing discretion within each agency about how to spend those dollars. Although some service deliveries within departments are mandated by law, other activities are discretionary. Units can determine within the target provided what they believe to be the most important, and what personnel and operating expenditures are necessary to carry out these services and activities.

If the target provided is above the historical dollars provided to the agency, discretion increases. However, if targets are at or below previous targets, discretion is reduced. Departments with high mandated activities may be at a competitive disadvantage relative to units with fewer mandated service levels. County health departments may be mandated by the state to carry out inspections of restaurants, so some expenditures will have to be dedicated to this activity within the target set by the county commissioners. Similarly, departments with access to external funding through fees, grants, or donations may be able to shift expenditures covered by such revenue sources rather than having them covered by tax-based revenue set as part of the budget process. Also, because some tax revenue is earmarked for particular purposes, some departmental targets may be adjusted to reflect trends in revenue generation in these areas. County road departments may have revenue generated through gasoline taxes because such revenue has to be used for the construction or reconstruction of roads, while county health departments may receive revenue generated through inspection fees. Such earmarking must be taken into consideration as the chief executive and budget office develop targets for each department.

Linking targets to revenue forecasts is a tricky business, but is essential to making TBB work. Revenue forecasts must be set early enough to ensure time for target expenditures to be set and departmental responses to be developed. Forecasting property tax receipts may be quite easy because past patterns are good predictors of future revenue. Predicting sales

or income tax is much more difficult because they are based on underlying economic conditions and consumer confidence rather than on mandatory fee payments. White suggested that forecasts be set at least 6 to 8 months in advance.^[1]

Target-based budgeting is different from lump sum budgeting (LSB), which assumes no line item discipline within the agency or departmental budget. It also does not necessarily involve setting target ceilings for expenditures based on forecasted revenue. Target-based budgeting can be combined with LSB to create only a ceiling on expenditures. In this case the department has total discretion to shift funds between personnel and operating line items as long as they remain within the target ceiling. Target-based budgeting can also be linked to specific line items. In such cases, permission would be required for shifts among personnel or operating categories once the target budget has been established.

Target-based budgeting is also different from zero base budgeting (ZBB), which requires extensive development of budget scenarios within each department through the development of “packages” of services and the costs associated with each of those packages. Target-based budgeting is much less paper and time intensive because it only requires departments to provide overall requests as they might typically do in more traditional budget processes as long as they do not exceed the ceiling.

The amount set aside for contingencies or discretionary spending is determined by the chief administrator and the policy board, and is likely to range from 10% to 15% of the overall revenue projected for the agency. Agencies can also pass down the contingency estimates to each individual department but this will require targets to be adjusted for expenditure predictability. Public works departments, for example, may require greater contingencies for such things as salt, sand, and overtime due to inclement weather than would be true in a department such as human resources or planning. The amount set aside for discretionary spending increases the ability of policy

boards to set priorities but it also decreases the flexibility and discretion at the unit level. Policy makers must be willing to allow 85% to 90% of the budget to be decentralized to individual units. However, the remaining 10% to 15% can be ranked to meet overall organizational priorities. Table 1 provides an example developed by Wenz and Nolan about how the process would work.^[3]

HISTORY AND USE

According to Rubin,^[4] the first mention of TBB was a case study of Berkeley, California, in the 1920s. Although the exact term was not used, the process used in Berkeley was similar to what we think of as target budgeting. In this case, each department was provided a “maximum amount” to spend during the fiscal year, while providing supplemental requests in priority order to the city manager.

Some form of TBB has been practiced since the 1920s. The emergence of ZBB in the 1970s in Georgia, and later at the federal level, provided a new twist on target budgeting. However, the enormous complexity and paper work requirements of ZBB made it fall from favor even when TBB was continuing to be used throughout the United States.

As noted earlier, most documented uses of TBB have been in city and county governments. Cincinnati, Tampa, Rochester, Phoenix, Salina, Kansas, and Douglas County, Nebraska, have all used the target budgeting approach. In Cincinnati, the city council developed decision criteria to allocate discretionary funds centrally once the target budgets had been established. The criteria included such things as maintenance of the city’s infrastructure, legal mandates, agency capacity, employment potential for the private sector, environmental impact, preventive maintenance, and support for long-range plans.^[5]

Rochester developed a target for each unit, but subtracted all one-time costs that occurred in the previous fiscal year and the addition of new one-time costs for

Table 1 Calculation of targets

Total operating fund budget	\$151,840,000
Recreation department budget	7,956,740
Proportion of total operating fund budget	$(7,956,740)/(151,840,000) = 0.0524$
Total projected operating fund budget resources	153,850,000
Less 10% discretionary reserve	15,340,000
Net projected total operating fund resources ^a	138,510,000
Recreation department target	$0.0524 \times \$138,510,000 = \$7,258,200$

^aThe projected \$138 million did not involve certain expenses, such as fringe benefits, that were not included in the target.

(Source: From Ref.^[3].)

the coming fiscal year. Phoenix along with many of the other cities used some form of performance measures either at the unit level or individual level in conjunction with TBB. Outcome measurement is seen as an important linkage to target budgeting.^[6] Salina, Kansas, closely tied its TBB efforts to strategic planning with the city council developing performance goals that help to direct strategic priorities for the city.^[11]

STRENGTHS

Target-based budgeting has several benefits. First, target budgeting can be carried out without extensive burdens on departments. It requires some analysis, but not the extensive kinds of detail that other budget approaches require. Second, it can provide constructive policy involvement by elected officials, but leave considerable discretion to unit managers. The target budgeting approach allows prioritization of a percentage of the budget at the policy level, while delegating decisions for most of the funding to the unit level. This mix of centralized and decentralized decision making often creates a healthy balance in the budget process. Third, by having the targets set by revenue estimates, TBB avoids much of the expenditure haggling that occurs with other budget approaches. Departments know the parameters they will be working within and this predictability can improve management decisions throughout the fiscal year. Finally, the ability to identify discretionary funds that can be used to reallocate toward priorities of the organization allows chief executives and policy boards the ability to have substantial affects on meeting agency goals.

WEAKNESSES

Target-based budgeting must have the total support of the policy board and the chief executive. Pressure will exist to vary the targets set, and it is likely that some departments will attempt to manipulate the process in other ways. Only strong support by organizational decision makers will make the process work. Conversely, government officials must avoid the temptation to micromanage decisions made at the departmental level. The true incentive for unit managers is the discretion to control budgetary allocations. Target budgeting without such discretion ties the hands of managers, and reduces support and incentives to innovate and produce efficiencies that benefit the entire organization.

Revenue projections must be as accurate as possible. The whole basis of TBB is that ceilings are set based on proportionate allocation of known revenue. Inaccurate forecasting can undermine the entire approach.

Although improvements can result with TBB, the process is still likely to be an incremental one. To some extent, this is determined by how the discretionary dollars are allocated across the agency. Placing substantial portions of the revenue into new initiatives can result in substantial changes. This is not common. If the culture is developed where innovation is supported, non-incremental change may be possible at all levels of the organization. However, TBB does not guarantee that such chance will occur.

Target-based budgeting works well with general fund tax dollars. It does not work as well with departments that generate substantial dollars externally through fees or external grants or donations. Public utilities are not good candidates for TBB because most of their resources are based on customer payments. In this case, these units operate through self-funded activities. Another complexity of TBB is sorting through tax revenues that are earmarked for particular activities within departments. Public works departments who receive earmarked funding of gasoline or cigarette taxes rely on different allocation and forecasting processes than those departments that depend exclusively on property, sales, or income taxes.

CASE STUDY OF TBB

The University of Nebraska (NU) system operates under a modified target budgeting approach. As a result of a state Supreme Court decision, the Nebraska Unicameral Legislature (NUL) is required to provide NU with funding in a "lump sum" based on revenue forecasts without the ability to earmark funding for specific purposes. The University of Nebraska submits line item detail to NUL in the late summer or early fall of each year. Nebraska Unicameral Legislature approves the budget in May or early June and, based on revenue forecasts, developed by a state forecasting board. The governor can approve the recommendations of NUL or line item veto elements that are oppositional. Nebraska Unicameral Legislature can then sustain or override the governor's vetoes. Once approved, the university has total discretion in how those tax dollars are spent beginning July 1 of each year.

Once NUL and the governor have acted, the elected Board of Regents in conjunction with the president of NU allocates target lump sum budgets to each of the four campuses of the system. Certain campuswide expenditure priorities are identified by the top administrators of each campus and funding is set aside for these priorities. At The University of Nebraska at Omaha, however, the vast majority of funding is allocated in the form of "target" amount to each unit on the campus. Certain assumptions are built into these target allocations. These are initial established as NU

prepares its budget for submission to the state and are revised based on final action taken. Estimates are made for personnel and operating based on existing lines and previous state-funded operating expenditures. Any increases in personnel costs or fringe benefits are added to the targeted amount for each unit.

Once this target budget is established, unit directors have complete discretion in shifting dollars between established personnel and operating line items as long as they remain within the overall budget target established at the beginning of each fiscal year. In reality, the amount of discretion is a function of changes in personnel status within the unit. If all personnel lines are filled, discretion becomes almost non-existent because most assumptions concerning operating budgets are considerably below actual need. However, if personnel lines are open and remain unfilled, discretion increases substantially.

Each unit is charged a "tax" by NU, based on legislative assumptions about the number of open lines that will exist within each campus and unit each year. These charges are passed down to each unit in proportion to their overall personnel budget, irrespective of the number or dollars associated with actual open positions. In some cases, these charges can be offset during the fiscal year by allocations made centrally for such things as equipment requests, funding for part-time instructors, and so forth. However, the allocation of such funds remains at the discretion of upper-level administrators and may or may not materialize depending on priorities that exist at a campuswide level. Unit discretion will also vary based on how many external dollars can be generated from grants, fees, and endowment support. Units with significant revenue generated from such sources have much greater discretion than units without such revenue streams.

Budget and finance offices for the campus track line item expenditures to compare with estimated expenditures, but their primary focus is the bottom line expenditure budget for each major unit. If expenditures by month in total begin to exceed estimated expenditures, they notify the manager in charge of a unit to ensure that total expenditures are in line with budgeted amounts. Overages within individual line items are generally not an area of concern because of the discretion available to each unit manager to shift funds as needed. If total unit expenditures appear to be exceeding the proposed budget, unit managers must identify additional revenue sources that will be provided to cover these overages.

CONCLUSION

Target-based budgeting is a little-recognized but extensively used approach to public budgeting. It provides both executive control and decentralized flexibility. Advantages include reduced bureaucratic reporting and procedural systems, increased policy involvement by elected officials, and reduced conflict over expenditure decisions. Problems with the approach are often associated with limited support by elected officials and chief executives, lack of accurate revenue forecasting, and extending the approach beyond general fund tax dollars to units generated substantial dollars externally.

As the previous case study demonstrates, TBB can work effectively if it is designed to accommodate the governmental structure in which it operates. As with any budgeting approach, failure is likely unless that budgeting approach fits within the organizational environment.

REFERENCES

1. White, J. *Target-Based Budgeting: A New Approach to an Old Activity*; Kansas Government Journal, August, 1994. <http://www.co.saline.ks.us/Budget/TBB.htm> (accessed June 2001).
2. Rubin, I. Target-based budgeting. In *Defining Public Administration: Selections From The International Encyclopedia of Public Policy and Administration*; Weston Press: Boulder, CO, 2000; 369 pp.
3. Wenz, T.W.; Nolan, A.P. Budgeting for the future: target based budgeting. *Public Budg. Finance* **1992**, 2 (2), 88.
4. Rubin, I. Target-based budgeting. In *Defining Public Administration: Selections From The International Encyclopedia of Public Policy and Administration*; Weston Press: Boulder, CO, 2000; 368–369.
5. Wenz, T.W.; Nolan, A.P. Budgeting for the future: target based budgeting. *Public Budg. Finance* **1992**, 2 (2), 90.
6. Rubin, I. Budgeting for our times: target base budgeting. *Public Budg. Finance* **1991**, 3 (11), 5–13.

BIBLIOGRAPHY

1. Koehler, F.U.; Reed, B.J. Target budgeting in Lincoln county. In *Public Budgeting and Financial Management*; Khan, A., Hildreth, B.W., Eds.; Kendall/Hunt: Deubuque, IA, 1994; 101–112.

Budgeting: Allocation Formula

Patricia Moore

Public Administration Department, Kean University, Union, New Jersey, U.S.A.

INTRODUCTION

Allocation by formula is one of the decision rules used in the United States by budgeters to allocate scarce resources in the public sector. Other rules include budget allocation in terms of existing base plus a fair share of the increment or decrement, top-down containment rules, and targets and relative merit.^[1] Mikesell described an allocation formula as a quantitative mathematical equation used to distribute grant funds to eligible recipients.^[2] Wildavsky suggested that allocation formulas for public-sector program expenditures evolved out of recurrent budget processes.^[3] Usually, an allocation formula is specified in legislation, but sometimes it is provided by regulation.

In the Western world, the underlying theory surrounding formula allocation seems to be a calculative logic that formulas are sensitive to the concerns of democratic government. These concerns include equity, efficiency, and effectiveness of service provision on behalf of the (voting) population. In the United States, prior to World War II, budget allocations were often based on historical precedent or negotiation; however, since that time, federal and state governments have progressively shifted in favor of distributing aid to states and localities by formulas.^[4] This shift to formula allocation may have dampened the political problems naturally arising from the annual competitive budget appropriation process in Congress at the federal and state levels of government.^[5]

A good allocation formula should be free from perverse manipulations, open to review, and should distribute resources equitably across governmental units. Moreover, it should be understandable, equitable, adequate, and predictable. It should also be reasonable in reporting requirements, fiscally accountable, and politically acceptable.^[6] Grizzle suggested that the design of an allocation formula should give primary consideration to program objectives, operational measures of objectives, data availability, and simple methods.^[1] Although formulas can be used to direct funds to individual service providers, historically, they have been directed at aggregations of service providers: state governments, hospitals, counties, municipalities, and school districts. In the United States, formulas are frequently used to determine allocation levels for various intergovernmental fund transfers, such as general

revenue sharing, block grants, and categorical grants.^[7] While formulas are used to allocate resources for many different governmental functions, they are most widely used in education, health, housing, and commerce.^[8]

WHY ARE ALLOCATION FORMULAS USED WIDELY IN BUDGETING?

Formula allocation is appealing in public budgeting, because it has several advantages. Formula allocation makes justification for budget requests easier. It also has an appearance of neutrality and rationality, because a formula allocation is expressed as a mathematical equation. Using a precise mathematical formula can make the allocation plan less controversial and, therefore, more politically acceptable. "A multitude of (redundant) statistical variables may be politically necessary to rationalize a program, to at least give the appearance of responding to special circumstance."^[9]

Furthermore, allocation formulas allow for greater precision and accuracy and thus save legislative time and energy in arguing about the intended appropriation. Allocation formulas provide local government a reasonable assurance of the general level of funding they are to expect; thus, they keep planning from being chaotic. Another advantage of formula allocation is that it suggests consistency and fairness. When the objective is equalization, allocation formulas seem to be an effective instrument for redistributing public resources. Their transparency and stability tend to reduce the possibility of manipulation, therefore, facilitating equity.

A well-designed allocation formula simplifies program administration, reduces paperwork, as well as provides state and local governments a reasonable degree of fiscal certainty. In addition, formula allocations reduce the discretion of federal administrators in the review process, allowing public administrators to insulate themselves from political pressures.

TECHNICAL NATURE OF ALLOCATION FORMULAS

The typical allocation formula has a complex structure entailing the identification of several variables. For example, the United States Department of Agriculture

allocates funds to school districts for their Breakfast Program using the following formula:

$$M = \text{NAP}_1 * \text{NPB} + \text{NAP}_2 * \text{NFB} + \text{NAP}_3 * \text{NRB} + (\text{NPB} + \text{NFB} + \text{NRB}) * 0.06 + \text{AD}$$

M is the amount of federal funds given to the school district, NAP_1 a national average payment (NAP), NPB a NAP for paid breakfasts, NAP_2 a NAP prescribed for free breakfasts, NRB the number of reduced priced breakfasts, and NFB the number of free breakfasts.

Allocations to school districts are calculated by multiplying the number of paid breakfasts served to eligible children during the fiscal year by a NAP; multiplying the number of breakfasts served free to eligible children by a NAP prescribed by the Secretary of Agriculture for free breakfasts; and multiplying the number of reduced price breakfasts served to eligible children by a NAP prescribed by the Secretary for reduced priced breakfasts. School districts also receive an additional six cents for every breakfast served. Schools with a high percentage of needy children and high breakfast costs may receive additional payments.^[10]

In adopting allocation formulas, the U.S. Congress commonly uses measures that closely represent at least one of the following characteristics of the recipient government: “need,” “capability,” and “effort.”^[11] Krepes and Slater suggested that few allocation formulas include all three measures, however, more than half include a measure of need. “Need” often refers to the population to be served, while “capability” refers to the ability of a jurisdiction to meet the defined need from local or state funds. “Effort” refers to the actual amounts of local or state revenues available for the need. One common measure of “capability” is per capita income. School districts often use per pupil property values as a measure of “need,” while per pupil expenditures are frequently used to measure “effort.”

TYPES OF ALLOCATION FORMULAS

Formulas may be demand driven or resource constrained.^[1] The main characteristic of a demand-driven formula is that allocations are driven by the need of the different units, e.g., school districts, states, or municipalities. It is mainly used to stimulate action by the recipient government as much as possible. Demand-driven formulas are more appropriate when policy makers value controlling agency behavior more than agency spending.^[12] One of the disadvantages of a demand-driven formula is its unpredictability in terms of the level of total spending that will result. Medicaid in the states and the federal government shares an example of a demand-driven formula. Total spending

depends on the number of people entitled to have their medical care paid for by the government, those who apply for help, and the basket of services for which the states reimburse health-care providers.

On the other hand, resource-constrained formulas place a limit on the resources available for a specific purpose. Each unit’s allocation is determined by its percentage of the total appropriation. An example of a resource-constrained formula is the equation used to allocate resources to the airline industry in the aftermath of the September 11, 2001, attack on the World Trade Center in New York City. According to the Air Transportation Safety and System Stabilization Act, the federal funding formula was based on each carrier’s share of the industry’s flying capacity, measured in available seat miles.^[13] Specifically, the allocation for each airline was the total aid provided by the federal government multiplied by the percent of the industry’s total flying capacity, i.e., number of seats multiplied by total miles. This formula determined what fraction of the total appropriation each airline received. When controlling spending is more important, the resource-constrained funding formula is more likely to be effective in containing spending than demand-driven formulas.^[12] The main difference between the two is that demand-driven formulas may not have funding ceilings but resource-constrained formulas do.

WEAKNESSES OF ALLOCATION FORMULAS

Allocation formulas are often incredibly complex.^[14] Frequently, they require a bewildering combination of mathematical calculations, i.e., addition, multiplication, subtraction, squaring, and even quadratic equations that all too frequently produce unsatisfactory results. Quite often, they are not clearly understood by policy makers, program designers, or even the statisticians. Mathematical models, from which policy makers derive allocation formulas, fail, even with the best available data, to produce the desired allocation pattern, because models generally represent an oversimplification of the real world.

Although proponents of allocation formulas suggest that predictability is one of their strengths (especially when compared with policy makers’ caprice), in many instances, public administrators find they cannot predict the results the formula will produce. Also, often, the formula or model does not reflect the complexity of natural and social phenomena. Unpredictability, as a result, cripples long-term planning and budgeting at all levels of government. This inability to plan, ironically, makes for poor policy execution because of the simplicity of the models and formulas.^[4]

It is not an easy task to design a formula that closely approximates congressional intent; however, the job

gets more difficult when the legislature does not make the goals and objectives of a particular program clear. As one might expect, a decision to adopt a specific formula involves a series of distinct prior choices. An inappropriate decision at any of these steps in the process may lead to a formula, which results in skewed allocations. Generally, in the United States, Congress does not legislate the use of direct measures in the mathematical formula, so statisticians have the arduous task of finding measures that adequately operationalize legislative intent. Quite often, this is not achieved.

A central dilemma for formula allocation is that while it simplifies justifications for budget requests, it frequently presents difficulties in reconciling various policy objectives. For instance, the U.S. Department of Transportation Highway Funding Formula attempts to meet a relatively large number of objectives, some of which are in conflict with one another. For example, one objective is to return funds to the states. At the same time, the program must address national goals and deal with “externalities,” which often require redistributing resources from one state to another.^[15]

PROBLEMS RELATING TO DATA SOURCE

Allocation formulas rely on data that need to be reliable and valid. But frequently, bias and corruption in the gathering and use of data create serious problems, which might not be easily corrected. In other words, it is suspected that other airlines might not have submitted valid data.

Another concern with formula allocation is that irrelevant or outdated factors often undermine the funding calculations for certain services. For example, the formula for apportioning federal highway funds among states is derived from a complicated set of calculations. The U.S. General Accounting Office (GAO) described the formula as cumbersome, relying on factors and data, which, in some cases, are outdated and irrelevant.^[16]

In many cases, allocation measures are selected solely on the basis of familiarity, plausibility, and ease of administration. One measure that is frequently used is per capita. States in the northeast tend to have an advantage when formulas are based on per capita, because they are more densely populated. Those in the south tend to fare better when the formula requires units of \$1000 of personal income. On the other hand, states in the west topped distribution figures when the formula relied on a percentage of local general revenue.^[17]

Furthermore, allocation formulas frequently require census data; however, census data have problems. Census data tend to undercount persons living in rural

areas, ghetto areas, and large cities. Critics of the use of census data maintain that large numbers of Hispanics, African Americans, American Indians, and Asians are frequently missed by the census, with the undercount resulting in a loss of federal funds to minorities. But, others argue that the debate over the significance of the undercount has been highly distorted, even misleading.^[17]

In terms of the fiscal stakes, the GAO 1998 study found that only 0.5% of the \$185 billion distributed by the federal government was affected by the overall undercount.^[17] The reason suggested is that population is only one of several factors in most federal grant formulas, and many programs designed to help distressed communities actually reduce funding when population increases. Even when grant increases are pegged to population gains, the critical factor for a given jurisdiction is not merely its absolute population, but its population relative to other jurisdictions—a result that can obviously hurt as well as help minorities.

However difficult it may be to understand and evaluate the performance of a formula, the task of foreseeing and assuring good performance is even more difficult.^[11] There seems to be three issues. First, the formula may require data that cannot be updated frequently. Policy makers often argue that the formula makes allowances for the data. But often, the degree of distortion caused by the use of obsolescent data cannot be estimated precisely at the time allocations are made. Second, data, which can be updated, may depart from their historical behavior and from their assumed stable relationships with other variables. Third, the social or economic problem to which the program is directed may evolve in such a way that the measure chosen may no longer be the most relevant measure available.

CONCLUSION

Allocation formulas reflect political and economic relationships among federal, state, and local governments. They are effective and convenient tools used to translate congressional intent. When used correctly, they give a predictable framework for government organizations to use to meet needs through goal setting, careful planning, control, oversight, and review. Formula allocation generally guarantees funds to state and local governments based on economic or demographic factors outlined in the formula. Citizens tend to be accepting of allocations that are based on formulas, because they have an aura of objectivity and fairness. Nevertheless, their effectiveness can be severely hampered by data that are stale, inaccurate, and irrelevant.

REFERENCES

1. Grizzle, G. Five issues in budgeting and financial management. In *Handbook of Public Administration*; Rabin, J., Bartley, H., Miller, G., Eds.; Marcel Dekker Publishers: New York, 1989; 197–206.
2. Mikesell, J. *Fiscal Administration: Analysis and Applications for the Public Sector*; Harcourt Brace Publishers: Orlando, FL, 2006.
3. Wildavsky, A. *The Politics of the Budgetary Process*; Little Brown: Boston, MA, 1984.
4. Fessler, W. *The Politics of the Administrative Process*; Chatham House Publishers: New Jersey, 1996.
5. Dugdale, P. The better practice program funding formula. *Aust. J. Public Admin.* **1997**, *56* (3), 65–67.
6. Parish, T. *Removing Incentives for Restrictive Placements. Fiscal Issues in Special Education*; series: CSEF Policy Paper 4 ED.377608; US Department of Education: Washington, DC, 1995.
7. Bowman, A.; Kearney, R. *State and Local Government*; Houghton Mifflin Company: Boston, MA, 2005.
8. U.S. General Accounting Office. Federal Funding 2001 (<http://financenet.gov>).
9. Spencer, B. Technical issues in allocation formula design. *Public Admin. Rev.* **1982**, *42*, 524–529.
10. General Service M Administration 11.553 School Breakfast Program www.cfda.gov.
11. Kreps, J.; Slater, C. *Report on Statistics for Allocation Formulae*; series: Statistical Policy Working Paper (NTIS PB86-211521/AS); Federal Committee on Statistical Methods: Washington, DC, 1978.
12. Cothran, D. Some sources of budgeting accountability. *Public Budg. Finance* **1986**, *6*, 45–62.
13. United States Public Law 107–42, Washington, DC, September 22, 2001.
14. Parry, D. *The Highway Trust Fund Uses a Complicated Allocation Formula* Economic Research Service; series: Agriculture Information Bulletin; <http://www.ers.usda.gov/publications/aib753b>; USDA: Washington, DC, 1999; 3 pp.
15. Parry, D. *The Highway Funding Has Been Criticized*; series: Agriculture Information Bulletin; <http://www.ers.usda.gov/publications/aib753-1>; Economic Research Service: USDA: Washington, DC, 1999; 4–5.
16. U.S. General Accounting Office Highway Funding. *Alternatives for Distributing Federal Funds*, Report to Congressional Committees, 1995; GA1.13:RCED-96-6 ([0546-D(MF)]).
17. Skerry, P. *Race, Group Identity and the Evasion of Politics*; Brookings Institution Press: Washington, DC, 2000.

Budgets and Accountability

Jonathan Justice

School of Urban Affairs and Public Policy, University of Delaware, Newark, Delaware, U.S.A.

INTRODUCTION

From one perspective, public budgets and budgeting processes may be viewed as being, first and foremost, instruments of accountability. Accountability can mean almost anything, depending on who is defining it, but, at its core, involves the communication of governing expectations for the conduct of those held accountable and the subsequent communication of accounts of that conduct to those holding them accountable. Public budgeting, both in its decision-making processes and in the documents recounting the decisions made, determines the substance of what officials are expected to do with public resources, as well as the manner and extent to which particular individuals and groups in a society are involved in specifying those expectations. Budgets, together with after-the-fact financial reporting, also constitute accounts of decisions and action by means of which stakeholders, including the general public, can hold public officials accountable for their resource-using behavior.

A century of budget and financial reporting (or “accounting”) reforms in the United States have been aimed largely at making budgeting a more effective means of holding users of resources accountable to others. As reformers and other students of budgeting have learned, budget processes, formats, and the distribution of financial information are important factors in determining whether, to whom, and for what public organizations and administrators are actually accountable in their use and application of resources, and, just as budgets themselves are instruments for making tradeoffs among valued uses of resources, budget design choices often involve their own tradeoffs. Efforts aimed at improving budget and financial reporting formats and processes have borne considerable fruit, aided considerably by developments in information technology (IT). Efforts to educate and engage the public have been less successful, after the initial successes of the New York Bureau of Municipal Research in the early twentieth century.

Technical progress notwithstanding, significant challenges of accountability for public resource-use decisions exist at the dawn of the twenty-first century. In spite of our historically unprecedented wealth, contemporary Americans may, in our desire to get

something for nothing, be trading off our own welfare against that of future Americans. If we are to make it clear to ourselves as a polity whether and to what extent that is true, continuing progress toward solving technical problems of process and presentation design is as necessary as it is likely. Much greater progress in the practice of public attraction and education is equally or even more necessary, however, if we are to have any prospect of succeeding in holding ourselves accountable to future generations for wise stewardship of collective resources.

ACCOUNTABILITY

Accountability in the simplest, idealized, managerial-instrumental sense provides a means for one person or group (the principal) to manage (or at least influence) the actions of another person or group (the agent), as it relates to some expectation of performance. This model of accountability relies on the prospect of postperformance sanctions and/or rewards, applied after an examination of the agent’s behavior and/or the results of that behavior, as an efficient way to provide incentives for the agent to satisfy the principal’s expectations. Key elements in this model of accountability, then, are the principal–agent relationship, the ex ante stipulation of expectations, the ex-post account that enables comparison of actual with desired behavior and outcomes, and the actual/anticipated application and effects of sanctions or rewards. For this form of accountability to be effective, it is generally presumed that all four elements must be in place and recognized by all parties to the relationship. There must be a principal–agent relationship and a delegated task or responsibility such that the parties are able to establish a relatively unambiguous (and feasible) set of expectations for the agent’s performance. There must be an account of the agent’s performance through observation by the principal, reporting by the agent, or some other means of monitoring by which the principal can determine accurately and reliably whether the agent’s performance satisfied those expectations. There must also be at least the potential for sanctions to be applied in the event that performance is not

satisfactory. The agent's internalization of the principal's expectations (and acceptance of them as legitimate and/or likely to be enforced) can improve the efficiency of the arrangement by reducing the need for monitoring and application of sanctions by the principal.

Accountability, in practice, however, takes a nearly infinite variety of additional forms, including those based on indirect or mediated relationships linking principals and agents, relationships of agents with multiple principals who hold competing expectations, and other relationships in which there is not an obvious principal and agent—more non-hierarchical relationships such as those with collaborators or social and professional peers.^[1,2] Even in what are apparently relatively straightforward principal-agent settings, efforts by agents to manage the often ambiguous expectations of principals can lead to complications in the relationship and behaviors.^[3] Accountability of public servants in democracies is, in some ways, particularly problematic, as illustrated by the evergreen debate over professional discretion and expertise,^[4,5] and by the way in which principal-agent models become paradoxical when street-level bureaucrats are required to hold accountable the citizens who are presumed through a lengthy chain of delegated relationships to be holding the bureaucrats accountable.^[6] The complexity and scale of modern governance and the concomitant rise of professional specialization and expertise, together with increasing reliance on privatization and other indirect means of public-goods production, introduce additional challenges to the simple model of hierarchical accountability of street-level bureaucrats to managers, managers to agency heads, heads to elected officials, and elected officials to voters/citizens.

Even in all of accountability's variety and complexity, however, certain elements generally must remain if we are to speak of being held accountable or holding others accountable: expectations and accounts. First, there must be some form of social relationship, in which the parties to the relationship recognize that one (or more) of them has particular expectations *ex ante* about how the other(s) will behave in a given situation. All parties who will hold or be held accountable must know what those expectations are, if accountability is to exist in any meaningful form. That is, we must be able to determine *to whom* and *for what* some person(s) is/are accountable: Whose expectations legitimately establish normative standards for action, and what is the substance of those expectations? Second, in order for one to hold another accountable (or oneself to be held accountable) *ex post*, it must be possible for the behavior of the one(s) held accountable to be monitored by or, in some other fashion, made known to the one(s) holding them accountable. Practically speaking, this means

that when direct monitoring is not feasible, an account must be provided to the holder. It also means that the substance and intelligibility of that account, as well as of the benchmark statement of expectations, will influence the holder's response to the behavior of the party(ies) held accountable.

BUDGETS

Public budgets and budget processes are critically important in public policy and administration, even for those who despise them, because they are precisely the determinants of politics' "who gets what, when and how." Budgets are made necessary by the simple reality of *scarcity*. Households, governments, and other organizations generally do not have access to unlimited levels of resources and so must make choices, trading off some of their costly, valued goals against others: guns vs. butter; present vs. future consumption; tax expenditures vs. overall revenue capacity; internal waste or misappropriation vs. desired outputs; and so on. One of the benefits of replacing department-by-department legislative appropriations with the comprehensive, annual or biennial budget processes adopted in many U.S. state and local governments around the turn of the twentieth century was that it provided for explicit consideration of relative priorities and tradeoffs. In order to accomplish this trading off, budgeting performs at least three broad functions: *planning* (social resource allocation); *management* (bureaucratic resource allocation); and *control* (regulation of resource consumption).^[7,8] Government budgets, thus, can be understood as statements of cultural and political values and preferences for allocating societal resources toward desired social outcomes, as devices for managing allocated resources in the interest of efficiently and effectively achieving those outcomes (and their antecedent outputs), and as means to prevent the misuse of allocated resources.

As strategic plans and instruments of managerial and operational control, budgets stipulate *expectations* concerning the levels of resources available for particular public purposes during particular time periods and how and by whom those resources are to be employed in efforts to achieve those valued ends. Government budgets express those expectations in the form of laws—appropriations enacted by legislatures, approved by executives, and presumably legally binding on all concerned. A budget document also serves as an *account* of what has taken place over the course of a budgetary decision-making process. Budget processes normally include a series of activities encompassing, to some degree, the formulation and communication of expectations concerning resource allocations and use, as well as the production and

evaluation of accounts describing the use of resources. The form and content of budget documents as well as of budget formulation processes, budget implementation (or “execution”) activities, and various accounts thereof are, thus, key determinants of who among public organizations and officials will be accountable as well as how (and how well), to whom, and for what they will be accountable. Thus, the illustrative history of achieved, failed, and potential future reforms and practices of public budgeting, financial management, and reporting in the United States over the past century demonstrates some of the ways in which budget formats, processes, and reforms influence the quantity and quality of accountability, for the allocation and consumption of public resources.^[9,10]

BUDGETS AND ACCOUNTABILITY

Returning to our two questions concerning the nature of accountability in particular instances—accountable to whom and accountable for what—we can now consider a few of the implications for accountability of the structure and content of budget processes and formats. Under the “for what” heading, the characteristics of budgets as means for formulating and describing choices about tradeoffs, broad resource allocations, and specific uses of resources—that is, the substance of expectations—come to the fore. The question, “accountable to whom,” introduces complexities regarding the availability and usefulness of budgets and financial reports as accounts as well as matters of how the design of budget processes determine who participates in the formulation of the expectations embodied in budget decisions. Both areas of concern lead to the murkier, but no less important, matter of who is being held accountable for budget choices and practices.

Accountable for What

The history of U.S. budget reforms throughout the twentieth century played out in large part as a matter of budget format. It is customary to divide budget formats according to the three-part, planning-management-control distinction: Planning is associated with a variety of program-budget designs, management with “performance” budgeting, and control with line-item or object-of-expenditure formats. Each orientation implies a different for-what emphasis of fiscal accountability: outcomes; outputs and activities; or inputs. Within each broad category of formats fall a range of specific approaches, each of which can vary considerably in the level of detail to which expectations are specified.

Schick’s historical review of budget reforms in the first two-thirds of the twentieth century noted that early reforms focused on control as a primary objective to be advanced by budget formats, thus leading to line-item budgets that specified in considerable detail how appropriated funds could be spent.^[8] This was meant to reduce misappropriation in the form of fraud and waste, by holding public officials accountable for spending public money only on those inputs (“objects of expenditure,” such as certain types and quantities of specialized labor, particular items of equipment, or certain kinds of materials and supplies) specified by the budget. A second wave of reforms emphasized managerial control of organizational performance: activities and outputs. Thus, performance-based budget formats emphasize what is to be done or accomplished by allocating resources and imply a need for good measures of organizational performance relative to money expended. Budgets and ex-post reports in a performance budgeting regime lead, thus, to accountability of public officials for provision and production of particular desired goods and services and for productive efficiency. Finally, program budget formats such as the planning-programming-budgeting system or zero-based budgeting emphasize social planning and accountability for the accomplishment of desired broad social outcomes or even social welfare in general. Schick, in his historical review, pointed out that each version of reform reflected particular concerns of reformers about for what they wanted public officials and organizations to be held accountable. And indeed, at least one empirical study has found that the format of a budget presentation does, in fact, direct the attention of elected officials in just these ways.^[11]

Schick also noted that in the past, practical limitations on the ability to “cross-walk” budgetary allocations across the different formats had constrained reformers’ ability to implement performance- and, especially, program-budget reforms fully. Line-item formats and budgetary presentations aligned with administrative structures are better adapted to budget execution, by contrast with program budgets that may gather together in a single allocation category activities undertaken by multiple administrative units. This problem was largely soluble by using computers to store, retrieve, and manipulate budget and financial data, even by the time Schick wrote in 1966, and contemporary IT is exponentially more powerful in the early twenty-first century than it was in the 1960s. Still, the most grandiose reform schemes, aimed at holding government accountable through the budget for achieving broad programmatic goals of social welfare, largely came to naught for a variety of reasons: resistance to change; the failure of human decision makers to expand their information-absorption and information-processing abilities as rapidly as computers

have; and, to some degree, both the technological infeasibility of holding officials accountable for generating social outcomes from monetary inputs in the face of our imperfect understanding of how to transform money into social welfare and the political infeasibility of reaching agreement on unambiguously stipulated policy goals.^[12]

Still, U.S. governments have taken a number of steps in recent years to reform budgets from the perspective of making it possible to hold governments and public officials more accountable for more aspects of resource allocation and use, with much of the credit due to professional associations and standard-setting bodies as well as good-government groups. Many state and local budgets include at least some elements of performance and program information, and the Government Finance Officers Association (GFOA) promulgates standards for, and offers awards to recognize, exemplary budget presentations. Even the U.S. federal government has of late begun more or less systematically to assess the performance of programs on a regular basis, through its new program assessment rating tool, and to use that assessment in budget formulation. More traditional state and local accountability expectations exist in the form of reporting obligations, balanced-budget requirements, debt restrictions, and other constitutional and statutory prior constraints on budgeting and financial management. State and local governments respond as well to the reporting standards embodied in the generally accepted accounting principles (GAAP) and service efforts and accomplishments, reporting recommendations promulgated by the standard-setting Government Accounting Standards Board (GASB). Additional steps that can be taken fairly readily by municipalities in particular include being more explicit in their budget presentations about relevant legal obligations and constraints and about the economic contexts that influence revenue and expenditure choices, and providing information in budgets and annual reports that facilitates translation between budgetary and financial-reporting bases of accounting.^[9]

Progress notwithstanding, any number of “accountable for what” issues remain unresolved. One perennial discussion concerns the matter of budgetary time frames—the period of time for which budget formulators and implementers are to be held accountable. One question concerns the appropriate time period for budgetary planning and balance: Should it be annual, as is most common in the United States, or cover two to five years, or should it employ a multiyear perspective that provides for balance over the full business cycle, with operating deficits in recession years balanced by surpluses in expansion years?^[13] A related issue is the familiar “use it or lose it” problem associated with budgetary periodicity: the perverse incentive

for operating units to rush to spend appropriated but currently un-needed funds before the end of a fiscal period, rather than losing the available budget authority and facing reduced allocations in future periods. Additionally related to time is the question of delayed effects: Spending in the current fiscal period may be linked with program outcomes several years in the future. What then is the appropriate lag to use in holding officials and agencies accountable for final outcomes?

A final set of accountability-for-what issues involves “the relation of budget documents to reality” and implicates the expectations and accounts generated by budgeting practices and processes as well as formats.^[9] These mainly involve the comprehensiveness and accuracy of budget estimates and financial reports. One form of gamesmanship that complicates accountability is evident at this writing in the federal government’s practice of using supplemental “emergency” appropriations, rather than the regular budget to fund readily predictable war expenses on the order of \$100 billion per year. This, of course, means that estimates of budget (im)balance and tradeoffs among priorities in the regular budget process are misleading, because they do not account for this large group of expenditures. A related practice that muddies accountability is the use of extensive inter- and intrafund transfers of funds during a budget year: The actual use of resources may vary significantly from the initially articulated expectations and accounts embodied in the adopted budget. Another practice that undermines accountability by facilitating “rebudgeting” is the systematic underestimation of revenue and overestimation of expenditures, which provides elected officials and administrators with extra funds they can allocate during a fiscal period.^[14] After-the-fact reporting (accounting) that clearly highlights comparisons of actual to budgeting revenues and expenditures can make it easier to determine whether rebudgeting responds to unforeseeable exigencies or has become a habit for officials seeking to avoid accountability for resource allocation and use.

The proliferation of off-budget entities and means of incurring debt are other common ways of manufacturing “fiscal illusions” that can succeed *in the short term* in evading or deterring of deferring budgetary tradeoffs. However, to anticipate this essay’s concluding remarks somewhat, it is reasonable to ask under the heading of “accountable for what” whether some of public officials’ seemingly reckless and unaccountable practices of creating fiscal illusions with debt, rebudgeting, and selective presentations of information might, in fact, respond to deeper expectations, communicated directly or indirectly to administrators by elected officials and/or by voters to their representatives.^[15] At this point, we can turn to the second question of accountability: to whom?

Accountable to Whom?

Whose expectations form the basis for budget decisions and actions, and to whom are accounts of those decisions and actions available, comprehensible, and actionable? A fundamental tension of large, active, representative governments is between the competing governing principles of professional specialization and the expertise, on the one hand, and popular control, on the other hand.^[4,5] This tension is readily evident in a century of attempted and accomplished budget reforms in the United States. At the turn of the last century, the Progressive-era reformers espoused a range of accountability ambitions. They wanted to attract members of the public to “mind their own business” by becoming involved in making budget decisions, attending to financial reports, and holding public officials accountable for responding to their desires and interests, as activist owners of government.^[16] The budget exhibitions staged by the New York Bureau of Municipal Research (BMR) were famous, innovative attempts to popularize involvement in budgeting by attracting and educating non-specialists. The BMR understood that, in order for public budgeting to be meaningfully accountable to citizens, the public needed not only to have access to accounts and opportunities to communicate expectations, but also to be educated in order to understand those accounts and to formulate and communicate realistic fiscal expectations to officials responsible for the details of budget making and execution.

The subsequent direction of budgeting and financial reporting reforms, however, emphasized the roles of administrators and chief executives in particular. Reforms focused on improving the ability of responsible executives to hold the bureaucracy accountable and implicitly or explicitly relied on those executives in turn to be accountable to the larger public. Practices and standards moved more toward detailed and complex presentations that were very useful for specialists and insiders, but often opaque to non-specialists and sometimes not even readily available for inspection by the general public.^[10] By the end of the 1960s, public budgeting seemed, in large part, an exercise conducted almost in private, with public hearings serving more as ritualized accounts of decisions already taken than as opportunities for the public to convey expectations that would govern officials’ choices and actions. The first few years of research conducted and sponsored by GASB in the 1980s and 1990s reflected the existing bias toward seeing specialists as the key consumers of unfiltered government financial data. Categories of users were understood to include citizen groups (news media, researchers, and advocacy/informational groups), legislative and oversight officials, and investors and creditors; individual citizens were not among

the users of financial reports explicitly considered by the researchers.^[17–20] GASB did sponsor research aimed at identifying standards of good practice for preparing simplified popular financial reports.^[21] The number of jurisdictions preparing such reports is not large, however, and the degree of reduction and selection involved means that only the bravest (or most foolhardy) of officials would be likely to prepare reports that could result in their being held meaningfully accountable for any errors or omissions in their fiduciary roles.

For many in the field, the taxpayer revolts results and tax-and-expenditure limitations of the 1970s and 1980s—with California’s Proposition 13 perhaps the seminal instance—changed the game in some fundamental ways, as did the near-bankruptcy of New York City in 1975. The 1980s also brought President Reagan’s efforts to cut off Congress’s allowance by reducing federal tax revenues. Cutbacks, tradeoffs, and conflict came to be familiar concerns for public budget professionals, and the attendant revenue squeezes together with apparent contempt for government and public servants among the U.S. citizenry prompted a variety of professional innovations. The last quarter of the twentieth century brought, among other things, the creation of GASB; the “reinventing government” movement and Vice President Gore’s National Performance Review; the promulgation by GFOA and other professional bodies of a range of standards and recommendations for increasing public awareness and understanding of, and participation in, budget processes and decisions; improvements in the craft of performance measurement and renewed efforts to link performance measures to budgets; a raft of federal legislation such as the Government Performance and Results Act and the Chief Financial Officers Act; greater regulatory scrutiny and improved disclosure practices in the municipal bond market; and an almost sickening repetition of the mantra of “accountability” in public discourse. The latter part of the twentieth century also brought, however, increasing sophistication in the creation of “fiscal illusions,” using both traditional means such as debt-issuing public authorities and “one-time” revenues, and newer approaches such as permanent peacetime budget deficits at the federal level, the Orange County investment pool, the issuance of bonds to finance pension obligations, and load-shedding privatizations of operating responsibilities and infrastructure at all levels of government.

INTO THE TWENTY-FIRST CENTURY

Are we better off today, in terms of accountability for the allocation and use of public resources, than we were in the last century? It all depends on what

we mean by “accountability.” Public officials do appear to be accountable for more, and to more audiences, with respect to their budget and financial-management actions, with every passing year. At the same time, it is not clear that this increased accountability is encouraging improved stewardship of collective resources. Also unclear is whether contemporary movement away from fiscal sustainability is a failure of public servants’ accountability—perhaps due to a still insufficiently aroused and informed public—or a triumph of accountability to contemporary voters and interest groups that reflects their unconcern for the interests of future generations.

Certainly, there are some positive signs, particularly with respect to the quantity, quality, and availability of after-the-fact reporting and financial information; to some degree in the design and information content of budget formats and processes; and to a lesser degree in some apparently sincere efforts to engage the public in the formulation of budgets. The work of professional associations, regulatory bodies, special- and general-interest watchdog groups, and standard-setting organizations has certainly improved financial reporting and disclosures by state and local governments, as well as the potential for more comprehensive approaches to budgeting. The calculation and publication of performance measures, and their utilization in budget processes and decisions, while still modest for the most part, appear to be on the increase at all levels of government. Specialists in the field are increasingly able to use improved cost-accounting methodologies and data, many governments produce tax-expenditure budgets and multiyear financial estimates and plans, and fiscal notes often accompany legislative proposals. More and better accounts of financial condition and performance are available to those with the time, skills, and incentives to obtain, examine, and understand them, especially financial professionals. It can reasonably be hypothesized that the expectations of bond rating agencies and investors calling for prudent long-term financial management by state and local debtors influence budgeting and financial management practices to a significant degree. Recent developments in IT, particularly the World Wide Web, have brought with them the promise—and in a few instances the reality—of “providing better information in better ways” to more stakeholders, including the general public.^[22] By 2006, more than 90% of the 1069 jurisdictions that earned the GFOA’s 2005 Distinguished Budget Presentation Awards had adopted the GFOA’s recommended practice by posting their budget documents on the Web, up from less than 60% of a 2004 sample of winners of the award for 2003.^[23,24] The website Guidestar.org continues to expand the scope of its information on the finances of non-profit organizations. Refinements to GAAP

and other professional standards continue to improve the technical comprehensiveness and comparability of government financial reports and disclosures. The journals published by the GFOA and the International City/County Management Association routinely feature exhortations and practical recommendations for meaningful performance reporting, the use of performance information in budgeting, active citizen participation in budgeting, and popular financial reporting.

Technology and professional standards, however, have not yet proved and probably never will prove to be silver magic bullets capable of simply erasing the challenges to accountability in public budgeting, which continue to bedevil would-be reformers. Political processes, almost certainly, will never operate with perfect efficiency nor generate the economically optimal results that experts’ cost-benefit calculations prescribe. Budgeting is no more likely—in fact, it is probably less likely, alas—to be a topic that overcomes the competing day-to-day demands on citizens’ time and attention that constrain their active involvement in political and policy-oriented activities. Thus, large portions of the public may fail to avail themselves of even the best and most widely disseminated information and opportunities to participate in budget formulation. An additional problem of accountability in using program and performance-oriented models of budgeting is how to respond to results. If a program crosses organizational lines, which administrative organizations should be held accountable and which held harmless for program failures? Even when good performance measures are computed and reported, the paradoxes of performance measurement makes using such data in budget decisions less than perfectly straightforward: Does poor performance call for budget cuts as a way to stop wasting money, or does it mean that additional resources are needed? Does good performance call for budget increases on the assumption that more money will yield more of a good thing, or does it indicate that a program’s budget can safely be reduced or held steady?

Finally, it may be that we in the United States—ordinary citizens, power brokers, and public servants alike—are, at least in part, failing to hold *ourselves* sufficiently accountable *to* future generations in the United States, *for* responsible stewardship of our collective wealth. At all levels of government, we seem to be satisfied, in practice if not in our public exclamations, only by the continuing creation of more fiscal illusions—resulting in growing external debt, neglected infrastructure, and other burdens on our descendants—as we demand tax cuts without reductions in collective benefits. Lewis writes of a prevailing “paradoxical public attitude whereby the public advocates specific programs but rejects the aggregate cost,” and observes, “Neither increasingly stringent professional standards

nor advances in technology and technique appear to have altered public perceptions.”^[25] That is, what we have here may be less a failure to communicate expectations and accounts, and more a failure to acknowledge the scarcity of resources and the consequent need to accept tradeoffs that together constitute the fundamental rationale and logic of budgeting in the first place.

Gardiner has written of the three-part “perfect moral storm” posed by climate change: a problem of collective action among contemporary actors seeking to gain benefits for themselves and shift costs to others; technical problems that render imperfect our understanding of the likelihood and consequences of climate change, and of how it may be prevented; and an intergenerational problem of collective action, as contemporary actors seek to enjoy benefits while passing costs forward to a future generation unable to speak for itself.^[26] The structure of that problem is precisely analogous to that of fiscal stewardship, as Gardiner notes in quoting Samuelson’s nicely phrased observation:^[27]

Americans dislike deficits but dislike them less than the alternatives—higher taxes or lower spending. There’s a quiet clamor for hypocrisy and deception; and pragmatic politicians respond with massive borrowing schemes that seem to promise something from nothing. Please, spare us the truth.

In short, much has been done to make public budgeting accountable over the past hundred years and much remains to be done. Public servants, their professional associations, watchdog organizations, and academics concerned with public budgeting and financial management have made great strides toward addressing the technical third of our perfect moral storm, so that it is increasingly feasible for citizens as well as creditors and investors to hold public officials and organizations accountable for their use of resources. Efforts to improve officials’ abilities to grasp and incentives to respond to public expectations as well as the usability and availability to all audiences of meaningful accounts of resource use should continue. But such efforts must be complemented by renewed efforts to engage and educate the non-specialist public, since even the best account may be ignored by or incomprehensible to an ill-equipped or inattentive audience. By making the implications and consequences of collective choices clearer to all concerned, public administrators can at least make widespread public self-deception and hypocrisy less easy and more apparent. In so doing, we can make ourselves more accountable to future as well as present generations. Then it has to be up to the various individuals and constituencies that constitute “the public” to determine whether and how to discharge our own fiscal responsibilities to each other and to our descendants.

ARTICLES OF FURTHER INTEREST

- Accountability and Ethics: Reconsidering the Relationships*, p. 10.
Achieving Productivity Through Budgeting, p. 38.
Auditing, p. 123.
Budget Execution and Management Control, p. 158.
Budget Reform, p. 166.
Citizen Participation in the Budget Process: Exit, Voice, and Loyalty, p. 305.
Control Systems, p. 404.
Ethics and Administrative Reform, p. 686.
Motives and Incentives: Rethinking Budget Theory in a New Institutional Context, p. 1275.
Performance Measurement: Citizen-Driven, p. 1459.
Performance Reporting, p. 1465.
Popular Reporting, p. 1516.
Public Budgeting, Politics in, p. 1605.

REFERENCES

1. Romzek, B.S.; Dubnick, M.J. Accountability in the public sector: lessons from the Challenger tragedy. *Public Admin. Rev.* **1987**, *47* (3), 227–239.
2. Weber, E.P. The question of accountability in historical perspective: from Jackson to contemporary grassroots ecosystem management. *Admin. Soc.* **1999**, *31* (4), 451–494.
3. Kearns, K.P. Accountability and entrepreneurial public management: the case of the Orange County investment fund. *Public Budget. Finance* **1995**, *15* (3), 3–21.
4. Friedrich, C.J. Public policy and the nature of administrative responsibility. *Public Policy* **1940**, *1*, 3–24. Friedrich suggested here that, in the increasingly complex and professionalized environment of mid-twentieth century government, professional norms and peer monitoring and competition among technical experts could constitute the most appropriate model of bureaucratic accountability.
5. Finer, H. Administrative responsibility in democratic government. *Public Admin. Rev.* **1941**, *1* (4), 335–350. Finer responded to Friedrich by arguing that professionals in representative democracies ought, more appropriately, to respond to the explicit instructions provided to them by the people’s representatives through the medium of legislation, rather than by interposing their own professional values.
6. Lipsky, M. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*; Russell Sage Foundation: New York, 1980.
7. Anthony, R.N. *Planning and Control Systems: A Framework for Analysis*; Division of Research Graduate School of Business Administration Harvard University: Boston, 1965.
8. Schick, A. The road to PPB: the stages of budget reform. *Public Adm. Rev.* **1966**, *26*, 243–258.

9. Rubin, I. Budgeting for accountability: municipal budgeting for the 1990s. *Public Budget. Finance* **1996**, 16 (2), 112–132.
10. Kahn, J. *Budgeting Democracy: State Building and Citizenship in America, 1890–1928*; Cornell University Press: Ithaca, NY, 1997.
11. Grizzle, G.A. Does budget format really govern the actions of budgetmakers? *Public Budget. Finance* **1986**, 6 (1), 60–70.
12. Wildavsky, A. A budget for all seasons? Why the traditional budget lasts. *Public Adm. Rev.* **1978**, 38, 501–509.
13. Hou, Y. Budgeting for fiscal stability over the business cycle: a countercyclical fiscal policy and the multiyear perspective on budgeting. *Pub. Adm. Rev.* **2006**, 66 (5), 730–741.
14. Forrester, J.P.; Mullins, D.R. Rebudgeting: the serial nature of municipal budgetary processes. *Public Adm. Rev.* **1992**, 52 (5), 467–473.
15. Justice, J.B.; Miller, G.J. *Debt Networks and Accountability: The Case of the New York Metropolitan Transportation Authority's Capital Program and Debt Restructuring*, Paper presented at the annual meeting of the Association for Budgeting and Financial Management, Washington, DC, November, 10, 2005.
16. Schachter, H.L. *Reinventing Government or Reinventing Ourselves: The Role of Citizen Owners in Making a Better Government*; State University of New York Press: Albany, NY, 1997.
17. Hay, L.E. *A Study of the Usefulness of Disclosures Required by GASB Standards*; Governmental Accounting Standards Board: Stamford, CT, 1988.
18. Jones, D.B.; Scott, R.B.; Kimbro, L.; Ingram, R. *The Needs of Users of Governmental Financial Reports*; Governmental Accounting Standards Board: Stamford, CT, 1985.
19. Patton, J.M. *An Empirical Study of Governmental Financial Reporting Entity Issues*; Governmental Accounting Standards Board: Stamford, CT, 1987.
20. Wilson, E.R. *Financial Reporting by State and Local Governments: A Survey of Preferences Among Alternative Formats*; Governmental Accounting Standards Board: Stamford, CT, 1990.
21. Carpenter, F.H.; Sharp, F.C. *Popular Reporting: Local Government Financial Reports to the Citizenry*; Governmental Accounting Standards Board: Stamford, CT, 1992.
22. Boydston, R.; Aaron, W. The citizen's guide to the Nashville budget: providing better information in better ways. *Gov. Finance Rev.* **2004**, 20, 12–16.
23. Exemplary budgets. GFOA Newsletter, November 10, 2006; 1.
24. Justice, J.B.; Melitski, J.; Smith, D.L. E-government as an instrument of fiscal accountability and responsiveness: do the best practitioners employ the best practices? *Am. Rev. Public Adm.* **2006**, 36 (3), 301–322.
25. Lewis, C.W. The field of public budgeting and financial management, 1789–2004. In *Handbook of Public Administration*, 3rd Ed.; Rabin, J., Hildreth, W.B., Miller, G.J., Eds.; CRC Press: Boca Raton, FL, 2007; 151–225.
26. Gardiner, S.M. A perfect moral storm: climate change, intergenerational ethics, and the problem of moral corruption. <http://faculty.washington.edu/smgard/GardinerStorm06.pdf> (retrieved September 13, 2006).
27. Samuelson, R.J. Lots of gain and no pain! *Newsweek* **2005**, 145, 41.

Bureaucracy, Economic Theories of

James R. Simmons

Department of Political Science, College of Letters and Sciences, University of Wisconsin-Oshkosh, Oshkosh, Wisconsin, U.S.A.

INTRODUCTION

Economic theories of bureaucracy begin with the assumption that people who act in a self-interested way in the marketplace are the same people who vote, run for office, and are employed in the public sector. Government officials and civil servants are portrayed as acting to advance their own private agendas in the name of the “public interest.” The primary implication drawn from this perspective is that the collective political process fails because participants are responding to incentives other than those assumed by conventional theory. Economic reasoning promises a logical deductive model of how public agencies behave with clear directions for policy analysis and simple normative recommendations for reform.

Classical management theory envisioned public sector bureaucracy as a modernizing form of organization that would ensure precision, speed, reliability, clarity, efficiency, and regularity in the provision of high-quality goods and services. In contrast, modern political economists have developed theories of bureaucracy, which see in these same hierarchical structures institutions that foster self-serving behavior that corrupts the natural functioning of markets and mask the political pursuit of private gain by those who manipulate them. Self-interest in the private sector is said to produce optimal outcomes. However, in all government agencies, this same motivation generates command-and-control monopolies operated by an intrusive, problem-finding elite that overemploys, overpays, overbuilds, and overproduces. Thus, whenever possible, the reliance on the public sector should be reduced and market forces restored.

HISTORY

Much of the prescription for the model organization comes down to us from the writings of Max Weber. In Weber’s bureaucratic world, public employees are rational professionals who follow the routines and duties of the agencies for which they work. Their values are expected to be limited to the neutral pursuit of their organization’s objectives or official norms, while their status is derived from their training, title,

and job performance. Although organization theory has been dominated by adherents of the Weberian perspective, modern political economists have developed a competing diagnosis of bureaucracy that emphasizes public failure in the absence of market discipline. This literature on public bureaucracy can be traced back to Ludwig von Mises and Frederick Hayek. However, neither their works is a recognizable economic account because there is little abstract modeling in either contribution. Formal theorizing about government is best represented by contemporary authors such as Anthony Downs, Gordon Tullock, Vincent Ostrom, and William Niskanen.

Economic theories of bureaucracy begin with the assumption that people who act in a self-interested way in the marketplace are the same people who vote, run for office, and are employed in the public sector. Government officials and civil servants are portrayed as acting to advance their own private agendas in the name of the “public interest.” The primary implication drawn from this perspective is that the collective political process fails because participants are responding to incentives other than those assumed by conventional theory. Economic reasoning promises a logical deductive model of how public agencies behave with clear directions for policy analysis and simple normative recommendations for reform.

The distinguishing feature of economic theory is its concentration on a narrow conception of individuals as inherent utility maximizers. This account of what bureaucrats want is a direct analogy of the standard neoclassical assumption that the managers and owners of private firms maximize profits, because their compensation is tied to performance (materialistic motivation) and because high profits safeguard management against aspiring underlings or potential hostile external takeover bids that would jeopardize their positions (survival motivation). But because public bureaucracies are “non-profit organizations,” what do bureaucrats maximize?

The basic answer (apart from Downs) has been that officials maximize the size of their agency, the size of their budget, or their job satisfaction. Some theorists have suggested that size is assessed primarily in terms of size. Bureaucrats from larger, well-financed agencies will have greater visibility, enhanced reputations for

influence, and higher status among their peers. Comparative budget increases across departments are seen as an indicator of winners and losers in the ecological competition for resources. Bureaucrats in larger agencies control more patronage and have a greater capacity to influence events or overcome resistance. Size also creates organizational slack and excess capacity that allows officials to more easily meet unusual risks or crises whenever they occur.

In addition, maximizing behavior is critical for an agency's survival in two ways. The organization's budget is allocated to it by a "sponsor"—a single, dominant collective entity such as a chief executive, a department secretary, or a committee of the legislature. Sponsors expect to be routinely presented with proposals for increased funding. During the budgetary process, attention will concentrate on the marginal increments being sought. If no such increase is requested, the funding body is likely to become suspicious or disoriented and these procedures become inoperative. Completely static or reduced budget requests are likely to be interpreted as evidence of the bureau's stagnation or loss of initiative.

Senior bureaucrats seek funding expansion as a balm that facilitates agency operations. In large or growing agencies, contentious or incompetent officials can be moved sideways into sinecures. Also, the costs of reorganization can be more easily absorbed and wage demands or staff resistance to change can be more easily bought off. Budget maximization facilitates the hiring of skilled staff, and improves wage and promotion prospects. Client groups also value top officials who expand programs because this will enhance their own agendas. They will even cooperate by mobilizing public support for the agency's policies, especially where key clients are well-organized interest groups.

The economic account says little explicitly about the internal operations of government agencies. Most writers assume that institutional and functional variations among government departments simply do not matter. They generally adopt the view of bureaus as command-and-control organizations run in a top-down manner. This contrasts starkly with markets (which they prefer) that they claim operate as decentralized preference systems. In addition, proponents employ a unitary theory of the firm that treats both private corporations and public agencies as if they are run by a single management team so that they can focus on their aggregate behavior vis-à-vis other actors. The basic differential in the behavior of these "firms" is explained by the profit driven "bottom line" of business and the less measurable "public interest" service imperatives of government.

Similar to other organizations, bureaus have an inherent tendency to expand, but with fewer than the

normal constraints on their capacity to do so. Agency growth helps in the retention and promotion of personnel, and it enables leaders to acquire more power, income, and prestige. It reduces internal conflicts, facilitates economies of scale, helps to insulate top officials from social pressures, and strengthens the agency's ability to survive. Larger agencies are also in a better position to manipulate their environment than are small ones. If top officials do not have to pay the costs of adding more personnel, they will be inclined to increase the size of the organization indefinitely because each new member adds incrementally to their total direct action capabilities.

Most economic analyses of bureaucracies posit a fundamental dichotomy between two basic ways of coordinating social activities in industrialized societies—markets and command structures. Anthony Downs labeled this contrast the "*law of hierarchy*." Coordination of large-scale activities without markets requires a hierarchical authority structure. This law results directly from the limited capacity of each individual, as well as the existence of ineradicable sources of individual conflict. In particular, the general self-interest axiom implies that actors will not cooperate unless compelled to do so by an effective command structure. Yet, this ultra-Weberian picture of government agencies is qualified by more stress on an underlying self-centered inclination to resist the demands of their superiors.

Max Weber looked for alternative mechanisms of administrative coordination through the inculcation of a common organizational culture or strong ideology that guides and justifies operations. Yet, economists view all such mechanisms as simply devices for defending collective interests, not as methods to ensure accountability. Bureaus develop ideologies, they argued, to rationalize their activities and insulate them from external scrutiny or criticism. All agencies develop peculiar cultures that enhance member identification with their activities. But none of these devices can substitute for authoritative guidance. Employing ideology to guide detailed policy making risks ossification that inhibits creative responses to new problems or circumstances. The strengthening of agency culture will exacerbate the inherent tendency toward organizational rigidity.

For one organization to monitor another or to coordinate the activity of multiple agencies also requires hierarchical interorganizational relations to overcome resistance from staff in the subordinate bureaus. The law of countercontrol combined with the stress on hierarchy produces duplication because any attempt to control one large organization tends to generate another new bureau. Of course, controlling organizations need not become as large as implementing agencies. But the law does mean that contrasts between

“line” and “staff” organizations often become blurred as supervisory agencies absorb increasing resources. A continuous spiral is likely as top-down control efforts are followed by subordinate agency evasion, which, in turn, produces renewed attempts to secure control.

In addition, strong internal pressures push monitoring bureaus to extend their control over subordinate agencies. The logic is that by placing one agency in a supervisory position over another promotes an “ever-expanding control” whereby the quantity and detail of reporting required by monitoring bureaus tends to rise steadily over time, regardless of the amount or nature of the activity being monitored. Agencies can also be significantly affected by their suppliers, by clients (interest groups or firms that benefit from their activities), and by those entities whose behavior they regulate. A “territorial drive” in all bureaus leads them to define core functions central to their organization’s mission and the more peripheral activities that surround this basic core and are used to insulate it from erosion or attack.

Most economic analyses of bureaucracies view them as uniformly hierarchical organizations for two reasons. First, early authors in the field posited a fundamental contrast between organizations run in a centralized command mode and markets, which are viewed as decentralized preference systems. “In the absence of profit goals, bureaus must be centrally managed by the pervasive regulation and monitoring of the activity of subordinates (von Mises).” For many polemicists, such as Hayek, it became axiomatic that government agencies in the West were comparable in their operation with the bureaucracies of communist states, particularly in their reliance on centralized planning and management systems. Second, methodological assumptions carried over from the neoclassical theory of the firm invariably included an undifferentiated management team that makes unitary rational decisions and is capable of centralized control over the organization.

By collapsing all non-market mechanisms of coordination into voting or command and control, economists exclude other potentially important options, such as:

- Incremental agreements that “balance interests”
- Behavior modification that alters motives
- Interactive “mutual adjustment”
- Participation, delegation, and salesmanship
- Voluntary cooperation between self-interested actors
- Non-hierarchical “kinship” systems of “mutual service”
- Persuasion through logic, charm, and reasoned argument
- Economic and ideological hegemony by a dominant class

- Consensus and accommodation among multiple stakeholders

This unitary control orientation rejects or ignores any solidaristic behavior, public interest orientation, or altruistic values by public sector managers or employees. Within this general framework, only Downs provides a complex, pluralistic perspective of motivation in the public sector. Unless economic theorists can show why these options do not function or will not work, their propositions about behavior in bureaus become tautological, claiming only that “hierarchical organizations must be run hierarchically.”

Variations in institutional form and function will produce differences in agency behavior and influence. Moreover, relationships between agencies are qualitatively different from the relationships between ranks, sections, or groups of actors within one relatively unified institution. These two propositions will seem obvious to a political scientist, but apparently not to those who write public choice accounts. Using simply specified parameters (e.g., perfect competition, oligopoly, or monopoly), economists can move easily from their analysis of how one enterprise behaves in a market to conclusions about the behavior of the industry as a whole. Within this tradition, writers seem bound to the claim that once they understand how “the bureau” works, they have also ipso facto discovered how the entire state apparatus works. Aggregate behavior across bureaus is typically described only in passing. Economic theorists usually concentrate their efforts on demonstrating why distortions in the behavior of public agencies stimulate their growth.

The conclusion they draw is that the most important government problem is the unnecessary and continuous expansion of its bureaucracy. From this perspective, the most corrupting impact of the behavior of top public officials is their tendency to exploit their positions to increase the size of their agencies so that they become much larger, less accountable, and more expensive than they need to be. This general proposition that government bureaus are continuously expanding monopolies leads directly to reform proposals that emphasize encouraging public entrepreneurship and the promotion of competition between bureaus. Thus, political economists favor such diverse measures as bonuses for bureau heads from the savings they uncover, breaking up large centralized agencies into regional entities, allowing bureaucratic duplication in the provision of public goods, and privatizing or contracting out many governmental activities.

Classical management theory envisioned public sector bureaucracy as a modernizing form of organization that would ensure precision, speed, reliability, clarity, efficiency, and regularity in the provision of high-quality goods and services. In contrast, modern

political economists have developed theories of bureaucracy, which see, in these same hierarchical structures, institutions that foster self-serving behavior that corrupts the natural functioning of markets and that masks the political pursuit of private gain by those who manipulate them. Self-interest in the private sector is said to produce optimal outcomes, but in all government agencies this same motivation generates command-and-control monopolies operated by an intrusive, problem-finding elite that overemploys, overpays, overbuilds, and overproduces. Thus, whenever possible, the reliance on the public sector should be reduced and market forces restored.

BIBLIOGRAPHY

1. Buchanan, J. Social choice, democracy, and free markets. *J. Polit. Econ.* **1954**, *LXII*, 114–123, April.
2. Buchanan, J. Why does government grow? In *Budgets and Bureaucrats*; Borcharding, T., Ed.; Duke: Durham, 1977.
3. Downs, A. *An Economic Theory of Democracy*; Harper & Row: New York, 1957.
4. Downs, A. Why the government budget is too small in a democracy. *World Polit.* **1960**, *XII*, 541–563, July.
5. Downs, A. *Inside Bureaucracy*; Rand: Prospect Heights, 1967.
6. Friedman, J. *The Rational Choice Controversy*; Yale: New Haven, 1995.
7. Green, D.; Shapiro, I. *Pathologies of Rational Choice Theory*; Yale: New Haven, 1994.
8. Hayek, F. *Individualism and Economic Order*; Chicago: Chicago, 1948.
9. Knott, J.; Baybrooke, R. *A Strategy of Decision*; Free Press: New York, 1965.
10. Knott, J.; Miller, G. *Reforming Bureaucracy*; Prentice-Hall: Englewood Cliffs, 1987.
11. Mitchell, W.; Simmons, R. *Beyond Politics*; Westview: Boulder, 1994.
12. Niskanen, W. *Bureaucracy and Representative Government*; Aldine: Chicago, 1971.
13. Ostrom, V. Public choice. *Public Adm. Rev.* **1971**, *31* (2), 203–216.
14. Ostrom, V. *The Intellectual Crisis in American Public Administration*; Alabama: Tuscaloosa, 1974.
15. Savas, E. *Privatization*; Chatham House: Chatham, 1987.
16. Stevens, J. *The Economics of Collective Choice*; Westview: Boulder, 1993.
17. Tullock, G. *The Politics of Bureaucracy*; Public Affairs: Washington, DC, 1965.
18. Tullock, G. *Toward a Mathematics of Politics*; Michigan: Ann Arbor, 1967.
19. Tullock, G.; McKenzie, R. *The New World of Economics*; Irwin: Homewood, 1975.
20. von Mises, L. *Bureaucracy*; Yale: New Haven, 1944.
21. Weber, M. *Theory of Social and Economic Organization*; Free Press: New York, 1966.
22. Weber, M. *Essays in Sociology*; Oxford: New York, 1968.

Bureaucracy: Perspectives

Larkin S. Dudley

Center for Public Administration and Policy, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, U.S.A.

INTRODUCTION

A brief encyclopedia article cannot capture the vast cornucopia of themes and studies that surround the concept of bureaucracy. To introduce some of the bureaucratic literature pertinent to public administration, this entry restricts itself to the following themes: the historical context of bureaucratic scholarship; characteristics of the organizational form; public bureaucracies in their political environment; public bureaucracies in an economic framework; and the case against and for bureaucracy.

THE STUDY OF BUREAUCRACY: HISTORICAL CONTEXT

The term bureaucracy is believed to have its origin as a reference to a cloth covering the desks of French government officials in the eighteenth century.^[1] Early usage referred to an official workplace (bureau), in which individual activities were routinely determined by explicit rules and regulations. From a plethora of conceptions, such as a synonym for civil service to a pejorative term for inefficient and corrupt systems, the usage that has dominated the social sciences is bureaucracy as a form of organization.

The work of Karl Marx and Alexis de Tocqueville documented the historical trend toward increasing bureaucratization throughout modern Western Europe. However, most link the term and tradition of study of bureaucracy to the intellectual legacy of Max Weber, who conducted his study at the turn of the twentieth century, but whose work was not available widely to English-speaking scholars until later translations in the 1940s.^[2] Weber's conception of bureaucracy is understandable in terms of his focus on the legitimization of power in societies, including Chinese, Egyptian, Roman, Prussian, and French administrative systems.^[3]

In Weber's analysis, the bureaucratic form of administration reflected one way in which power could be legitimated. Weber argued that power was the ability to direct others' behavior, regardless of whether those being directed believed it was proper

and acceptable.^[4] Authority, however, represented the legitimate use of power, power believed to be proper by those whose activities were being directed. Three different types of authority were then identified: traditional; charismatic; and legal-rational.^[4] Under a system of traditional authority, the use of power is believed to be legitimate because of belief in time-honored traditions and patterns of behavior, while charismatic authority is imbued in individuals who exhibit exceptional or heroic characteristics. Rational-legal authority derives from belief in the legitimacy of law, including the belief that rules and the authority of officials to perform mandated duties are legitimate. Thus, the legal authorization of individuals to behave in a certain manner (e.g., the use of coercion) is a crucial dimension under rational-legal authority.

According to Weber's historical analysis, distinctive types of administrative systems are associated with each type of authority. Rational-legal systems of authority stress impersonal rules and regulations, exemplified in modern rationalistic bureaucracy.^[5] However, the Weberian model of bureaucracy has often been misunderstood and erroneously explained. One source of the confusion is that Weber conceived of his model of bureaucracy as an "ideal type."^[1,6] Ideal in this sense does not mean desirable but, instead, means a simplification of many features to more clearly understand the essence of a construct. Weber understood that actual bureaucratic organizations might exhibit only some of these characteristics or might possess them in varying degrees. The construct identifies bureaucracy by a certain structure and certain characteristics, but the meaning of the term is neither pejorative nor complimentary but instead determined by the context.

CHARACTERISTICS OF BUREAUCRACY

In the Weberian model of bureaucracy, the importance of rules is paramount. Rules become the basis for decisions, displacing individual judgment, and the means to efficiently achieve specific organizational goals. Weber himself was skeptical of bureaucracy and even pessimistic that the bureaucratization of societies would devour the relationships connected to our values.

Following Weber, distinguishing characteristics of bureaucracies include:^[2]

1. *Division of labor*: A clearly defined division of labor with specialization and training for assigned tasks.
2. *Hierarchy*: Authority relations are hierarchically ordered with definite limited areas of responsibility for superordinate and subordinate personnel.
3. *Rationality of rules*: Explicit rules are designed to ensure uniformity of performance and treatment, displacing emotional and personal decision making.
4. *Technical expertise*: Individuals are recruited and promoted on the basis of technical expertise and competence, rather than on personal relationships or luck.
5. *General rules*: Prior publication and general imposition of rules were believed to replace administrative systems of unpredictable arbitrary decisions.
6. *Written documentation*: Official records become important because documentation would make it possible to trace actions and have them inspected by others.
7. *Full and continuous employment*: Career expectations are to be fulfilled within the organization in a full-time capacity. This assumes that fulfillment of one's duty will be the highest priority, thus assuring that clients' legitimate needs will be met if at all possible.

In addition, full and continuous employment contributes to the ability to develop an impersonal and impartial relationship among organization members and with those outside the bureaucracy.

Although Weber certainly saw bureaucratic organizations in relationship to their social, political, and economic environments, the most prevalent study of bureaucratic structure has tended to follow his discussion of bureaucratic characteristics. As Nickinovich, Borgatta, and Montgomery summarized, one focus of study has been the internal characteristics of bureaucracy (e.g., how organizational structure is related to technology and worker productivity, how individuals are integrated within organizations, and how the characteristics of bureaucracy identified by Weber may be inter-related).^[7] In this tradition, scholars have attempted to empirically assess the extent to which bureaucratic organizations exhibit bureaucratic properties or have inter-related dimensions. Other scholars focusing on bureaucratic characteristics, such as Hummel, articulated one of Weber's theses that humanistic values are too easily lost within the bureaucratic ethos.^[8]

One example of later work that combines a focus on the relationship between organizational structure of

bureaucracy and how operators, managers, and executives understand their duties is that of Wilson.^[9] He also furthers our understanding of the differences in private and public bureaucracies, particularly pointing out the importance of the political context in which bureaucracies operate, a subject to which we next turn.

PUBLIC BUREAUCRACIES IN THEIR POLITICAL ENVIRONMENTS

Three important themes relate the study of public bureaucracies to the political environment and political structures of which they are a part. First is the theme of bureaucratic politics, where bureaucracies are seen as important actors in policy decisions. Second is that of the control of the political system over bureaucratic choices and performance, and third is that of the relationship of bureaucracy and democracy, particularly in terms of being representative and responsive.

Two important works more than a decade apart illuminated the importance of the bureaucracy in relation to its political environment. In 1973, the bureaucratic politics approach was singled out by Allison as one of several lenses to view the Cuban missile crisis.^[10] Allison's ability to probe the relationship of bureaucratic processes and interests to the making of foreign policy challenged a view of rational policy making, determined by a monolithic entity that is goal directed and intentional. Instead, as summarized by Allison, the bureaucratic politics model sees policy making as being influenced by many players who act in terms of various conceptions of national, organization, and personal goals in the politics of negotiation and bargaining. The public bureaucracies involved in foreign policy were seen as having their own interests of budgets, power, prestige, and access to the White House, interests that they sought to enhance and protect. Players in the process were viewed as having relative influence in terms of personality and skill and in terms of where they were hierarchically in the chain of command.

The influence of the political system on bureaucracy is a different theme, one exemplified in the publication of Wilson's *Bureaucracy, What Government Agencies Do and Why They Do It* in 1989.^[9] Wilson stressed the need to understand the political context. He argued that much of what agencies do was defined from the outside in—by the political structure and culture in which agencies operate. Through several chapters on the relationship of the bureaucracy to the presidency and to Congress, Wilson carefully laid out the difficulties with a political system that asked agencies to serve a multitude of ends and a variety of masters. He noted that public officials and operators must often manage situations over which they have little control, on the

basis of a poorly defined mission and in the face of a complex array of growing constraints.

Finally, the theme of bureaucracy within a democracy has been undertaken by many serious scholars of public administration (e.g., see the scholarly works of Waldo, Kaufmann, and DeLeon).^[11] One manifestation of that theme has been the interest of scholars in both the normative and the empirical study of representative bureaucracy.^[12] This interest is premised on the idea that insofar as the personnel in a public bureaucracy reflect those of its jurisdiction in characteristics such as race, ethnicity, and gender, the bureaucracy will be more responsive. Authors argue that, through ensuring that all politically significant interests and values are represented in the formulation and implementation of policies and programs, the public interest will be better served. A second idea has been that of bureaucratic responsiveness. Scholars have examined the extent to which decisions made in public bureaucracies reflect the preferences of elected officials, interest groups, and private citizens.^[13] This idea has become even more pronounced in literature using an economic framework as discussed in the following section.

BUREAUCRACY IN AN ECONOMIC FRAMEWORK

The 1980s and 1990s brought increased attention to economic-based approaches to describe some of the themes of bureaucratic politics and the relationship of the bureaucracy to its political environment.^[14] Economic theory has stressed that rather than a focus on workers hired for their skills and controlled through authority, as in the traditional concept of bureaucracy, workers should be conceptualized as self-interested. The most noted economic approach, principal-agent theory, views organizations as relationships between a principal, who has a job to be done, and an agent, who agrees to do the job in exchange for compensation.^[15] In such relationships, principals and agents alike seek their self-interest. Although the approach focuses on the driving force of self-interest, it also helps to explain organizational pathologies. For example, information asymmetries plague relationships so that principals may not know if they have selected the best agents or may not be able to tell when agents' work matches the terms of contracts.

Agency theory has influenced both the theories of bureaucratic outcomes and the theories of institutional choice. The bureaucratic-outcome approach questions the traditional theme of bureaucratic politics that emphasizes how bureaucracies resist change.^[16] Instead, principal-agent theory suggests that bureaucracies can be viewed as agents for the principals' (elected

officials) wishes. Thus, part of the research focus has been on the principals' design of incentives and sanctions to enhance their control. As such, principal agency theory maintains elected leaders in the presidency and Congress can and do shape bureaucratic behavior in systematic ways.

Building also upon agency theory and the bureaucratic politics tradition, scholars identified as working within institutional choice theory also emphasize the relationships among agents in the regulatory bureaucracy and their political principals (e.g., legislators, presidency, and interest groups). However, they attempt to model these interactions formally to illustrate how bureaucracies are acted upon, as well as how they perform as actors.^[17] As such, these scholars argue that the design of bureaucracy does not promote efficiency but, instead, is the result of rules that are the result of political forces. Thus, organizations reflect the power of political interests. As Kettl explained, these political interests are understood as selecting among organizational options for the improvement of results, thus giving the approach the name of institutional choice.^[14]

Even more formal models, often within a game theory approach, have emphasized modeling the relationships between the bureaucrats and the rest of the system as a bargaining process. Conflicting arguments about the relationships between the bureaucrats and the rest of the system include one argument that the bureaucracy exploits its information advantages over politicians, leading to bureaucratic budgets too large, bureaucrats too numerous, and bureaucratic output too abundant, as in the work of Niskanen.^[18] Others contend that information asymmetries may actually favor politicians who can control the behavior of bureaucrats.

THE CASE AGAINST AND FOR BUREAUCRACY

A final theme in this brief overview of bureaucracy is that most academic treatments of bureaucracy contain not so hidden biases. Most studies are highly critical, according to Goodsell, who surveys the many kinds of criticisms by grouping them under three broad categories.^[19] The first criticisms Goodsell reviews are those that focuses on unacceptable performance, including arguments of general distrust of public ownership, fear of budget maximization, and the rigidity of bureaucratic forms that produce goal displacement along with other pathologies. A second theme Goodsell sees is the accusation that bureaucracies mobilize dangerous political power; that is, they may be totalitarian or, at the least, elitist in distorting the democratic political process. Finally, Goodsell notes that many academics condemn bureaucracy for its oppressive

qualities. In other words, bureaucracies are portrayed as treating humans, at a minimum, in a callous manner and are often seen as crushing employees and clients. Finally, Goodsell states that a considerable literature has now emerged that attempts to assess bureaucratic performance and power in a democracy in a balanced fashion, pointing out the many positive contributions of bureaucracy as well.

CONCLUSION

Scholarship on bureaucracy will continue to develop somewhat around the themes surrounding the internal characteristics of bureaucracy, particularly concerning structure and performance. However, as is true of organizational theories in general, more literature will focus on how bureaucracy affects and is affected by its environment, particularly its political environment in the case of public bureaucracies. One can speculate that research will also highlight the changing role of bureaucracies in a networked and globalized world. Such scholarship will emphasize that bureaucracies are one type of organizations among others that are a mix of traditional and network forms in private, non-profit, and public sectors, all involved in governing.^[20]

REFERENCES

1. Bendix, R.; Sills, D.L., Eds. Bureaucracy. In *International Encyclopedia of the Social Sciences*; Crowell, Collier and Macmillan, Inc.: New York, 1968.
2. Weber, M. Bureaucracy. In *From Max Weber: Essays in Sociology*; Gerth, H.H., Mills, C.W., Eds.; Oxford University Press: New York, 1946.
3. Weber, M. *The Religion of China, Confucianism and Taoism*; Free Press: Glencoe, IL, 1951.
4. Weber, M.; Shils, E.; Rheinstein, M., Eds. *Law in Economy and Society*; Harvard University Press: Cambridge, MA, 1954.
5. Turner, J.; Beeghly, L.; Powers, C. *The Emergence of Sociological Theory*; Wadsworth Publishing Company: Belmont, CA, 1998; 137–188; Mouzelis, N.P. *Organization and Bureaucracy: An Analysis of Modern Theories*; Aldine Publishing Company: New York, 1979.
6. Hummel, R.; Shafritz, J., Eds. Bureaucracy. In *International Encyclopedia of Public Policy and Administration*; Westview Press: Boulder, CO, 1998.
7. Nickinovich, D.; Borgatta, E.; Montgomery, R.J.V., Eds. Bureaucracy. In *Encyclopedia of Sociology*; MacMillan Reference: New York, 2000; 229–236.
8. Hummel, R.P. *The Bureaucratic Experience*; St. Martin's Press: New York, 1982.
9. Wilson, J.Q. *Bureaucracy, What Government Agencies Do and Why They Do It*; Basic Books: New York, 1989.
10. Allison, G.T. *Essence of Decision*; Little, Brown: Boston, 1971; Art, R.J. Bureaucratic politics and American foreign policy: a critique. *Policy Sci.* **1973**, *4*, 467–490; Bendor, J.; Hammond, T. Rethinking Allison's models. *Am. Polit. Sci. Rev.* **1992**, *86*, 301–322.
11. Waldo, D. *The Administrative State: A Study of the Political Theory of American Public Administration*. Holmes & Meier Publishers: New York, 1984; Kaufman, H. Major Players: Bureaucracies in American Government. *PAR* 2001, *61*, 18; DeLeon, P. *Democracy and the Policy Sciences*. State University of New York Press: Albany, 1997.
12. Greene, V.; Selden, S.C.; Brewer, G. Measuring power and presence: bureaucratic representation in the American states. *J. Public Admin. Res. Theory* **2001**, *11*, 379.
13. Verba, S.; Nie, N.H. *Participation in America: Political Democracy and Social Equality*; Harper and Row: New York, 1972; Bendor, J.; Moe, T.M. An adaptive model of bureaucratic politics. *Am. Polit. Sci. Rev.* **1985**, *79*, 755–774.
14. Kettl, D.F. Public administration at the millennium: the state of the field. *J. Public Admin. Res. Theory* **2000**, *10*, 7.
15. Williamson, O.E. *Markets and Hierarchies: Analysis and Antitrust Implications*; Free Press: New York, 1975.
16. Wood, B.D.; Waterman, R.W. The dynamics of political control of the bureaucracy. *Am. Polit. Sci. Rev.* **1991**, *85*, 801–828.
17. Moe, T.M.; Williamson, O.E., Eds. The politics of structural choice: toward a theory of public bureaucracy. In *Organization Theory; From Chester Barnard to the Present and Beyond*; Oxford University Press: New York, 1995.
18. Niskanen, W.A. *Bureaucracy and Representative Government*; Aldine: Chicago, 1971.
19. Goodsell, C.T. *The Case for Bureaucracy: A Public Administration Polemic*; Chatham House Publishers, Inc.: Chatham, NJ, 1994.
20. Meier, K.J.; O'Toole, L.J., Jr. *Bureaucracy in a Democratic State: A Governance Perspective*; The John Hopkins University Press: Baltimore, MD, 2006.

Bureaucratic Agencies: Delegation of Authority

Scott R. Furlong

*Public and Environmental Affairs, University of Wisconsin-Green Bay,
Green Bay, Wisconsin, U.S.A.*

INTRODUCTION

Throughout the history of the United States, there has been an evolution in the role that the federal government plays in our daily lives. For much of our history, this role was somewhat limited. Many point to Article I, section 8 of the Constitution, the enumerated powers section, in describing much of what the federal government did. These powers include, for example, the power to lay taxes, regulate commerce, coin money, and many others. This evolution of the role of the federal government has also occurred within the executive branch itself. The role of the bureaucracy in our daily lives is much different than it was 200 year ago, or even 50 year ago. The expanded role of the federal government, specifically the bureaucracy, is interconnected and has been facilitated by the idea of delegation of authority in the United States. This article examines the evolution of delegation of authority or power, especially since the 1940s and how this evolution has had a major impact on public policy making in the United States.

ISSUES OF DELEGATION OF AUTHORITY

Delegation of authority or power is simply the Congress providing the bureaucracy with law-making powers that are typically held by itself. Congress, for various reasons, has decided that most of the specifics in policy making should be left to the agencies to develop and has increasingly written their statutes in such a way as to provide agencies with expanded authority. This delegation raises numerous issues that political and legal scholars have been debating for years.

One of the primary arguments against delegation of authority concerns the constitutional issue of separation of powers. According to this major component of the Constitution, power is to be divided among the three branches of government. The legislative branch would be responsible for making the law, the executive branch for executing the law, and the judicial branch for interpreting the law. Much of what Congress delegates to bureaucratic agencies is “law-making” power. In other words, when agencies develop regulatory policy that dictates or prevents

some forms of behavior, that is, in a sense, law making. In addition, these same agencies are often given the primary responsibility to interpret and adjudicate challenges against their policies and thus have judicial power as well. Many consider these forms of delegation to be blatant violations of the separation of power principles and thus unconstitutional.

Another issue raised, often in regards to delegation, concerns the issue of to whom the power is actually being delegated. By giving policy-making power to the executive branch agencies, Congress is actually allowing policy to be made by a branch of government that is somewhat removed from the general public and electoral politics. When Congress, or more specifically a member of Congress, votes for a law that we as citizens may not agree with, we have the option of voting against that individual in the next election. This representative role of a member of Congress is the basis of our representative democracy. When we believe that our congressperson is not representing us, we can vote for someone else. But that option is not available when someone within the bureaucracy makes a policy. Bureaucratic officials are either appointed or, more likely, hired based on merit considerations. Voters have some minimal electoral recourse regarding an appointed official in that they could hold the president accountable for the people appointed (although unlikely). If a career official from the Occupational Safety and Health Administration (OSHA) decides to issue a regulation requiring more stringent indoor air quality in the workplace and a person disagrees, he or she does not have the option to vote for another OSHA official. There is no direct representation function and therefore no electoral consequences.

Another issue raised by those concerned about delegation of authority concerns the relationship between executive agencies and interest groups. The issue arises if one is concerned with issues of agency capture. According to capture theory, administrative agencies are beholden to those interests in which they have been created to regulate.^[1,2] In other words, policies adopted by agencies are heavily influenced by the very organizations the agencies were legislated to control. If agencies are captured by the interests they should be regulating, then one has to wonder about the desirability of delegation of authority. As delegation

increases and agencies become responsible for developing more public policy, organizations have more opportunities to dictate the terms of these policies, especially if it is a captured agency. Critics of delegation argue that this provides advantages to the best-organized groups, and not what might be the best public policy.^[3] Others state that delegation eliminates accountability of our elected leaders.^[4]

A potential example of a captured agency was the Civil Aeronautics Board (CAB), which was responsible for the setting of airline rates for the industry. Yet, many believed that the CAB was basically carrying out the desires of the airline industry itself and not necessarily protecting consumers from monopolistic-type practices. This issue was one of the reasons that the CAB was disbanded when deregulation occurred in the airline industry in the 1970s.

Of course, positive issues are also associated with the delegation of authority that many scholars recognize. One of the primary arguments in support of delegation is that Congress is delegating policy making to the experts in the field.^[5] These agency officials can develop, use, and understand the best information (science, economics, risk, etc.) available in the making of the policy. In addition, agency bureaucrats, unlike members of Congress, should be further removed from the political interests associated with policy making because they do not have to worry about re-election.

Another argument made by those supporting delegation relates to the simple time and resource constraints faced by our elected leaders to be able to develop specific policies. Given the growth of the federal role in our daily lives, it would be difficult, if not impossible to expect our elected leaders to be able to develop the vast amount of policy necessary in our daily lives. Of course, this might be an argument in support of a major de-emphasis of the role of the federal government, but that is perhaps a topic for another article. It is unlikely that we will see that type of reduction in the role of the federal government at any time.

WHY DELEGATION OF AUTHORITY?

An important question to ask when discussing delegation of authority is why would Congress willingly abdicate its law-making powers and give it away to the bureaucracy. According to the Constitution, Congress is the law-making branch of government, which means that it has certain authority in our system of government. Why would Congress as an institution be so willing to share this authority with another branch?

Fiorina discussed two major theories as to why Congress may be so willing to delegate authority to

the bureaucracy, the “good government” and the “shift-the-responsibility” models of delegation.^[6] Under the “good government” model, Congress delegates for various reasons associated with the making of better public policy. Delegation allows policy to be made by, theoretically, nonpolitical actors who are experts in the field. Because they are not beholden to constituents through elections or the campaign money provided by special interests, the bureaucracy does not have to be concerned about these outside influences in policy development. They can select the most rational policy and not worry about the political fall out.

In addition, people working within the bureaucracy are careerists and have much information and expertise in the specific policy area. Therefore, Congress delegates to take advantage of this expertise in policy development. We do the same in our daily lives. We seek out doctors, auto mechanics, and electricians to do certain jobs that they are trained to do. Should we do anything less with the development of our nation’s policies? The final element of the “good government” model is that the bureaucracy may have more discretion and flexibility to make changes to a policy given changing circumstances. Congress can write a vague law and allow the agency to apply it to the appropriate situations. As technology, science, or the economic conditions change, the bureaucracy can make changes to a policy more easily, and typically more quickly, than the legislature. These reasons would then support the delegation of authority because the end result should be a better-construed policy, the interpretation of which can be readjusted as necessary.

Fiorina’s other model, the “shift-the-responsibility” model, appears to be a pessimistic view of why Congress delegates policy-making power to the bureaucracy. Under this scenario, Congress delegates for numerous reasons. One reason is that Congress does not have the necessary time and resources to develop specific laws, so they pass general laws and allow agencies to fill in the details, which may be quite extensive. Legislators would rather not spend time on the details of policy for many reasons. First, more conflict will likely arise between colleagues in hashing out the details of a policy. It is sometimes relatively easy to agree on the broad ideas or goals of a policy (e.g., a policy to improve air quality), but how the nation gets to that goal will generate a lot of conflict. In other words, the devil is in the detail. Second, working on the details of legislation will take more time and effort. Delegation allows Congress to potentially pass more legislation because they let agencies work out the specifics of policy, thus freeing congressional time and resources. In addition, they can use the time to do other congressional duties, such as constituent service, to help in their re-election goals.

Politically, the “shift-the-responsibility” model is also rational for members of Congress. By delegating decision making to agencies, Congress can avoid the responsibility of making a “bad” decision. The ultimate decisions that will affect citizens will come out of the bureaucracy through regulations or other administrative policy tools. People may complain about these decisions, but Congress can shield itself from these complaints by basically passing the buck to the bureaucracy. Not only can Congress avoid the responsibility for these decisions but they are also in a position to take advantage of the bad decision and make some political points. Complaints about a policy decision will come to a member of Congress, who can then look into the decision for a constituent through their casework networks. This member of Congress can serve their constituent by fixing this problem created by the agency decision and then claim credit for being a responsive legislator. These behaviors and activities by members of Congress follow closely with Mayhew’s premises discussed in *Congress: The Electoral Connection*, as well as Fiorina’s in *Congress: Keystone of the Washington Establishment*.^[7,8]

Between these two models for why Congress may delegate to the bureaucracy, there are many potential reasons why they are so willing to abdicate their power and responsibilities in policy making. Of course, varying situations may lead to different reasons as to why Congress delegates authority to the bureaucracy. In some cases, a lack of information and a need to use the experts available in an agency might represent a good and perhaps appropriate decision to delegate. In other cases, delegation may occur simply because members of Congress cannot agree on the specifics of a policy; therefore, they pass a general policy direction and give the agencies the responsibility of actually making the law. Whatever the reason, delegation is accepted by many as a fact of life in today’s governmental system. Congress has given agencies broad discretionary authority to make a policy that affects our daily lives, but it has not always been that way. Delegation of authority has gone through some evolution, especially since the 1930s.

EVOLUTION OF DELEGATION

The best way to discuss the evolution of delegation is to start with the definition of the delegation doctrine. The doctrine states, “congressional delegations of power are constitutional so long as they are accompanied by sufficient standards or guidelines so that the executive branch’s exercise of legislative power is not unbridled.”^[9] Note that the doctrine does not prohibit delegation of power, but rather states that delegations need to occur with “sufficient standards or guidelines,”

in other words, rules by which an agency is governed in its use of the delegation. For a large part of our nation’s history, the courts upheld the delegations that Congress provided to bureaucratic agencies due to the standards provided. However, the New Deal era ushered in a tremendous expansion of federal government agencies and responsibilities that brought the delegation issue to a head. This expansion of government authority is most directly tied to the enumerated power of regulating commerce among the states. Because so many of our goods and services are transported between state borders, now more than ever but even in the early-1900s, Congress could use this power to govern many activities.

The Supreme Court became involved in the issue of delegation during the New Deal, specifically in the ruling of three separate cases, in which the court ruled that the congressional delegation was unconstitutional. The three cases, *Panama Refining Co. v. Ryan*, *A.L.A. Schechter Poultry Corp. v. United States*, and *Carter v. Carter Coal Co.*, examined the issue of congressional delegation.^[10–12] In *Panama Oil*, the court ruled that the delegation that authorized the president to prohibit interstate shipments of contraband oil was overly broad. The court ruled in the *Schechter* case that there were not clear policy directives from the legislature in the issuance of “codes of fair competition” for particular industries. The issues were similar in *Carter Coal*, with the additional concern that the delegation had, in effect, been made to committees of industry representatives and not government officials.^[13] In general, the court was very concerned about the vagueness of these delegations and the amount of authority provided to nonlegislative personnel to make policy.

This time period was also one of the last times where the court ruled so conservatively on the delegation doctrine. President Roosevelt, whose policies these court rulings were overruling, attempted to increase the size of the Supreme Court in response to the court’s actions. His hope was to add additional justices, with beliefs similar to his own, to the current slate of justices to get more favorable rulings in the future. Although Roosevelt’s court packing scheme failed, a strong message was sent to the court. Some justices shifted their voting, and others retired, giving Roosevelt the opportunity to fill those vacant positions.^[14]

Since these three cases, the court has been much more accepting of congressional delegations of power, and in many cases, these delegations have been much broader and with less guidelines than the ones in the New Deal cases. The gradual acceptance of delegation made it clear that administrative agencies would play a much larger role in the development of public policy. This was not lost on Congress, which in 1946 passed the Administrative Procedure Act as one tool to control

the expanding power of agencies brought on by delegation and the use of the rule-making process.

The 1960s and 1970s saw another large increase in federal government involvement as Congress passed a wide range of regulatory laws focused on areas of public health, safety, and the environment. Laws such as the Clean Air Act, Consumer Product Safety Act, and the OSHA all delegated a tremendous amount of policy-making authority to the administrative agencies responsible for carrying out these laws. The vagueness and lack of guidelines associated with these laws is quite telling in the delegation controversy. For example, the OSHA directs the Secretary of Labor to issue regulations that requires employers to protect their workers “to the extent feasible” from harm in the workplace. This delegation provides OSHA with a wide range of options in which to interpret and implement the law. In the Supreme Court case of *Industrial Union Department, AFL-CIO v. American Petroleum Institute*,^[15] otherwise known as the *Benzene* case, this form of delegation was called into question by one of the justices in the majority. Justice Rehnquist believed that the statute represented an unconstitutional delegation of authority to the secretary. He restated his position in another case involving an OSHA rule, *American Textile Manufacturers Institute v. Donovan*.^[16] Although, the concern over a overly broad delegation was not the basis of the Supreme Court’s ruling on these cases, the issue of what constituted an appropriate delegation was once again being raised by the court.

The election of Ronald Reagan and then George H. Bush also led to the appointment of more conservative Supreme Court justices that would likely be more receptive to the idea of the non-delegation doctrine. Although those opposing broad delegations of power to agencies would perhaps have reasons to cheer when you examine the current Supreme Court, it has to date not gone so far as to reinstate the non-delegation doctrine and has also upheld congressional delegations. The court continues to follow a principle in administrative law that if Congress does not speak clearly about a particular policy, then a wide latitude is provided to agencies in their interpretation and implementation of the law. Such was the case in *Chevron U.S.A., Inc. v. NRDC*, where the court allowed for the Environmental Protection Agency’s interpretation of what constituted a stationary source, as discussed in the Clean Air Act.^[17] However, if Congress is clear on issues of agency jurisdiction, then Congress has the last word.

This last point was the issue in a Supreme Court case *Food and Drug Administration v. Brown and Williamson Corp.*, where the Food and Drug Administration (FDA) attempted to regulate tobacco products under the authority granted to it by the Food, Drug,

and Cosmetic Act.^[18] The court ruled that Congress was clear in developing the jurisdiction of the FDA in that it did not have the authority to regulate tobacco. Although the basis of the ruling is not specifically an issue of undue delegation, it is clear that the result of the ruling limited the FDA’s authority in interpreting a statute as they see fit. In fact, news accounts of this court ruling explicitly discussed the issues of delegation and agency power in their reporting and also referenced another court case concerning the role of delegation more specifically.

This second case dealt with the EPA’s ability to set clean air standards under the authority granted to it by the Clean Air Act. Upon challenge, a DC Circuit Court of Appeals struck down the agency’s ozone and small airborne particle standards under the non-delegation doctrine. The appeals court believed that EPA lacked any criteria in developing their standards and thus raised the issue of unbridled discretion. This ruling by the appeals court brought the issue of the non-delegation doctrine explicitly to the forefront on the debate. In addition, it raised some very real substantive concerns regarding environmental policy making, as well as other regulatory policy making within other agencies, because of the nature of many of the regulatory laws passed by Congress that had similar levels of non-specificity. Although it appeared that the Supreme Court might have been ready to implement the non-delegation doctrine more thoroughly upon an appeal based on past rulings and opinions, it did not do so. The Supreme Court ruled, unanimously, in *Whitman v. American Trucking Associations* that EPA’s standard-setting authority was not an unconstitutional delegation of power.^[19] In fact, one of the strongest supporters of the non-delegation doctrine, Justice Scalia, wrote the opinion of the court stating, “Even in sweeping regulatory schemes we have never demanded, as the Court of Appeals did here, that statutes provide a ‘determinate criterion’ for saying ‘how much of the regulated harm is too much.’”^[19] Although the question of what might be considered unbridled delegation may still exist, the Supreme Court still seems very willing to accept the idea of broad delegations of authority to administrative agencies and to rely on agency expertise in regards to general language such as “an adequate margin of safety.”

CONCLUSION

This article examines the issue of congressional delegation of authority from various perspectives. It explored the issues surrounding delegation, reasons why Congress has chosen to delegate power to the bureaucracy, and also the evolution of delegation

within the United States. It is clear that congressional delegation is a way of life in today's society.

Agencies are very much involved in the policy-making business due to the level of broad delegation provided to them by Congress. In fact, Stewart argued that such broad delegation is necessary for them to do their job adequately.^[20] This delegation provides the bureaucracy with the power to develop policy through administrative procedures. Although Lowi^[3,21] argued that these delegations and the use of the administrative process to make policy gives interest groups undue influence, Stewart^[22] countered by saying that interest representation is fundamental to our modern bureaucracy. Increased participation by these interests allows the agency personnel to understand what alternatives are desired. The conflict that occurs between interests ultimately leads to an agency policy.

The debate over delegation of authority will likely continue. In fact, in the *Whitman v. American Trucking Associations* case, Justice Thomas wrote in a separate opinion that he would be open to consider other non-delegation challenges. So, the court has not yet closed the door on this important constitutional and policy-making question. Therefore, we can expect to see further evolution of the delegation issue in the future.

REFERENCES

1. Bernstein, M.H. *Regulating Business by Independent Commission*; Princeton University Press: Princeton, NJ, 1955.
2. Stigler, G.J. The theory of economic regulation. *Bell J. Econ. Manage. Sci.* **1971**, 2 (1), 3–21.
3. Lowi, T. *The End of Liberalism: The Second Republic of the United States*; W.W. Norton: New York, 1979.
4. Schoenbrod, D. *Power Without Responsibility: How Congress Abuses the People Through Delegation*; Yale University Press: New Haven, 1993.
5. Rourke, F. *Bureaucracy, Politics, and Public Policy*; Little Brown: Boston, 1984.
6. Fiorina, M. Legislative choice of regulatory forms: legal process or administrative process. *Public Choice* **1982**, 39, 33–66.
7. Mayhew, D. *Congress: The Electoral Connection*; Yale University Press: New Haven, 1974.
8. Fiorina, M. *Congress: Keystone of the Washington Establishment*; Yale University Press: New Haven, 1977.
9. Cann, S.J. *Administrative Law*; Sage Publications: Thousand Oaks, CA, 2002.
10. *Panama Refining Co. v. Ryan*. 293 U.S. 388, 1935.
11. *A.L.A. Schechter Poultry Corp. v. United States*. 295 U.S. 495, 1935.
12. *Carter v. Carter Coal Co.* 298 U.S. 238, 1936.
13. Gellhorn, E.; Levin, R.M. *Administrative Law and Process in a Nutshell*; West Publishing: St. Paul, 1997.
14. Kerwin, C.M. *Rulemaking: How Government Agencies Write Law and Make Policy*; Congressional Quarterly Press: Washington, DC, 1999.
15. *Industrial Union Department, AFL-CIO v. American Petroleum Institute*. 448 U.S. 607, 1980.
16. *American Textile Manufacturers Institute v. Donovan*. 452 U.S. 490, 1981.
17. *Chevron U.S.A., Inc. v. NRDC*. 467 U.S. 837, 1984.
18. *Food and Drug Administration v. Brown and Williamson Corp.* 529 U.S. 120, 2000.
19. *Whitman v. American Trucking Associations*. 531 U.S. 457, 2001.
20. Stewart, R.B. Beyond delegation. *Am. Univers. Law Rev.* **1987**, 36, 323–343.
21. Lowi, T. Two roads to serfdom: liberalism, conservatism and administrative power. *Am. Univers. Law Rev.* **1987**, 36, 295–322.
22. Stewart, R.B. The reformation of American Administrative Law. *Harvard Law Rev.* **1975**, 88, 1667–1711.

Bureaucratic Discretion: Political Control and Influence

Scott R. Furlong

Public and Environmental Affairs, University of Wisconsin-Green Bay,
Green Bay, Wisconsin, U.S.A.

INTRODUCTION

Delegation of authority is predominant in our democratic society today. Congress, when writing laws and policies, has given the bureaucracy leeway in developing policies and regulations that interpret these statutes. Political scientists often refer to this leeway as bureaucratic discretion—the amount of control that bureaucratic agencies have in making public policy. This is an issue primarily because of the information asymmetries that arise between the bureaucracy and the elected leaders of Congress and the President. Left unchecked, many believe that the bureaucracy can use information advantages to set policy that is not what our elected institutions and, ultimately, the general public desire. (See *Bureaucratic Agencies: Delegation of Authority*, p. 225.)

This essay discusses some of the primary elements used by our elected leaders in their attempts to maintain control or to influence the decisions of agencies. Within the political arena, these controls can be placed into three major categories: administrative procedures, scientific and economic decision rules, and political oversight.^[1] Elected officials use these controls to influence bureaucratic agencies in their decisions, even though these agencies have an informational advantage. While these tools are presented as separate entities, it is important to recognize that, in many cases, they are interrelated. For example, a presidential directive that requires agencies to conduct a cost-benefit analysis on their regulatory activities is a procedural requirement to use a type of economic analysis or decision rule.

ADMINISTRATIVE PROCEDURES

Administrative procedures have been one of the primary ways in which our elected leaders have attempted to control and influence bureaucratic policy making. As bureaucratic policy making became more prevalent in the mid-twentieth century, our lawmakers recognized the potential concerns and began to develop some statutory requirements to ensure a sense of fairness and participation when dealing with issues of bureaucratic discretion. The major result of this

activity was the passing of the Administrative Procedure Act (APA) in 1946. This was an important law because of the procedures required that ensured that bureaucratic agencies could not regulate out of control. Some of the major components of the law were as follows: 1) granting the right to seek judicial review to “any person suffering legal wrong because of agency action, or adversely affected or aggrieved by agency action within the meaning of the relevant statute;”^[2] and 2) setting administrative procedures for how agencies would conduct rule making and adjudication. These two parts of the APA limited what agencies could do and allowed elected leaders and general citizens to be part of the administrative process. Agencies, in general, must follow certain procedures when developing policy, which will limit their discretion. The threat of judicial review also will make agencies think seriously about their policies and the possible impacts on individuals.

A large amount of agency policy making occurs through the rule-making process, and specifically through informal or notice-and-comment rule making. The APA requires three minimum steps when agencies conduct informal rule making. First, they must provide notice of a proposed policy. This typically occurs in the *Federal Register*. Second, agencies must provide an opportunity for interested parties to comment on these proposals. Finally, agencies must consider the comments before issuing the final regulation, which is again published in the *Federal Register*.

Kerwin comments that the core elements put forth in the APA are information, participation, and accountability, and these are basic to our democracy.^[3] Information is crucial to a democratic system and to decision making in general. The APA requires that notice be provided to the public when a rule is being developed and when it becomes final. Participation is another important element in democratic decision making. In the case of the APA, agencies must allow for the submission of written comments to any rule being developed. Finally, as mentioned earlier, the provision of judicial review ensures some element of accountability that we typically associate with the electoral process—in other words, if we as a society are not happy with our policies coming out of Congress, then we can elect someone different. While these

elements are not full protections against unbridled bureaucratic discretion, they provide a minimum level of criteria that all agencies must follow. In a number of cases, our elected institutions and the agencies have developed even more stringent administrative requirements to facilitate bureaucratic control.

As Congress increased the level of regulatory policy making in the 1970s, additional procedural elements to help control bureaucratic discretion were also provided. One concrete example of this was the passage of the National Environmental Protection Act in 1969, which required environmental impact statements for all federally funded projects expected to impact the environment. Other examples have included minimum public comment periods when conducting informal rule making, requirements for public hearings, and the maintenance of a rule-making docket.^[1]

While these procedures may appear to be mundane and neutral, the writings of McCubbins, Noll, and Weingast suggest that they are anything but.^[4] According to “McNollGast,” these procedures provide political principals, in this case Congress, with the necessary tools to help monitor and perhaps influence the activities of bureaucratic agencies. In addition, interest groups with similar parochial interests will benefit from these procedures. The passing of the National Environmental Protection Act provides such an example. Organizations and politicians were interested in affecting programs from a number of different policy areas, and this procedural law gave them entrée and influence into these decisions. Requiring longer public comment periods provides similar benefits to organizations that may have had difficulty responding to a solicitation of comments from an agency in a short period of time.

So, administrative procedures have been an important tool in controlling bureaucratic discretion and influencing agency decision for a number of years. These procedures establish boundaries as to how much leeway agencies have in making decisions and can also push agencies to make decisions in a certain direction. In addition, they establish procedural protections to address the potential dangers to delegation of authority.

ECONOMIC AND SCIENTIFIC DECISION RULES

Another popular tool used in influencing agency decision making is the use of economic and scientific decision rules. This has been especially true in those policy areas dependent on a large amount of technical or economic expertise, such as health, safety, and environmental policy making. Examples abound in this area of discretionary control. According to Bryner, statutory requirements that impose “best available”

technology for pollution sources or that require agencies to focus on the hazards at hand regardless of the costs involved are the ways that technical analyses are used to limit agency discretion.^[1] A similar requirement by the Reagan Administration stated that agencies had to show that the benefits of the rule outweighed the cost imposed before issuing a regulation. Another way of limiting discretion is by using of science advisory boards that must review agency decisions prior to issuing policies.

One of the more common forms of analysis used, especially in recent history, has been some form of economic analysis. Typically, presidents have issued executive orders requiring agencies to conduct one type of analysis or another to justify the policies at hand. These types of analyses began in earnest during the Carter Administration, which required rules to discuss economic consequences of their alternatives. It was the Reagan Administration, though, that took the use of economic analysis to a new level.

Upon taking office, President Reagan issued Executive Order 12291, which was the most direct requirement for economic analysis of agency rules. Under the order, agencies were required to conduct regulatory impact analyses that showed that the potential benefits to society would outweigh the potential costs; the regulatory objective chosen would maximize net benefits; and among a series of alternatives, the one involving the least net cost to society would be chosen. The Office of Management and Budget (OMB) was given the responsibility of overseeing this requirement placed upon agencies. Requiring such analyses, and dictating how they were to be used, limited the agencies’ decision-making discretion by dictating the information they had to use when regulating in their policy area. Certain categories of policies were exempt from the decision-rule criteria included in the executive order (e.g., certain statutes do not allow cost to enter into the decision-making process). Even so, the writing was “on the wall” for agencies. Even for policies that were exempt from the benefit–cost decision rule, agencies collected economic data and information, and they were important elements of the regulatory proposals.

OVERSIGHT

The final major way that elected institutions can influence agency policy making is through a variety of oversight techniques. Oversight is typically associated with a function performed by the legislative branch, and in fact, oversight is usually discussed as those activities performed by Congress to ensure that the bureaucracy is following through on congressional intent. For purposes of controlling agency behavior, though, one can also consider the President as overseeing

the bureaucracy as well. Oversight demonstrates the interesting situation of an agency having multiple principals, or in other words, these agencies must respond to two different “bosses.” On the one hand, Congress provides statutory and budget authority to the agencies, but on the other, the President is the head of the executive branch and ultimately the top of the bureaucratic hierarchy. Agencies need to somehow respond to both actors.

Presidential Oversight

The President has a number of tools at his disposal to try to control and influence agency policy making. In some cases, these tools are based on constitutional powers, and in others, they have evolved as the powers of the presidency have evolved. The most common and well-known tool that a president can use is his appointment power. According to Article II of the Constitution, the President has the power to appoint officials to government offices. One typically thinks of the President appointing a secretary to a cabinet office such as the Secretary of Transportation, but these appointments are much more extensive. A president will want to appoint officials with similar ideological views in order to improve the chances of a certain agenda being followed by an agency. Ideally, such an official might be considered an extension of the president in the policy arena in question. President Reagan was especially successful in the use of his appointment power, placing people such as James Watt into the Department of Interior and Ann Gorsuch into the Environmental Protection Agency. These individuals shared Reagan’s conservative beliefs in regards to the government being too involved in regulating business. If an appointee fails to follow the presidential agenda, the president can remove that person in favor of someone more likely to follow the president’s views.^a Presidents are often concerned that their appointees will “go native,” meaning that they will take on the views of their agencies rather than the administration. This appointment power provides a tool for which the president may guard against this concern.

Another tool that a president can use is to reorganize agencies or agency functions. Restructuring agencies internally, or a series of agencies, has the effect of emphasizing or de-emphasizing particular areas of policy. So, for example, President Carter’s splitting off

of the Department of Education from what was once the Department of Health, Education, and Welfare focused attention on this area of policy. Likewise, President Nixon’s reorganization of many agencies to create the Environmental Protection Agency was an attempt to highlight this policy arena. A related organizational tool that presidents have used relates more to internal reorganization so that agency heads have more control over enforcement activities.^[1] Different interpretations of compliance and the strictness of the related penalties, where there is some discretion, will influence agency activities and impact the policy area. If regulated entities believe that agency enforcement activities will decrease or that the penalties will be less, then there is an incentive to perhaps not spend the resources necessary to comply with the policy.

Another tool at the president’s disposal is the use of the budget. Because the president and his economic staff initiate the budget process, they can send signals to agencies about policy direction. In addition, agency budgets are submitted to the OMB for their review of the dollars and the related programs. OMB will not look favorably on agency programs that are not within the framework of the administration’s agenda. OMB will likely suggest or dictate changes that influence agency policy making.

One final, but particularly valuable, form of presidential oversight and influence is the actual presidential management of rule making and regulatory policy. In some cases, rule-making management has been the result of executive orders issued by presidents over the past 30 yr (e.g., Reagan’s Executive Order 12291). In other cases, organizational structures within the executive branch have provided presidents (or their designates) with the ability to influence these agency actions (e.g., President George H. Bush’s Council on Competitiveness).

There have been a number of executive orders issued in the recent history that have been directly aimed at influencing agency rule-making behavior and outcomes. Many of these historically have dealt with the economic side of rule-making activities. In other words, President Ford required inflationary impact statements for each major rule, and President Reagan’s executive order requiring benefit–cost analysis has been mentioned earlier. These executive orders required the gathering and consideration of economic data in rule development and passage, but other executive orders have focused on different types of information. Executive Order 12612 required agencies to conduct a “federalism assessment” of their rules. Other executive orders required agencies to consider the impact of their policies on families, the judiciary process, and private property rights.^[3] These types of requirements serve at least two purposes for presidents wanting to influence agency policy making. First, they

^aPresidential removal of agency officials is allowed for executive departments but not for commissioners working for independent regulatory commissions, because commissioners serve for fixed terms and cannot be removed due to personal differences in policy direction (see Refs.^[5,6]).

may slow the rule-making activity in general, because agencies must collect and consider additional information. For the Reagan Administration, this was an important goal. Second, agencies now must use additional pieces of data to make their decisions, and the data being required affects the type of decision being made.

Upon taking office, President Clinton also issued an executive order concerning rule making and associated analysis. Executive Order 12866 superseded Reagan's 12291 but kept many of the same principles. In essence, Clinton's executive order, which is still the governing order under President George W. Bush, attempted to decrease the amount of conflict between agencies and the OMB by decreasing the number of rules needing to be reviewed, opening up the process more for public scrutiny, and allowing agencies to consider other areas besides benefit-cost analysis. Still, this executive order continues the trend of presidents using tools to get more involved in agency policy making.

Presidents have also centralized their management structures in recent history, which has led to their ability to influence agency policy. The OMB under the Reagan Administration is the best example of this centralization. OMB's Office of Information and Regulatory Affairs had the ability, theoretically, to review almost every rule making coming out of the executive agencies. These reviews, provided by the authority of Executive Order 12291, gave OMB access to all agency rule-making decisions. While Executive Order 12291 was clear in stating that OMB reviewed these regulations and did not have to approve them prior to their issuance, politically, agencies wanted the go-ahead from OMB prior to publishing the regulation in the *Federal Register*. Executive Order 12498 was another tool of presidential management that required agencies to provide OMB with a yearly calendar of regulatory activities, and hence, giving OMB the ability to see the upcoming agency agenda and to attempt to influence the agenda or the policies within that agenda. Executive review is now an accepted part of the regulatory process, and although President Clinton decreased the role of OMB in terms of the number of regulations reviewed, it is clear that no president will ever eliminate this form of review now that it has been created.

Presidents have used other managerial tools to help them control and influence agency policies. For example, President Carter created the Regulatory Analysis Review Group (RARG) to help improve the quality of regulatory analysis. President Reagan created the Vice President's Task Force on Regulatory Relief to consider the impacts of regulations on certain sectors of the economy.^[3] President George H. Bush, perhaps adopting from his past role as head of such a task force, created the Council on Competitiveness,

which was chaired by Vice President Quayle, to act as another level of regulatory review for selected, major-impact rules. All of these ad hoc councils and groups provided presidents with another tool with which they could attempt to influence agency rules, but not without some controversy. The Council on Competitiveness, for example, was harshly criticized for its lack of openness and secret meetings. Critics charged that the group was violating the APA and other procedural protections when having their discussions.^[7] Of course, it did not help when it was revealed that the interest groups that appeared to have entrée into the Council happened to be strong financial and electoral supporters of the Bush Administration.

The president has a number of formal and informal tools to help influence and control agency policy making. These tools have become more prevalent during the past 30 yr with the growth of the administrative presidency. During this growth, presidents have taken their role as the chief administrative officer very seriously.

Congressional Oversight

Congress also has a series of tools available to influence and control agency policy making. From the congressional standpoint, oversight is important, because it is their way of ensuring that the policies coming from the agencies are consistent with the statute initially passed. Scholars have paid much more attention to the area of congressional oversight vis-à-vis presidential oversight. Given the other duties of Congress and the personal preferences of the members, many believe that there is not much time left for congressional oversight. Others have found oversight to be prevalent, if not by the congressional members, then by the committee and personal staff of the members.^[8,9] Still others may argue that we may need to think about what types of activities constitute oversight.^[10]

Congress has many oversight tools at its disposal. One of the more common is the use of a committee or subcommittee hearing. These hearings can examine a number of different elements regarding agency activity or programs, such as whether the programs are working correctly, the budgetary issues surrounding the program, or even the efficiency of the program. While it is typical to think of oversight as occurring after the fact, it is important to recognize that committees may call hearings in order to give direction or preferences of ongoing policies. In either case, Congress has the ability to send signals to an agency about what its members would like to happen in a policy area.

Another form of this type of oversight is when Congress asks for an investigation or evaluation of an agency or its programs. The Government Accountability

Office (GAO) will often conduct such a study and then report to the committee or subcommittee that made the request. This investigation will provide more detailed information about the program that Congress will focus upon in another committee hearing with the agency.

Much congressional oversight occurs informally. Members and their staff may make phone calls, have meetings, or provide some passing commentary to agency personnel. Of course, these forms of oversight are outside of public scrutiny and can always be backed up by other, more formal, oversight tools.

Agencies will respond to oversight due to many of the formal tools discussed earlier, in addition to some others. As discussed in the presidential oversight section, budget power is shared between the President and Congress, and so Congress may turn toward the budget as a way of influencing agency policy. Increases or decreases in agency budget could send strong signals to agencies in regard to the policies they are implementing. Of course, Congress' legislative power provides, in many ways, the ultimate form of oversight. The initial enabling statute that provides agencies with their statutory authority can influence and control agency behavior. The earlier section on administrative procedures discussed some of the procedural requirements Congress has placed within certain statutes. Congress has also used statutory deadlines as a way to force agency action by a certain date. And "hammers"—provisions that take effect if an agency fails to act by a certain date—have been used. A more specific piece of legislation will also dictate certain policy directions. Theoretically, if unhappy with an agency policy, Congress can attempt to pass a law that overrules the agency.

This last comment about overruling an agency policy brings up the issue of the legislative veto. For a number of years, Congress included in some statutes the ability to veto or disallow an agency action before it went into effect. The Supreme Court in *Immigration and Naturalization Services v. Chadha* ruled the legislative veto unconstitutional due to separation of powers and bicameralism violations.^[11] The legislative veto was a potentially powerful management tool that Congress created, because it allowed broad authority to be delegated to administrative agencies but still provided Congress the ability to stop an action with which it was uncomfortable.

Since the Supreme Court's ruling in *Chadha*, Congress has attempted to recapture some of its ability to prevent agency policy with which it disagrees. In 1996, Congress passed the Congressional Review Act (CRA) as a way to circumvent some of the concerns stated in the Court's ruling. The CRA provides Congress with oversight ability, not with a legislative veto, but rather with a "report-and-wait" provision. A report-and-wait provision is one in which the agency

provides Congress its proposed action and gives Congress the opportunity to review the action before it becomes effective and to pass legislation barring its effectiveness if Congress is dissatisfied with the agency action.^[12] The CRA states that Congress can bar the effectiveness of an agency action with the passage of a joint resolution presented to the President for signature. The CRA had not been invoked until the election of George W. Bush as president, when Congress used it to review and prevent regulatory actions passed during the waning days of the Clinton Administration. It would appear that the only time this act might be used to stop an action would be during this transition phase, or if there was a large majority within Congress not from the president's political party. Symbolically though, a member of Congress may threaten to invoke the law as a way of pressuring an agency decision.

CONCLUSION

Many are concerned about the issue of bureaucratic discretion and the ability of agencies to develop policy. In many ways, the primary issue is that unelected bureaucrats are making policy, and they are not beholden to the people of the United States through an electoral process. In addition, information asymmetries provide an advantage to the specialized agencies compared to the elected leaders who tend to be generalists. But, as this article has discussed, our elected leaders have traditionally had a number of avenues in which to control these unelected officials, and over the past 50 yr, have developed additional ways to influence agency policy.

Administrative procedures, economic and scientific decision rules, and statutory requirements are a priori tools available to give direction to agency decisions as they are being made. In the event these fail to persuade agencies, there are a number of post hoc mechanisms available to Congress and the President to bring an agency or its policies in line after the fact. Will every agency policy receive the ultimate amount of scrutiny from the political principals? No, but we should have some assurance that those areas especially important to the elected leaders and their constituency groups will be adequately monitored and changed if necessary.

REFERENCES

1. Bryner, G.C. *Bureaucratic Discretion: Law and Policy in Federal Regulatory Agencies*; Pergamon: New York, 1987.

2. Administrative Procedure Act of 1946. 5 U.S.C. pp. 551–706.
3. Kerwin, C.M. *Rulemaking: How Government Agencies Write Law and Make Policy*; CQ Press: Washington, DC, 1999.
4. McCubbins, M.D.; Noll, R.G.; Weingast, B.R. Administrative procedures as instruments of political control. *J. Law Econ. Organ.* **1987**, 3 (2), 243–277.
5. *Humphrey's Executor v. United States*. U.S. 602, p. 295, 1935.
6. *Wiener v. United States*. U.S. 349, p. 357, 1958.
7. Friedman, B.D. *Regulation in the Reagan-Bush Era: The Eruption of Presidential Influence*; University of Pittsburgh Press: Pittsburgh, 1995.
8. Foreman, C.H. *Signals from the Hill: Congressional Oversight and the Challenge of Social Regulation*; Yale University Press: New Haven, 1988.
9. Aberbach, J.D. *Keeping a Watchful Eye: The Politics of Congressional Oversight*; Brooking Institution: Washington, DC, 1990.
10. McCubbins, M.D.; Schwartz, T. Congressional oversight overlooked: police patrol versus fire alarms. *Am. J. Polit. Sci.* **1984**, 28, 165–179.
11. *Immigration and Naturalization Services v. Chadha*. U.S. 919, p. 462, 1983.
12. Korn, J. *The Power of Separation: American Constitutionalism and the Myth of the Legislative Veto*; Princeton University Press: Princeton, 1996.

Bureaucratic Politics

Gene A. Brewer

*School of Public and International Affairs, University of Georgia,
Athens, Georgia, U.S.A.*

INTRODUCTION

Bureaucratic politics occurs when bureaucrats, who are non-elected government officials, enter the political arena and try to influence or shape the policy-making process. In most cases, this process is characterized by conflict and bargaining as multiple actors vie for control of the same policy space. Most theories of bureaucratic politics portray bureaucrats as self-interested and self-serving individuals who steer policy decisions in directions that protect and advance their personal interests and those of their employing agencies.

THE RISE OF BUREAUCRATIC POLITICS

Politics and administration have always been messily intertwined, but this fact has not always been clear. The rise of bureaucratic politics in the United States is a good case in point. During the Progressive era, the academic field of public administration was born, and government was reformed on the premise that politics and administration are separate spheres. However, this make-believe separation became strained and was finally rejected in the 1940s when the nation experienced two successive crises of global magnitude: the Great Depression and World War II. The scope and severity of these crises required the best efforts of political leaders and public administrators working in tandem. Thereby, scholars and practitioners discovered that politics and administration are one, and that this can be a good thing.

Since then, two broad streams of research have spotlighted the relationship between bureaucratic agencies and other participants in the policy-making process. The first stream of research has investigated the relationship between bureaucratic agencies, client groups, and special interests, while the second has focused on the relationship between bureaucratic agencies and overhead political actors such as the president and Congress. These two streams of research are described in more detail below.

Bureaucratic Agencies, Client Groups, and Special Interests

In the post-war years, scholars described bureaucratic politics as a process in which officials and agencies

become “captured” by client groups and other special interests. For example, David E. Lilenthal explained that organizing and managing the Tennessee Valley Authority (TVA) was essentially a political enterprise, and Phillip Selznick described how TVA cultivated political constituencies to accomplish its mission and ensure its survival.^[1-2] Selznick labeled this process “cooptation”—a Faustian-type bargain in which agencies receive political support from client groups and special interests by giving them a voice in the agency’s decision-making process.

These types of alliances often turned into “iron triangles”—cooperative relationships between an agency, legislative committee, and interest group. The opportunities for quid pro quo in such a relationship are obvious, as members of the legislative committee receive political support from the pressure group, while providing favorable legislation and funding to the agency. These iron triangles seemed ironclad: they dominated specific policy areas to the extent that even presidents could not penetrate them.

In the 1960s and early 1970s, economic deregulation and liberalizing social changes in American society made iron triangles and agency capture theories seem somewhat outdated. Hugh Heclo observed that iron triangles were giving way to “issue networks,” which he described as loose alliances between agencies and other parties interested in a particular policy or issue.^[3]

Bureaucratic Agencies and Overhead Political Actors

Meanwhile on the political front, there was growing concern over the president and Congress’s ability to control the bureaucracy. This concern was exacerbated during periods of divided government, which prevailed after 1968. For example, the Nixon Administration feared that liberal bureaucrats would side with Congressional Democrats and sabotage Nixon’s conservative policy agenda. Thus, the administration mounted an aggressive campaign to regain political control of the bureaucracy.^[4] Their strategies included applying litmus tests to presidential appointees and weakening the career civil service by bending merit system rules.^[5] This battle for control of the bureaucracy grew fiercer in the Carter and Reagan Administrations. President

Carter secured passage of the Civil Service Reform Act of 1978, making the bureaucracy more accessible and vulnerable to overhead political control. President Reagan seized this opportunity to protect and advance his conservative policy agenda: he significantly increased the number of political appointees, further politicized the appointments process, and instituted policy reforms that weakened the career service in most regulatory, social service, and central management agencies.^[6]

Scholarship on bureaucratic politics increasingly reflected this concern about bureaucratic control. Researchers from several different academic fields began applying the principles of neoclassical economics to the problem. Most of these treatments suggested that bureaucrats and their agencies were on a feeding frenzy, acting in their self-interest rather than the public interest.^[7] One particular application—principal-agent theory—has become quite influential in recent years.^[8] Studies using this framework have shown that elected officials can exert control over bureaucratic outputs, but that doing so is costly and time-consuming. Excessive overhead political control seems to breed bureaucratic inefficiency.^[9]

In the United States, politics and administration have always been intermingled, and bureaucratic politics has always been an important part of the policy-making process. This fact was made clear in the 1940s, when elected officials and bureaucrats worked together to rescue the nation from economic bankruptcy and despotism. Ironically, scholarship on bureaucratic politics has become increasingly disconnected from this heroic era.

SCOPE OF BUREAUCRATIC INFLUENCE

According to James Q. Wilson and John J. DiIulio Jr., three features of the American political system make American bureaucracy distinctive. These features also promote bureaucratic politics: first, political authority over the bureaucracy is shared between the executive, legislative, and judicial branches; second, agencies at the federal, state, and local levels often share functions; and third, the American political system is adversarial in nature. In this highly charged political environment, bureaucratic politics is likely to thrive.^[10]

Most observers agree that bureaucrats' preferences are important, and that their behaviors are consequential. Yet the extent of bureaucratic influence in the policy-making process is not always clear. Francis E. Rourke contends that bureaucratic influence has declined in recent years. He says the main culprits are divided government, transformed constituencies, and increased policy volatility.^[11] Robert F. Durant disagrees in part. He provides numerous examples

showing that bureaucratic influence is alive and well in some quarters, and he concludes that it is highly variable. Durant lists more than a dozen factors that affect the degree of bureaucratic influence. These factors include individual and organizational attributes, type of constituents, and variables describing the policy arena and political environment.^[12] Both authors are probably correct. Bureaucratic influence has declined in recent years, but it remains quite potent in certain situations and may reassert itself in the future.

SOURCES OF BUREAUCRATIC POWER

Bureaucrats exercise considerable discretion in the modern democratic state. This fact is inevitable and desirable. It is inevitable because administration is about exercising discretion. Even if bureaucrats choose to be neutral, they are electing a course of action with sure and important consequences. Bureaucratic discretion is desirable, because elected officials need help. Bureaucrats have the expertise to provide this help, and they are strategically situated. That is, they work in close proximity to the problems elected officials face and the people they serve.

Bureaucrats also wield considerable influence in the policy-making process. They make many decisions in their own right and work behind the scenes to influence others. In fact, bureaucratic influence is found in every stage of the policy-making process. Bureaucrats help set policy agendas, frame issues, formulate alternatives, calculate costs and benefits, secure funding, make decisions, and implement those decisions. Furthermore, bureaucrats are involved in policy evaluation and provide crucial feedback to policy makers.

Political scientist Norton E. Long once explained that "power is the lifeblood of administration."^[13] To be effective, bureaucrats must enter the political arena and take an active role in the policy-making process. The bureaucracy is a political institution just like the president, Congress, and the Supreme Court. Kenneth J. Meier explained: "It must build political support for its actions, and it uses that political support to interact with other political institutions to make public policy."^[14]

There are many other sources of bureaucratic power. In the American political system, governing coalitions come and go. Yet the bureaucracy is relatively permanent. Bureaucrats administer programs and execute laws that are the cumulative legacy of past and present governing coalitions. Thus, the bureaucracy has a vivid institutional memory that spans several hundred years of self-government. Its stability and permanence, and the continuity it provides, are primary sources of bureaucratic power.

Francis E. Rourke emphasized the ability of bureaucrats to mobilize political support and lend professional expertise.^[15] James Q. Wilson contended that bureaucratic power is dependent on the type of task an agency performs and the ease of monitoring its operations and activities.^[16] Kenneth J. Meier provided yet another view. He explained that resources and autonomy are the two principal sources of bureaucratic power, and that this power can be enhanced or diminished by the type of policy environment, level of political and public support, specialized knowledge, cohesion of bureau personnel, and skill of bureau leadership.^[14] In sum, there are many sources of bureaucratic power, and its exercise is situation-dependent.

BUREAUCRATIC MOTIVES AND THE “AGENCY VIEW”

An important part of bureaucratic politics is the belief that officials are motivated primarily by self-interest and tend to promote their agency’s view on most policy issues. The notion that individuals and their employing agencies have similar interests comes from Mile’s Law: where you sit determines where you stand. Bureaucrats want to keep their jobs, advance their careers, and increase their budgets. Their agencies are likely to have similar goals: they want to maintain their current staffing levels, extend their jurisdiction and reach, and increase their budgets.

Another reason bureaucrats tend to have an agency orientation is that they spend most of their careers in a single agency. After years of indoctrination and investment in the agency, they naturally come to embrace its cause.

So far, the study of bureaucratic politics has been dominated by the assumption that bureaucrats are self-interested and self-serving, and that their agencies are similarly motivated. One of the most promising avenues for future research is to relax this assumption and recognize other important motives. These motives may include the desire to perform meaningful public service and to protect the public interest.

AN EXAMPLE: THE CUBAN MISSILE CRISIS

In his classic study of the Cuban Missile Crisis, Graham T. Allison examined the decision-making processes of the United States and Soviet governments during the tense international standoff that occurred in October 1962.^[17–18] Allison used three “conceptual lenses”: the rational, organizational process, and bureaucratic politics models. Thus, his study is particularly enlightening because it compares and contrasts the bureaucratic politics model with the two other models.

Summary of Events

The Cuban Missile Crisis was the most dangerous U.S.–Soviet confrontation of the Cold War. After the U.S.-sponsored Bay of Pigs incursion failed to depose Cuban Premier Fidel Castro, relations between the two nations steadily worsened. In October 1962, a U.S. reconnaissance flight discovered that the Soviet Union was secretly building ballistic missile launch sites in Cuba—approximately 100 miles from the Florida coastline. President John F. Kennedy was informed on the morning of October 16, and he immediately called the first of a series of meetings with his top advisors. After several days of gut-wrenching deliberations, Kennedy decided to impose a naval blockade on Cuba. He simultaneously warned Soviet Premier Nikita Khrushchev that any missile launches from Cuba would prompt a full-scale U.S. retaliatory attack against the Soviet Union. Khrushchev called the naval blockade “an act of aggression which pushes mankind toward the abyss of a world nuclear war.” On October 24, Soviet ships carrying missiles to Cuba approached the quarantine line. After several tense hours, the Soviet vessels turned back averting a direct military confrontation. Soon afterward, Khrushchev ordered the missile sites dismantled and the missiles removed from Cuba. Thus, the 13-day crisis ended.

The Rational Actor Model

The rational actor model assumes that states are unitary actors, and that they make rational decisions. Two types of rationality are possible: procedural and substantive. Procedural rationality focuses on the decision-making process. This process consists of the following steps: defining the problem, searching for alternatives, estimating the costs and benefits of each alternative, and finally, selecting the best option. In contrast, substantive rationality focuses on the result. The preference of the decision maker can be inferred from the outcome, as all actions are viewed as intentional.

In the present case, Kennedy and his advisors tried to utilize the rational model. They searched for alternatives and identified several: an Army invasion of Cuba; an Air Force attack on the missile sites; a naval blockade to prevent missiles from entering Cuba; a State Department diplomatic protest; and simply ignoring the situation because it did not change the overall balance of power. Each alternative was evaluated against a set of criteria that included potential loss of American lives; monetary cost; probability of starting a nuclear war; domestic political fallout; and others. After careful study, Kennedy decided that the most prudent course of action was his twin response: a naval

blockade of Cuba and the threat of massive retaliation against the Soviet Union. Thus, viewing the decision-making process through the rational model lends important insights.

The Organizational Process Model

Next, Allison considered the organizational process model. All large organizations develop bureaucratic structures and decision rules to avoid uncertainty and improve their decision-making ability. These structures and rules represent organizational learning from past successes and failures, and they provide continuity in the event of employee turnover. That is, bureaucratic organizations use decision rules and link decision-making power to offices rather than particular individuals.

The main difference between the rational and organizational models is in how the decision-making process is depicted. The rational model assumes that individuals go through the lengthy process of winnowing down alternatives and arriving at the one-best-solution. The organizational process model counters that rules—often referred to as standard operating procedures (SOPs)—guide decision making. Many participants with different preferences are involved, but they usually unite behind the SOP.

Allison provided the following example. President Kennedy ordered a naval blockade to prevent the Soviets from shipping more missiles into Cuba. Defense Secretary Robert McNamara wanted to draw the quarantine line close to the Cuban coast. He felt that a “tight blockade” would reduce the chances of a U.S.–Soviet confrontation. However, the Navy’s SOP called for a more relaxed 500-mile quarantine line. Such a line would serve several purposes for the Navy. It would increase their chances of intercepting ships headed for the coast and give them ample space to maneuver away from hostile naval forces and thwart air attacks from the mainland. Even though McNamara ordered the Navy to implement a tight blockade, they did not heed his order. The type of blockade was determined by the Navy’s SOP. Thus, the organizational process model provides important insights about the decision-making process.

The Bureaucratic Politics Model

The bureaucratic politics model (Fig. 1) emphasizes that many officials are involved in policy decisions, and their preferences are derived from their organizational backgrounds. The Secretary of the Army will stress the Army’s position in interexecutive bargaining, while the Secretaries of the Navy and Air Force will represent their respective service branches. Hence,

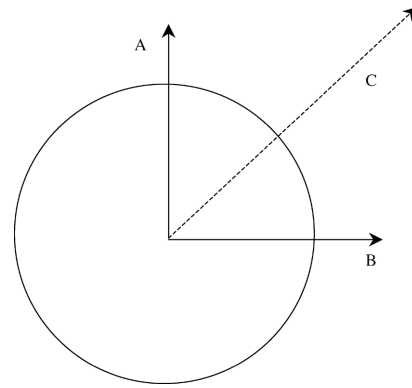


Fig. 1 The bureaucratic politics model of decision making. In the bureaucratic politics model, the decision-making process is characterized by conflict and bargaining. The participants engage in a “tug-of-war” trying to steer policy in the direction they prefer. Imagine two participants with equal power tugging in different directions (lines A and B). The outcome is what Allison called the “resultant” (line C).

the decision-making process involves multiple actors with different preferences, and it is marked by conflict and bargaining. The participants engage in a “tug-of-war” to pull policy toward their preferred outcome. Allison called the final decision a “resultant.”^[19–20]

During the Cuban Missile Crisis, many different groups were involved in the White House decision-making process: the Defense Department, State Department, Joint Chiefs of Staff, Central Intelligence Agency, and others. These actors had different preferences, largely because of their formal roles and agency ties. Furthermore, some had more influence than others. For example, Attorney General Robert F. Kennedy was highly influential because he was President Kennedy’s younger brother and close confidant.

An intense tug-of-war ensued. Defense Secretary McNamara originally proposed a “do nothing” option, but the President deemed this response politically unacceptable. A hard-line coalition led by the Service Chiefs, former Secretary of State Dean Acheson, and others emerged and pushed for a surgical air strike. McNamara opposed this option and countered by proposing a naval blockade. He convinced the Attorney General and Presidential Counsel Ted Sorenson to support his plan. In the end, this moderate coalition prevailed.

Allison said the decision-making process was a bargaining game, and the participants knew it. They formed strategic coalitions and entered the fray. The outcome was a political result hammered out by these coalitions. Allison explained that even when an outcome is mostly the triumph of one individual or group, bureaucratic politics may help explain the context and inner workings that made the victory possible.^[18]

	Rational Model	Organizational Process Model	Bureaucratic Politics Model
No. of Actors	One	Many	Many
No. of Preferences	One	Many (United)	Many (Divergent)
Decision-Making Process	Rational	According to Rules and Procedures	Via Conflict and Bargaining
Outcomes	Intentional	Pre-Determined	Resultant
Major Drawbacks	Hard to Attain	Not Very Flexible	Hard to Predict

Fig. 2 Comparison of Allison's three models of decision making.

In conclusion, the rational model assumes that states are unitary actors with fixed preferences. Officials acting for the state try to maximize the state's utility by making decisions rationally. The organizational process model assumes that each organization has unique preferences and that decisions are made according to predetermined bureaucratic routines called SOPs. The bureaucratic politics model assumes that decision makers have a unique set of preferences derived from their positions in government and that the decision-making process is a tug-of-war between these individuals. Overall, the bureaucratic politics model seems more plausible than the other two models because it conveys the dynamism and "real politik" of the policy-making process.

Fig. 2 summarizes the key differences between these models.

In the end, Allison is not advocating one best approach. He suggests using a variety of conceptual lenses.

CRITIQUES OF BUREAUCRATIC POLITICS

Theories of bureaucratic politics are limited for several reasons. First, Mile's Law does not always hold true. Officials are not always self-interested and self-serving, and they do not always promote their agency's view on important policy issues. Second, the model does not lend itself to making predictions. Third, bureaucratic politics raises the important question of bureaucratic accountability in democratic government.

How can we be sure that bureaucrats, who are non-elected government officials, will subordinate their self-interest and act in the public interest? This question is the Achilles heel of modern-day public administration,

and it seems to defy any simple answer.^[21] Most contemporary thinkers contend that a combination of internal and external control mechanisms is the best strategy. Accordingly, bureaucrats are urged to act responsibly, and they are monitored by overhead political officials.

CONCLUSION

The term bureaucratic politics has several different meanings in the literature. These meanings vary from the narrow decision-making model described by Graham T. Allison to the broader, more encompassing notion of bureaucratic influence on policy making described by political scientists Francis E. Rourke, James Q. Wilson, and Kenneth J. Meier. Another stream of scholarship portrays bureaucrats and their agencies as threats to the public interest. Thus, this portrayal casts a negative light on bureaucratic politics. Overall, these various perspectives have several things in common. They all acknowledge that bureaucrats have a significant amount of discretion, and that these officials are active participants in the policy-making process.

Accordingly, bureaucratic politics emphasizes that bureaucrats, who are non-elected government officials, enter the political arena and try to influence or shape the policy-making process. These officials are assumed to be self-interested and likely to promote their agency's view on specific policy issues. That is, they are expected to protect and advance their own interests when engaged in inter-executive or inter-agency competition. As a result, theories of bureaucratic politics suggest that policy outcomes are the result of conflict and bargaining rather than the more rational or incremental processes described elsewhere.

REFERENCES

1. Lilenthal, D.E. *Democracy on the March*; Harper and Row: New York, NY, 1944.
2. Selznick, P. *TVA and the Grass Roots*; University of California Press: Berkeley, CA, 1949.
3. Heclo, H. Issue networks and the executive establishment. In *The New American Political System*; King, A., Ed.; American Enterprise Institute: Washington, DC, 1978; 87–124.
4. Lynn, L.E., Jr. The reagan administration and the renitent bureaucracy. In *The Reagan Presidency and the Governing of America*; Salamon, L.M., Lund, M.S., Eds.; The Urban Institute Press: Washington, DC, 1984; 339–370.
5. White house personnel office the malek manual. In *Classics of Public Personnel Policy*; Thompson, F.J.,

- Ed.; Moore Publishing Co., Inc.: Oak Park, IL, 1979; 159–187.
6. Nathan, R.P. *The Administrative Presidency*; John Wiley: New York, NY, 1983.
 7. Niskanen, W.A., Jr. *Bureaucracy: Servant or Master?* Institute of Economic Affairs: London, 1971.
 8. Perrow, C. *Complex Organizations: A Critical Essay*, 3rd Ed.; Random House: New York, NY, 1986.
 9. Wood, B.D.; Waterman, R.W. *Bureaucratic Discretion: The Role of Bureaucracy in a Democracy*; Westview Press: Boulder, CO, 1994.
 10. Wilson, J.Q.; DiIulio, J.J., Jr. *American Government: Institutions and Policies*, 8th Ed.; Houghton Mifflin Company: Boston, MA, 2001.
 11. Rourke, F.E. American bureaucracy in a changing political setting. *J. Public Adm. Res. Theory* **1991**, *1* (2), 111–129.
 12. Durant, R.F. Whither bureaucratic influence? A cautionary note. *J. Public Adm. Res. Theory* **1991**, *1* (4), 461–476.
 13. Long, N.E. Power and administration. *Public Adm. Rev.* **1949**, *9* (3), 257–264.
 14. Meier, K.J. *Politics and the Bureaucracy: Policymaking in the Fourth Branch of Government*, 3rd Ed.; Brooks/Cole Publishing Co.: Pacific Grove, CA, 1993; 1.
 15. Rourke, F.E. *Bureaucracy, Politics, and Public Policy*; Little, Brown and Co.: Boston, MA, 1969.
 16. Wilson, J.Q. *Bureaucracy: What Government Agencies Do and Why They Do It*; Basic Books: New York, NY, 1989.
 17. Allison, G.T. *Essence of Decision*; Little, Brown: Boston, MA, 1971.
 18. Allison, G.T.; Zelikow, P. *Essence of Decision: Explaining the Cuban Missile Crisis*, 2nd Ed.; Addison-Wesley: New York, NY, 1979.
 19. Allison, G.T.; Halperin, M.H. Bureaucratic politics: a paradigm and some policy implications. In *Theory and Policy in International Relations*; Tanter, R., Ullman, R.H., Eds.; Princeton University Press: Princeton, NJ, 1972; 40–78.
 20. Rousseau, D.L. *Bureaucratic Politics Module*; 2000. Available online: <http://www.polisci.upenn.edu/psci150/modules/bur/index.htm>.
 21. Waldo, D. *The Administrative State*; Ronald Press: New York, NY, 1948.

Bureaucrats and Politicians in Southeast Asia

Scott Fritzen

Public Policy Programme, National University of Singapore, Singapore

INTRODUCTION

Ever shifting relationships between bureaucrats and politicians have played a direct or indirect role in several major debates in Southeast Asia since colonization. One theme concerns the *contradictions of the colonial inheritance*, with its ideology of a politics–administration separation coupled with extreme executive dominance in practice. A second theme involves the role of the bureaucracy in *promoting economic transformation* in several Southeast Asian states modeling themselves after the successful newly industrialized countries (NICs). Finally, more recent calls for good governance in the wake of the East Asian financial crisis coupled with democratization trends in the region have raised the salience of the concept of *democratic accountability* of both politicians and bureaucrats. This review examines each of these areas, concluding that the dynamics of the bureaucrat–politician relationship are likely to continue to be a central point of theoretical and practical contestation in the coming years.

THE COLONIAL LEGACY

A seminal concept dominating the “modern” democratic notion of the bureaucrat–politician relationship harkens back to Woodrow Wilson’s politics–administration dichotomy: “The field of administration is a field of business ... removed from the hurry and strife of politics” (1887, quoted in Ref.^[1]). Both politics and administration, it was understood, were to be underpinned by accountability to the public and to the rule of law. This normative ideal, whatever the realities in the home country context, did not travel well to Europe’s Southeast Asian colonies. There, the appointed (European) “man in the field” was the ultimate bureaucrat-cum-politician ruling with great discretion with no democratic accountability.^[2] This was, in practice, usually consistent with indigenous patterns of rule in kingdoms of Siam and Burma or the sultanates of the Malay Peninsula.^[3–5]

The task of crafting modern polities, whether as part of the decolonization agenda or revolutionary struggles, was one of developing and cementing elite consensus within functioning organizations capable

of commanding a sufficient degree of legitimacy and of exercising state power. This task expressed itself in different forms, depending on the contingencies of the political process. Three models are evident.

The first pattern, found in Malaysia, Indonesia, the Philippines, and Burma, involved a semisponsored, more or less orderly handover of power from the colonial authorities to a local elite. Colonial authorities looked to the civil service as the logical—often the only—source of “acceptable” nationalist leadership. In some places, these authorities explicitly endeavored to leave behind more ethnically representative bureaucracies (reversing earlier policies that favored Indians and Chinese in Malaysia, for instance) and political movements.^[6] When such promotion led, in practice, to uncomfortably close ties between aspiring politicians and bureaucrats (because of the shallow pool of politically acceptable talent that colonial administrators could tap), principle gave way to political expediency. These former colonies were left with weakly institutionalized political systems that had but a semblance of political–administrative separation—and none at all following the military coups that subsequently occurred in all of the above countries except Malaysia.

Thailand, which avoided colonization, must stand in its own category. Here, at least until recent years, a military–bureaucratic alliance dominated. The 1932 coup d’état that overthrew the absolute monarchy was motivated as much by bureaucratic–clientelistic objectives as by calls for greater modernization or democracy per se.^[4] The ensuing pattern of Thai institutional life as “a matter of competition between bureaucratic cliques for the benefits of government” (David Wilson, quoted in Ref.^[4]) remained, in some ways, institutionally consistent for several decades.

The third pattern—most costly in terms of human suffering and developmental trajectories—involved prolonged violence and instability in Indochina. It began in the immediate postwar period with the North Vietnamese struggle to unify the country under communist rule and ended only in 1993 with UN-sponsored elections initiated a decade of political consolidation and a semblance of stability.^[7] Prolonged strife, combined with attempts to implement radical socialist programs, left Vietnam, Laos, and Cambodia without a professional bureaucracy or even

rhetorical norms limiting regime powers. The bureaucracy was subsumed under the overriding necessity of developing revolutionary administrations utilizing neo-Stalinist forms of political organization.

To summarize, the internal contradictions of colonialism, coupled with the political turbulence that followed decolonization and modernization period, left bureaucracies in all of these countries (save Thailand) decisively dominated by political masters. The systems that emerged, however, varied greatly in their developmental effectiveness, for reasons explored in the next section.

THE BUREAUCRACY IN NEWLY INDUSTRIALIZING CONTEXTS

Beginning with the Japanese transformation, several East Asian countries, together with one Southeast Asian—Singapore—were able to rapidly transform their economies, attaining within some 20 years levels of per capita income that placed them in the ranks of advanced, industrial countries. Several Southeast Asian countries, which collectively became known as the new “tigers”—Malaysia, Thailand, and Indonesia—appeared to model certain aspects of the NIC’s developmental recipe and appeared to achieve considerable success in doing so.

How had the NICs achieved such a feat? This became the dominant question in practical and theoretical debates that continue to this day and in which the relationships between bureaucrats and politicians came to play a key role. Contrasting explanations were offered by those^[8,9] who advocated that the state’s role in economic decision making was decisive and others^[10] claiming that state success in achieving macro-economic stability and an external orientation was far more important than bureaucratic interventions. An influential explanation among the former group^[11,12] held that the key to developmental effectiveness, both in the NIC case and potentially in other countries, was the “embedded autonomy” of the bureaucracy. Bureaucracies could, by this theory, successfully serve as “midwives” to economic development if they enjoyed four conditions. The first was a political leadership determined to pursue a transformational agenda with little fear of electoral backlash. The second was a meritocratic, technically competent bureaucracy (or an elite, “piloting” segment of it). The third and fourth requirements were special characteristics of this bureaucracy that would allow it to successfully intervene in the economy. The bureaucracy was to be both “embedded”—enjoying dense informational links to the companies and market sectors to be promoted—and “autonomous,” i.e., not captured by any special interest and therefore being capable of “disciplining”

capital by, for instance, stopping subsidies where this was necessary.

To what extent did Southeast Asian countries put these purported conditions of success in place? Singapore, as one of the original NICs, was in many ways the archetypical example of the process. One must only make the proviso that the bureaucratic “autonomy” in question was not autonomy from the ruling People’s Action Party itself. Rather, it reflected the overall autonomy of the regime, with its technically competent bureaucrats clearly in position as “implementers” of policies.^[13,14]

Malaysia, Thailand, and Indonesia were all held up at one time or another as exemplary “new tigers” primarily because of their success in achieving, for extended periods, high rates of economic growth. Malaysia and Indonesia shared considerable regime stability (at times enforced by repression), a commitment of the leadership to economic modernization, and development outcomes that were broadly based, whether through programs of outright redistribution (as in the case of Malaysia’s New Economic Policy) or successful strategy to boost agricultural productivity (as in Indonesia).

The bureaucracy in all of these cases was thought to be reasonably capable and “technocratic.” Yet economic interventions were of a generally smaller scale than in the NICs themselves, and the “disciplining” of capital was heavily constrained by the need to promote the “illegitimate” (i.e., non-transformational) interests of the ruling clique itself, as expressed in soaring levels of corruption. Only in Malaysia did the high degree of unity of the Malay political-bureaucratic establishment consistently promote an agenda of redistribution from the Chinese economic elite to ethnic Malay majority. As shown in the next section, however, the developmental model employed by the “tigers” clearly involved less of a “steel-frame” bureaucracy than that implied by the embedded autonomy theory.

Vietnam has more recently emerged as arguably a vital member of the “tiger” community, with economic growth rates consistently among the highest in the world throughout the 1990s. Scholarly debates continue as to whether the far-reaching *doi moi* (renovation) reforms it introduced were driven by local bureaucratic responses to failed central planning or by a far-sighted central leadership.^[15] Whatever the case, the rapid growth that ensued initially reflected gains brought on by the one-off removal of “artificial” constraints on growth imposed by poor institutions (such as collective agriculture) but was later sustained by a political-bureaucratic Communist Party elite calculus that rapid growth and modernization was the best chance the party had to maintain its legitimacy.^[16] This strategy of economic reform with continued

political authoritarianism has generated some tensions in Vietnam, as it has in countries such as Singapore and Malaysia, a trend reflecting of democratic pressures for reform.

PRESSURES FOR REFORM: THE GOVERNANCE AGENDA COMES OF AGE

The last 15 years or so have seen a shift in the parameters of the debate over the politician–bureaucrat relationship throughout Southeast Asia. Two overlapping categories of influence are particularly important in this context: democratization and the rise of the “good governance” agenda.

Calls for greater political accountability have been on the increase throughout the region. A “third wave of democratization”^[17] saw civilian, multiparty rule return in Thailand (1992), Philippines (1986), Indonesia (1998), and Cambodia (1993). Public anger over the closed workings of the political–bureaucratic elites dominating the first three countries was a key factor in bringing down autocratic rulers. This increased mobilization and citizen consciousness has carried over into democratic politics in these countries; use of the bureaucracy for the enrichment of the political elite continued to feature prominently in calls for both political and bureaucratic reform in several contexts, for instance, the ouster of President Joseph Estrada during “People’s Power II” (2002) and more positively, in the introduction of an ambitious new constitution in Thailand in 1997. Citizens throughout the region showed themselves eager and willing to envision democratization as a check on both unaccountable political and bureaucratic power.

In Singapore, Vietnam, Laos, and Brunei, governance remained both authoritarian and (for practical purposes, although via very different mechanisms) uncontested, whereas in Myanmar (Burma), a small group of generals continued to block any political normalization. Even here, the *ersatz* vocabulary of democratization—articulated in terms of improved governmental responsiveness to citizen feedback—was an increasingly prominent feature in officially sanctioned political discourse, although practical reform efforts betrayed the significant contradictions of their contexts. For instance, Singapore’s establishment of district-based “Community Development Councils” in 1996 was touted as an attempt to promote greater citizen involvement in local decision making, despite the fact that council members were all appointed by the ruling party.^[18] In Vietnam, much-hyped “grassroots democratization” reforms introduced in 1997, in practice, meant a clarification of standards (e.g., regarding local government budget transparency) that higher levels of the party-state

apparatus could use selectively to “discipline” lower levels.^[16]

The “good governance” agenda that arose contemporaneously with the democratization movement above was given impetus primarily from donors, academic work highlighting the impact of governance qualities on development outcomes, and various think tanks and non-governmental organizations. One variant of the agenda, driven by increasing financial integration of the region, was concerned corporate governance. Explanations for the Asian financial crisis beginning in 1997, which afflicted Indonesia, Thailand, and the Philippines more than other Southeast Asian countries, came to center on a pernicious set of incentives—of politicians to use capital markets for personal purposes; of banks to lend funds for unsound investments; and of bureaucratic regulators to overlook all of the above.^[19] The financial crisis led to a fundamental reassessment of the supposed strengths of the region’s political–bureaucratic systems. More specifically, the call was for greater transparency and disciplined oversight systems that reinforced the call for a clear separation of bureaucratic from political authority and for both to be underpinned by a reinforced rule of law.

Another “good governance” agenda lies in the promotion of decentralized decision making and management, whether in a democratic, fiscal, or administrative context. Here there was also much to discuss on the regional scene, with virtually all countries in the region engaged in some form of formal decentralization policy. For most countries, this went beyond rhetorical support for the concept as an ideological smokescreen. In Cambodia—arguably the least hospital environment for devolution in the region—authorities astounded observers with the speed with which it established democratically elected Commune Councils, even before presumably easier administrative reforms supporting such arrangements were in place. Through the rapid introduction of two decentralization laws (Nos. 22/1999 and 25/2000) in the immediate aftermath of Soeharto’s resignation, Indonesia moved “from being one of the most centralized countries in the world to one of the most decentralized.”^[20] Thailand and the Philippines both continued to decentralize decision making to provinces and districts, the latter quite vigorously.^[21]

There is, finally, a managerialist variant of the good governance agenda. The New Public Management (NPM) can be seen as an attempt to reengineer relationships between politicians, bureaucrats, and the public.^[22] Key strategic elements in this broad family of reforms include: 1) delinking policy formulation (or “steering”) from implementation functions to concretize the specific accountability borne by all actors; 2) introduction of competitive pressures to the

bureaucracy using market mechanisms such as privatization where possible and systems of performance measurement where not; and 3) deregulation of internal operations to promote flexible, creative implementation.^[23] Application of such ideas to developing country contexts, where the predictability and transparency of bureaucratic processes is limited, has been hotly contested.^[24] But attempts to apply NPM principles to governance reforms are standard policy in the higher capacity countries covered here (Singapore and Malaysia), increasingly, if hesitantly, being applied in the newly democratic countries of Southeast Asia and are prominent in public administration reform strategy.

CONCLUSION

In the aftermath of colonialism, bureaucratic-politician relationships were, like government systems themselves, weakly institutionalized in Southeast Asia, regardless of whether they had inherited the ideological veneer of a politics-administration dichotomy. The “tigers” of Southeast Asia were able to grow quickly by drawing on a combination of transformational leadership, an external orientation, and core bureaucratic agencies of at least moderate capacity. The politician-bureaucrat relationship was driven in these cases by a substantial unity of elite interests (as in Malaysia) or domineering executives or parties (as in the Philippines, Indonesia, and Vietnam). Political systems were, in many cases, highly fragile and, in others (such as Myanmar), rigid but non-transformative; in all cases except Singapore and possibly Malaysia, however, bureaucratic capacities and incentives fell well beneath the threshold levels of “embedded autonomy” that theorists argued were necessary to sustain economic transformation over a period of not years but decades, as in the newly industrialized countries.

Democratization and good governance pressures, driven by the increased sophistication of societies and by global financial integration, have propelled the politics-administration link into the center of political discourse throughout the region. In a few countries (notably Myanmar, Laos, and Brunei), there is no discernable movement toward a more institutionalized, professional bureaucracy. Other authoritarian contexts (such as Singapore and Vietnam) are attempting to make their bureaucratic systems more responsive to their publics without threatening (indeed, while bolstering) the position of ruling elites. In more democratically competitive settings, a wider range of reforms—from formal regulation of civil servants via asset disclosure requirements to democratic decentralization to increase the direct accountability of service providers and local politicians to the public—is being

introduced. Capacity to implement such ambitious reforms is, in all cases, limited. The outcomes of attempts to establish transparent relationships between politicians and bureaucrats, where these have never sat easily within the region’s political economy, will be hotly contested in the years to come, both theoretically and on the ground.

REFERENCES

1. Starling, G. *Managing the Public Sector*, 6th Ed.; Harcourt College Publishers: Fort Worth, TX, 2002.
2. Tarling, N. The establishment of colonial regimes. In *The Cambridge History of Southeast Asia, Vol 2. Part 1. From c.1800 to the 1930s*; Tarling, N., Ed.; Cambridge University Press: Cambridge, 1992; Chapter 1.
3. Crouch, H. *Government and Society in Malaysia*; Cornell University Press: Ithaca, NY, 1996.
4. Riggs, F. The bureaucratic polity as a working system. In *The Political Economy of East Asia*; Ravenhill, J., Ed.; Edward Elgar: London, 1994.
5. Silverstein, J. The evolution and salience of burma’s national political culture. In *Burma: Prospects for a Democratic Future*; Rotberg, R.I., Ed.; Brookings Institution Press: Washington, DC, 1999.
6. Puthucherry, M. *The Politics of Administration: The Malaysian Experience*; Oxford University Press: Kuala Lumpur, 1978.
7. Roberts, D. From ‘communism’ to ‘democracy’ in Cambodia: a decade of transition and beyond. *Communist Post-Communist Stud.* **2003**, *36*, 245–258.
8. Wade, R. *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*; Princeton University Press: Princeton, NJ, 1990.
9. Amsden, A. *Asia’s Next Giant: South Korea and Late Industrialization*; Oxford University Press: New York, 1989.
10. World Bank. *The East Asian Miracle: Economic Growth and Public Policy*; Oxford University Press: New York, NY, 1993.
11. Evans, P. *Embedded Autonomy: States & Industrial Transformation*; Princeton University Press: Princeton, NJ, 1995.
12. Evans, P. Transferable lessons? Re-examining the institutional prerequisites of East Asian economic policies. *J. Dev. Stud.* **1998**, *34* (6), 66–86.
13. Leong, H.K. *The Politics of Policy-Making in Singapore*; Oxford University Press: Singapore, 2000.
14. Hamilton-Hart, N. The Singapore state revisited. *Pac. Rev.* **2000**, *13* (2), 195–216.
15. Fforde, A.; de Vylder, S. *From Plan to Market: The Economic Transition in Vietnam*; Westview Press: Boulder, CO, 1996.
16. Fritzen, S. Growth, poverty and the future of poverty reduction in Vietnam. *J. Asian Econ.* **2002**, *13* (5), 635–657.
17. Diamond, L. *Developing Democracy: Toward Consolidation*; Johns Hopkins University Press: Baltimore, 1999.

18. George, C. *Singapore: The Air-Conditioned Nation: Essays on the Politics of Comfort and Control 1990–2000*; Landmark Books: Singapore, 2000.
19. Dwivedi, O.P. Challenges of culture and governance in Asian public administration. In *Handbook of Public Administration*; Peters, B.G., Pierre, J., Eds.; Sage: London, 2003; Chapter 41.
20. Hoffman, B.; Kaiser, K. The making of the big bang and its aftermath: a political economy perspective. Conference Paper: “Can Decentralization Help Rebuild Indonesia?,” Atlanta, Georgia, May 1–3, 2002; Georgia State University.
21. Brillantes, A.B., Jr. *Philippines: Sustaining the Gains of Decentralization*; Brillantes, A.B., Jr., Cuachon, N.G., Eds.; Asian Resource Center for Decentralization, 2002. Available at http://www.decentralization.ws/srcbook_main.asp (accessed December 29, 2003).
22. Peters, B.G.; Pierre, J. Civil servants and politicians: the changing balance. In *Politicians, Bureaucrats and Administrative Reform*; Peters, B.G., Pierre, J., Eds.; Routledge: London, 2001.
23. Osborne, D.; Plastrik, P. *Banishing Bureaucracy: The Five Strategies for Reinventing Government*; Addison Wesley: Reading, MA, 1996.
24. *The Internationalization of Public Management: Reinventing the Third World State*; McCourt, W., Minogue, M., Eds.; Edward Elgar: Cheltenham, UK, 2001.

Burnout in Public Agencies

Robert T. Golembiewski

Department of Political Science, University of Georgia,
Athens, Georgia, U.S.A.

INTRODUCTION

A favorite metaphor involves efforts to add tension to a system to energize it, which requires fine-tuning short of that point at which the system becomes overloaded and “snaps.” In this elemental sense, organizations are like rubber bands propelling a spitball. Increase the loading on the band by extending it, *but not too far*, and the missile will go further and faster. Apply too much loading, and the rubber band will just pop into two or more pieces, and that may sting the shooter rather than provide momentum for the projectile.

Some such metaphor underlies most downsizing or rightsizing. For example, a system is seen as having too many human resources, or the wrong kinds, and the goal is to eliminate some employees while increasing the workload of those who remain. All too often, however, the human system may be deteriorated in the process (e.g., Ref. ^[1]).

This common elemental fine-tuning strategy is not guaranteed to “work,” far from it, in fact. Guidelines exist for seeking that magical middle ground, where systems become more energized (e.g., Ref. ^[2]) and also avoid the breaking point. Failure can be costly, nonetheless, and perhaps is so in the majority of cases. Witness violence at the work site, which seems more and more a hallmark of our age. Frequently, these outbursts can be traced to a person or persons under stress, tragically beyond the point at which they “snap,” or “go around the bend,” emotionally speaking. In the most dramatic form, someone turns gunfire on fellow workers, on specific targets as well as those unlucky enough to be in the wrong place at the wrong time.

The burnout literature takes us a substantial distance toward isolating such “breaking-points” in practice, although we are not yet there. In the opinion of this author, and he has a clear vested interest (e.g., Ref. ^[3]), the “phase model of burnout” is the farthest along of several possible developmental models. A few details suggest the nature and range of support for this view.

PHASE MODEL OF BURNOUT

After a decade of development, the phase model permits a number of confident generalizations. In turn, six points will receive emphasis below:

- Conceptually, we have a solid idea of how to define a dimension reasonably labeled “burnout.”
- Operationally, for extensive purposes, we know how to measure the incidence and severity of burnout.
- Tactically, the phases of burnout are associated with a long list of undersized and undesirable outcomes, both personal and individual.
- Burnout seems to persist over extended periods of time.
- The incidence of burnout, as seen in various contrasts, seems to fall in the range of bad to worse.
- In general, incidences of burnout do not seem to vary robustly and regularly, in public versus business contexts.

Normal Definition of Burnout Subdomains

Most observers generally agree that burnout is comprised of three subdomains. Following Maslach,^[4,5] individual burnout is contributed to by the following:

- High scores on items relating to *depersonalization* or the tendency to think of individuals as things or objects rather than as sensitive, valuing beings.
- Low scores on *personal accomplishment* (*reversed*), which is tapped by items relating to how well a person does on work tasks as well as on how worthwhile these tasks are seen to be.
- High scores on *emotional exhaustion*, with items relating to the emotional slack available to individuals—as coping skills and attitudes are discounted by the stressors an individual experiences at a particular point in time.

Operational Definition of Phases

Items from the MBI (or Maslach Burnout Inventory) are used to estimate the three subdomains, scores on which for each individual are differentiated as high versus low using norms based on the population of a large public agency^[3] operating under substantial stressors and strain. A median-cut on each subdomain is utilized to make the high versus low distinction.

Beyond this point, a single decision rule applies: depersonalization is seen as more virulent than personal accomplishment, and both are seen as less determinative of burnout than emotional exhaustion. Overall, each individual is assigned to one of the eight phases in Fig. 1. Note that the high versus low assignments are based on norms from 1500 respondents in a U.S. federal agency high levels of strain.

Note that one significant point that the model does not imply is that progression to full burnout must take the path I→II...→VIII for individuals. Rather, however one arrives at a specific phase, the model predicts that worsening conditions for individuals will be associated regularly with I versus II, II versus III, III versus IV, and so on.

Indeed, there seem to be two basic flight paths to advanced burnout. In *chronic onset*, the basic pattern seems to be I→II→IV→VIII. In *acute onset*, various flight paths may exist. Thus, a person suddenly losing a beloved mate might quickly advance I→V. Beyond that point, continued growing might be variously reflected in advancing burnout. For example, a person might adapt by losing interest in work, in which case, the flight path becomes I→V→VII and might then extend to VIII. Alternatively, that person might try to “lose self in work,” in which case, the pathway would be I→V→VI, ultimately with a possible extension to VIII. Recovery could variously work its way back through (perhaps) phase I, in the cases of chronic and acute onset.

In sum, the burnout phases progress from least advanced→most advanced, or phase I→VIII, but not in a simple and sovereign progression. In sum, an VIII strongly tends to objectify personal relationships, believes self to be doing poorly on a task with little socially redemptive value, and experiences only a slight positive balance, if not an actual deficit, in the stressors

experienced as contrasted with coping skills and attitudes for dealing with them.

The overall results, in North America and world-wide, suggested that their distribution needed close attention.^[6] Only Phases I→VIII are regularly and robustly associated with a broad range of worsening conditions, phase by phase. In addition, 40% or more of all employees are assigned to Phases VI→VIII, which indicates a large proportion of “burned-out” individuals.

Phases as Breaking Points

The phases do not provide a map of the specific point when strain becomes “too great,” but the model nonetheless has tolerable utility. Table 1 suggests the point in contrasting the consequences of high versus low burnout, generally defined as phases I–III versus VI–VIII, respectively. Perhaps most notably, differences on virtually all variables are regular; thus, Phase I covariants are more desirable than those of Phase II, Phase II covariants are more desirable than those for Phase III, and so on for all paired comparisons. These mean differences for phases not only fall in the same directions, but they are also statistically-different in far more cases than one would expect.

Note that this general pattern of associations is all but universal, although variations in intensity exist that may reflect local conditions. The generalizations in Table 1 come from many studies in which those in Phase I were better off than those in II, those in II better than in III, and so on. To illustrate, the pattern is virtually the same in North American populations as well as those in 20-plus other nation states for which data are available.^[6] To a similar effect, the pattern is much the same for archival data as for self-reports. For example, those in Phase I spend the least on medical insurance; those in VIII spend the most; and the other phases have spending that progresses quite regularly through the phases.^[3]

Of great significance, both theoretic and practical, note that the high profile in Table 1 has a familiar character. Quite directly, that profile is like that associated with Participative Management or Industrial Humanism. This profile suggests that burnout is a

	Phases of Burnout							
	I	II	III	IV	V	VI	VII	VIII
Depersonalization	Lo	Hi	Lo	Hi	Lo	Hi	Lo	Hi
Personal accomplishment (reversed)	Lo	Lo	Hi	Hi	Lo	Lo	Hi	Hi
Emotional exhaustion	Lo	Lo	Lo	Lo	Hi	Hi	Hi	Hi

Fig. 1 Phase model assignments.

Table 1 Illustrated conditions characteristic of extremes of burnout

Least-advanced phases of burnout are characterized by:		Most-advanced phases of burnout are characterized by:
High	Participation in work decisions	Low
High	Involvement	Low
Low	Spending on medical insurance	High
Low	Physical and emotional symptoms	High
High	Performance evaluation	Low
High	Cohesiveness	Low
High	Autonomy at work	Low
High	Trust in management	Low
High	Trust in fellow employees	Low
High	Supervisory support	Low
Low	Family dislocation	High

centroid—the phases seem related to many aspects of nature relevant to management. This feature implies that the phases are a source of great leverage in planned change, once we understand them sufficiently to induce the effects we desire.

Virulence of Phases

The phase model is the only one that permits an estimate of how many people have which degrees of burnout. And, the overall message is alarming. On a large and growing panel of variables, self-reports and “objective” or “hard” variables, that is to say, persons in the progressive phases experienced diminishing and deteriorating conditions. Thus, those in Phase I reported the fewest or least severe symptoms of mental or physical illness, to illustrate; and those in VIII have the poorest record. In between, even more revealing, II are significantly worse off than Is; IIIs tend to report greater deficiencies than Is; and so on.

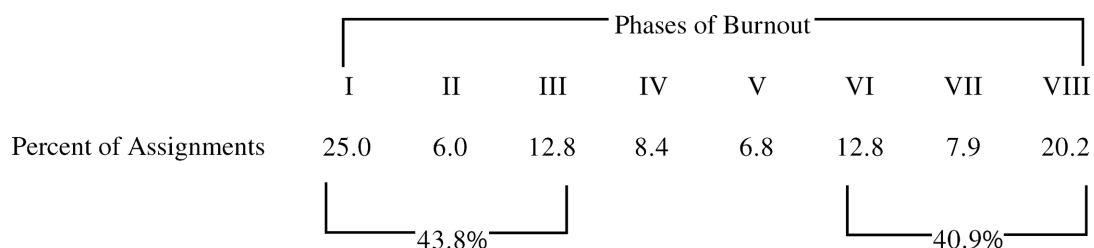
One example of the common pattern of relations will serve to illustrate the correlation of the phases with relevant variables. Consider physical symptoms, for

example (e.g., Ref. [3]). The 19 Michigan physical symptoms—from sweaty palms to chest pains^[7]—covary directly with the phases. For example, a national sample of respondents reflected this familiar pattern when compared to symptoms arrayed by the phases: those in Phase I had a small fraction of each symptom reported by the national panel; those in Phase VIII had five to 15 times to the incidence of symptoms reported by the national panel; and those in other phases fell in regular order.

The general sense of it is like a stepladder: as the phases vary, regularly and robustly, so do the conditions experienced by people tend to worsen or deteriorate. Table 1 provides useful illustration of these stable patterns of effects. As is almost universally true, all variables reflect such a good and bad pattern as burnout varies I→VIII.

Generalization About Incidence

The usual reservations about accumulating research apply, but the news is dominantly bad about the phase model, if not shocking. Several generalizations applied

**Scheme 1.**

to a panel of studies of incidence that numbered over 100 work sites from almost 20 nations, altogether including about 50,000 respondents.

First, overall, the United States and Canada report the least incidences of advanced phases along with New Zealand, but the base is high. In sum, over 24,000 cases reflect this distribution by phases among North American non-public work sites.^[6]

Second, the most favorable distribution of phases observed so far had over 22% in the three most advanced phases in any organization of any size ($N = 406$). Even this distribution will seem troubling to most observers.

Third, as a gross generalization, the global incidence seems worse. Roughly, about 10% more cases fell in the three most advanced phases; and 10% less are assigned to Phases I–III.^[6]

To be sure, the sampling of global work sites is not necessarily as representative as all non-North American work settings. So the generalization may have to be updated as the panel of national work sites is enlarged.

Fourth, there seem to be few outliers. Thus, phase data for New Zealand are substantially more attractive

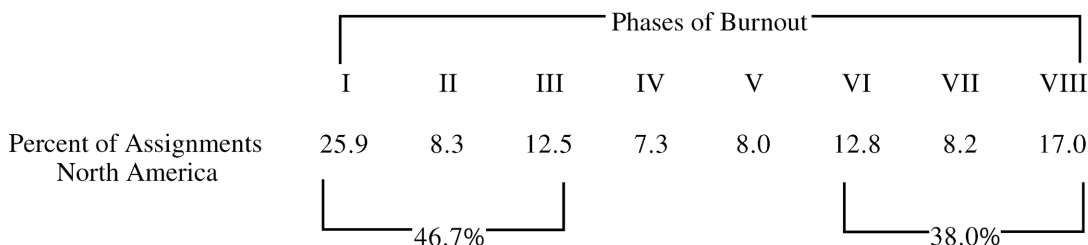
than for other non-North American settings, which is consistent with all other observations using various operational measures. Even this outlier estimate promises no bed of roses, however. To illustrate several recent estimates of the distribution of phases in six New Zealand work sites generated the following pattern:

Details on several of the New Zealand sites are conveniently available.^[8]

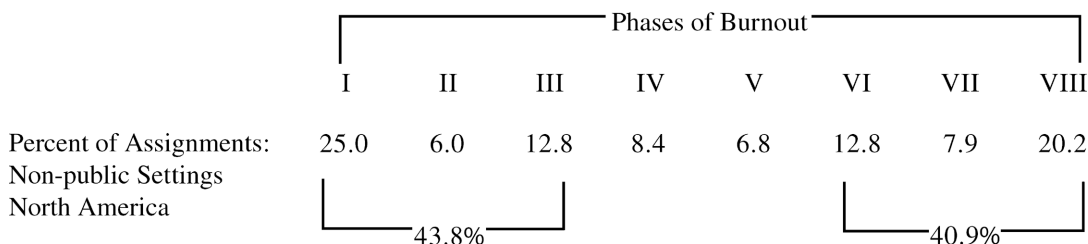
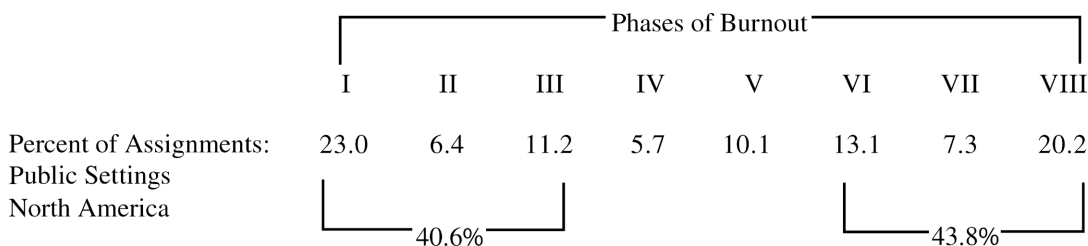
In sum, such estimates suggest that burnout poses serious issues even in New Zealand, a “good” environment for burnout. Had the distribution shown related to a physical disease like measles, for example, it would be interpreted as a pandemic, or far beyond “epidemic” status. Even “pandemic” might be too mild to label what exists, even in North America or New Zealand.

Arenas and Burnout Differences

Despite common opinion, public agencies do not appear to be markedly worse off in the matter of burnout than businesses. Available data do not permit a



Scheme 2.



Scheme 3.

fine-tuned conclusion on the point, especially because most public agencies for which incidence data are available come from North American contexts, i.e., the United States and Canada.

Basically, the public sector seems to suffer about the same, or perhaps a tad more, than available business work settings. To suggest the point, 9656 employees in 24 North American public-sector settings demonstrated this distribution in the eight phases of burnout,^[6] contrasted with 63 North American non-public settings (N = 24080).

Overall, this summary suggests no better or worse situation in public versus business employment.

CONCLUSION

These perspectives on the phase model of burnout are usefully placed in several broader contexts: a half-dozen-plus generalizations seem appropriate at this general level of analysis. A much longer and detailed list has been developed (e.g., Ref.^[2]).

1. Globally, it seems safe to conclude that we have a burnout problem, of substantial dimensions in many cases, and even worse in some work settings.
2. One set of items, the Maslach Burnout Inventory, is the most dominant instrument for assessing burnout. The phase model uses the MBI, essentially. Overall, the MBI items seem to tap three subdomains: depersonalization, personal accomplishment, and emotional exhaustion.
3. One basic disagreement exists among MBI users. The phase model assumes for testing that the three MBI subdomains tend to build to the most extreme burnout. Depersonalization builds with negative consequences on personal accomplishment. High levels of these two can lead to emotional exhaustion.

The major opposed view is that of Maslach and her supporters (e.g., Leiter and Maslach, 1999). The interest begins with heightening levels of emotional exhaustion, which if accompanied by high depersonalization, will generate decrements in personal accomplishment.

It is too early to make a decision concerning these two views. They may reflect different target phenomena, e.g., the full burnout sequence for the phases, and a basic focus on advanced burnout eruptions by Maslach and her supporters. For extreme cases, this distinction does not matter much, but practical and especially theoretical advances will occur once the issue is resolved.

4. Applications to ameliorate burnout stand in need of further development, to be sure, but the phase model has supported several successful efforts to reduce burnout. Thus, Golembiewski and his associates^[3] report designs for three type of situations: when burnout is rooted in interaction, policies and procedures, or in structure.^[9]
5. Evidence suggests that burnout will become an even greater problem over time, although this certainty is not appropriate. We are definitely in an era of downsizing or rightsizing, to suggest the point, and burnout predictably increases in work sites affected by adverse personnel actions (e.g., Ref.^[3]). Evidence establishes that advanced burnout seems characteristic of people against whom adverse personnel actions have been taken.
6. Institutional and policy linkages seem to influence burnout, but as of today, there has been no concerted effort to confront them (e.g., Ref.^[3]). For example, burnout seems most advanced among working single parents, especially mothers (e.g., Ref.^[10]), and one can envision supports in public policy for such subgroups.
The apparently high costs, both personnel and organizational, associated with advanced burnout suggest the value of attention to such linkages. However, no vigorous study of the economic costs of burnout now exists, as far as this author knows. The dour inventory of burnout covariants suggests these economic costs will be substantial; and those covariants also imply high social and economic costs.
7. Advanced burnout has been tracked in expected ways in its impact on a substantial panel of variables, but much work remains to be done. Consider only two points. Thus, most measures of burnout effects have relied on self-reports, i.e., on what are often labeled "soft data." Work with objective or hard measures (absence rates, costs of production, turnover, and so on) is growing. But, more attention to such data is necessary.

In addition, increases in burnout are reasonably enough associated with the outbursts of violence at work that seem to be stage-center so often these days. Testing of such linkages has not yet been made directly. Advancing burnout has been all but conclusively associated with affective states that seem related to violence. These states include depression, hostility, and a broader range of non-psychotic conditions that would warrant therapeutic interactions (e.g., Ref.^[3]). However, the specific linkages to outbursts of violence have not yet been made.

REFERENCES

1. Hager, G. Worker output plunges—productivity falls, labor costs soar. *USA Today* May 9, **2001**, 1A.
2. Golembiewski, R.T. Next stage of burnout research and applications. *Psychol. Bull.* **1999**, *84*, 441–446.
3. Golembiewski, R.T.; Boudreau, R.A.; Munzenrider, R.F.; Luo, H. *Global Burnout*; JAI Press: Greenwich, CT, 1996; pp. 83–128, 160, 239–244.
4. Maslach, C.; Jackson, S.E. *Maslach Burnout Inventory*; Consulting Psychologists Press: Palo Alto, CA, 1982.
5. Maslach, C.; Jackson, S.E. *Maslach Burnout Inventory*; Consulting Psychologists Press: Palo Alto, CA, 1986.
6. Golembiewski, R.T.; Boudreau, R.A.; Sun, B.C. Estimates of burnout in public agencies worldwide. *Public Adm. Rev.* **1999**, *58*, 59–63.
7. Quinn, R.F.; Staines, G.I. *The 1977 Quality of Employment Survey*; Survey Research Center, University of Michigan: Ann Arbor, MI, 1979.
8. Boudreau, R.A.; Everett, A.M.; Golembiewski, R.T. Burnout in New Zealand. *J. Health Human Resour. Adm.* **1999**, *23*, 37–43.
9. Golembiewski, R.T.; Rountree, B.H. System Redesign in Nursing, II: Action Planning and Its Effect on Worksite Stakeholders. In *Current Topics in Management*; Rahim, M.A., Golembiewski, R.T., Mackenzie, K., Eds.; JAI Publishers: Greenwich, CT, 1999; 197–211.
10. Ellay, D.; Anderson, K.S. An exploratory analysis of burnout among dual income and single income families. *J. Health Human Resour. Adm.* **1991**, *13*, 457–469.
11. Golembiewski, R.T.; Munzenrider, R.F. *Phases of Burnout*; Praeger: New York, 1998.
12. Leiter, M.P.; Maslach, C. Six areas of worksite: A model of the organizational causes of burnout. *J. Health Human Serv. Adm.* **1999**, *21*, 472–489.

Cambodia

Clay Wescott

*Government and Regional Cooperation Division, Asian Development Bank (ADB),
Manila, Philippines*

INTRODUCTION

After over 40 years of frequent, and unusually drastic, changes in its political and economic regimes, Cambodia has made headway since 1991 in rebuilding its governance institutions. Considerable progress has been achieved in maintaining public order, increasing revenues, rationalizing public expenditure policy and management, introducing new audit mechanisms, launching democratically elected commune councils, and facilitating a vibrant civil society, private sector, and free media.

PUBLIC SECTOR

Since independence in 1953, Cambodia has experienced frequent, and unusually drastic, changes in its political and economic regimes.^[1,2] Over the period, it has moved from constitutional monarchy, to republic, to extreme Maoist agrocommunism, to Communist party rule, and back to a constitutional monarchy with an elected legislature. The most unstable period was the period of Khmer Rouge rule from 1975 to 1979. During this period, the market economy and business activities were completely abolished, and there was no money or trade. No private ownership of any kind was allowed. Cambodia was cut off from the rest of the world except China, the regime's main supporter. The entire urban population was forcibly relocated to rural areas to perform agricultural work. Many educated civil servants, professionals, and military officers were executed or died from starvation or disease. People wearing glasses or able to speak a foreign language, both seen as symbols of higher education, were killed. Estimates indicate that more than 1 million people, or about 15% of the population, died of unnatural deaths.^[3]

Since the 1991 Paris Peace Agreement, Cambodia has rebuilt its key governance institutions. National Assembly members were selected in an election supervised by the United Nations Transitional Authority in 1993. Cambodia's sixth constitution was adopted in 1993 and amended in March 1999 to establish the Senate as a political compromise between Cambodia's two main political parties. The National Assembly

holds primary legislative power and has become more active since the formation of a new coalition government in late 1998. Although there was renewed political violence in 1997, the trend has been toward peaceful resolution of political disputes and increased democratic participation. The National Assembly has enacted several new laws critical for improving governance, such as the Financial Institutions Law and the Audit Law. In addition, television broadcasts of debates have increased the transparency of the legislative process. Meanwhile, the Senate is investigating complaints about illegal confiscation of land and has recommended amending the constitution to ensure that it has adequate time to review proposed laws.

The power to initiate legislation rests jointly with the Prime Minister and members of the National Assembly and Senate. Most draft legislation originates with the Royal Government of Cambodia (RGC), led by the Council of Ministers. The assembly has simply tended to review and enact bills drafted by the RGC, often without being given sufficient time and lacking the requisite expertise. In addition, there is no central repository for all regulations or subdecrees, and the legislature rarely receives copies of regulations or decisions by ministries related to laws enacted. Finding them consists of visiting each ministry to see what they have, making it difficult for citizens to know how to comply.

The imbalance between the legislature and executive is paralleled by an imbalance of financial and technical resources available, with far more donor resources going to the latter.

Executive power rests with the RGC, which directs the civil administration and armed forces. In 2003, Cambodia has 166,672 civil servants (including defense forces, but excluding police, whose numbers are unknown). In comparison to other low-income countries, the Cambodian civil service is about average in terms of cost (the wage bill as a percentage of GDP, current expenditures, and revenues) and size (civil servants as a percentage of population).^[4]

Low salaries in the public sector are one of the most fundamental structural problems, with direct implications for the sector's accountability and transparency (or lack thereof). If this issue is not addressed, it is likely to remain as a major obstacle to other governance reforms.

The RGC approved a *Governance Action Plan* in 2001, and the Council for Administrative Reform had made good progress on completing the civil service census and on a Priority Mission Group scheme (identification of staff groups to receive enhanced salaries linked to performance benchmarks). Perhaps the area of reform that had the clearest government commitment and leadership is decentralization. Following the announcement of the intention to hold elections in the country's 1,621 rural communes and urban sangkats in February 2001, the Law on Commune Elections and the Law on Administration of Communes were promulgated, and the interministerial National Committee for Support to Communes/Sangkats and its subcommittees responsible for implementing the decentralization reforms were established. The commune and sangkat elections were held in February 2002, and elected councils are increasingly taking on local government functions guided by the RGC's Decentralization Roadmap.

In addition, downsizing the huge defense and security forces is an urgent issue that requires a new vision for the armed forces and police. However, their ideal size is difficult to assess without determining their new roles in postconflict Cambodia. Military and security agencies need to promote professionalism, train their staff to observe new codes of conduct, and develop relevant skills. A reassessment of the roles of the military and police is under way.

The Governance Action Plan also identifies issues concerning the management of natural resources (including land, forestry, and fisheries) as critical to ensuring social peace, environmental sustainability, poverty reduction, and economic development. There are three important areas: 1) the resolution of land issues (i.e., ownership, classification, registration, and tenure); 2) combating corruption and mismanagement of forests; and 3) ensuring long-term sustainability of fisheries. Communities need to be empowered to play appropriate roles in the management of natural resources and to help ensure accessibility by the poor.

The RGC has undertaken several key reforms of public finance and administration since 1993. For example, the 1993 Organic Budget Law (basic law that defines the overall legal framework), combined with prudent fiscal and monetary policy, has helped fight inflation and bring about macroeconomic stability. The 1997 Law on Taxation has improved the government's revenue-raising capacity.

In 2001, total public revenues were only 11.7% of gross domestic product (GDP), one of the lowest proportions among countries, although up from 8.3% in 1998. The introduction of a value added tax in 1999 contributed to the revenue increases, while simplifying the tax structure, widening coverage, and reducing cascading (taxation of something more than once).

Improvements in tax administration have included better collaboration among government departments and strengthening tax auditing. Collection of tax arrears is expected to be stepped-up through several enforcement measures, including freezing bank accounts. However, further reforms in tax administration need to be considered, such as setting up a semiautonomous revenue authority and reducing the use of ad hoc tax and duty exemptions.

Over the period 1998–2001, the RGC has generally maintained fiscal discipline, and there has been a shift in expenditures away from the military and security toward social and economic development, with education spending doubled and health spending tripled; yet social spending is still too low to meet development goals. Furthermore, budget execution has suffered from delays and unpredictable release of funds because of cash constraints, thus undermining operational planning. The system is plagued by gate keeping, deficient accounting and reporting systems, and the use of cash rather than the banking system, leading to a weak control environment. The Ministry of Economy and Finance is currently undertaking several measures to improve budget execution and revenue-enhancing capacity, planning to raise total revenue to 14–15% of GDP by 2007.

The Governance Action Plan includes provisions to improve gender equity and to reduce systemic barriers to accessing services for women and girls. However, until an output-based, medium-term expenditure framework system has been implemented, it will be difficult to ensure that policies designed to increase gender equity are properly prioritized and funded.

Legislation establishing the National Audit Authority consistent with international standards was passed in March 2000. The authority is independent of the executive, reports to the National Assembly, enjoys financial and administrative autonomy, and is authorized to determine the scope and methods of audits. An auditor general and two deputy auditor generals are appointed for 5-year terms by royal decree at the recommendation of the government and approved by a two-third-majority vote of the National Assembly. The current auditor general is from the ruling party, and the deputies are from minority parties. Although these are promising developments, more training and additional equipment will be needed, along with considerable political will, for the authority to be fully effective.

The judiciary has four distinct components: the Constitutional Council, formed in 1998; the Supreme Council of Magistracy, established in 1994; the courts; and the prosecutors. The Constitutional Council safeguards the constitution and decides cases involving the election of National Assembly and Senate members. The Supreme Council is the prime guardian of

the judiciary's independence and is the only body empowered to discipline and appoint judges and prosecutors. The Supreme Court and Appeals Court are located in Phnom Penh, and each province and municipality has lower courts. There is also a military court. As of 1999, Cambodia had 117 judges and 54 prosecutors.

An independent, capable, and uncorrupted judiciary is the foundation of the rule of law and underpins the development of a market economy. Unfortunately, Cambodia's judiciary does not yet meet acceptable standards in this regard. The decimation of the legal sector in the 1970s and the socialist legal principles and processes instituted in 1980s still have a strong influence on the legal system.

Only 33% of judges and prosecutors have any formal legal education and like other civil servants receive a salary that is less than a living wage. Budget allocations are inadequate to cover other basic costs. Thus the general perception that the court system is riddled with corruption is hardly surprising.

The procedures for forwarding draft laws to the Constitutional Council are applied inconsistently. Only organic laws are consistently sent to the council for review. Contrary to the constitution, other laws are sometimes forwarded directly to the king for promulgation without prior council review. Rules governing how judges, prosecutors, and the judiciary in general function leave many issues of jurisdiction among courts unsettled and provide no standard for resolving jurisdictional disputes, although a modern set of rules has been drafted and is under discussion.

In 2002, the government made commitments to the following: 1) a time-bound legal and judicial reform strategy would be completed; 2) the Supreme Council of Magistrates would be restructured; 3) a Law on the Statute of Magistrates would be submitted to the National Assembly; 4) reported cases of corruption would be immediately investigated and prosecuted appropriately within the existing legal framework; and 5) an anticorruption law would be submitted to the National Assembly. In addition, Cambodia in 2003 became the 19th country to endorse the Anticorruption Action Plan for Asia-Pacific, designed in collaboration with the Asian Development Bank and the Organization for Economic Cooperation and Development. The goal of this initiative is to help countries work together to build effective anticorruption mechanisms.

CIVIL SOCIETY AND PRIVATE SECTOR

Since 1993, the government has vigorously promoted the transition to a market economy. Laws regulating

the private sector include the 1994 Law on Investment, the 1995 Law on Organization and Functioning of a Council for Development in Cambodia and the Cambodia Investment Board, and the 1997 Law on Taxation. New institutions were created to enforce these laws. The economy has started to recover since the downturn of 1997–1998, with real growth rate of 6.3% in 2001 and 4.5% in 2002. While the RGC's efforts thus far deserve credit, challenges remain in reducing transaction costs, both formal and informal. Weaknesses in land law and enforcement need to be addressed. Businesses need clear rules that guarantee their property rights, resolve commercial disputes, regulate anticompetitive conduct, and limit state interference. Promoting fair and consistent enforcement of laws is also important. Setting up monitoring systems to enforce newly enacted laws and to strengthen the enforcement capacities of relevant agencies may be useful.

Although non-governmental organizations (NGOs) had an important role prior to 1991 in the absence of other donor agencies, their numbers and the value of their assistance have increased dramatically since then, with NGO disbursements equal to over 11% of external assistance during the period 1999–2001. The government's attitude toward international NGOs is liberal while somewhat more restrictive toward national ones.

Other civil society organizations have also emerged since 1991. Establishment of the Cambodian Bar Association in 1995 was a notable development. Among civil society organizations, the financial position of Cambodian NGOs is generally weak and highly dependent on foreign sources of funding. Some NGOs appear to be under the strong influence of political parties, while others reportedly abuse their NGO status. However, many are providing much needed development services.

The lack of a legal framework for NGOs may allow some NGOs to abuse their status or give government officials the discretion to issue arbitrary decisions on the status of NGOs. To address these concerns, a draft legal framework is being finalized for discussion.

In the 1980s, media organizations were used for political party propaganda, and no privately owned media existed. Currently, Cambodia has more than 200 newspapers and magazines, including several foreign language newspapers, and a number of television and radio stations. The constitution guarantees the freedom of expression, press, and publication. The Press Law creates a legal framework.

While the Press Law guarantees more freedom than equivalent laws in other countries in Southeast Asia, journalists worry that vague definitions of terms present a risk that they could be used as excuses to suppress the freedom of the press. The media needs

mechanisms to curb irresponsible reporting, sometimes influenced by political parties.

CONCLUSION

Cambodia's experience in rebuilding its governance institutions since 1991 highlights that reform is a long-term process that requires setting achievable goals and providing consistent support. Considerable progress has been achieved in maintaining public order, increasing revenues, rationalizing public expenditure policy and management, introducing new audit mechanisms, launching democratically elected commune councils, and facilitating a vibrant civil society, private sector, and free media. While many challenges lie ahead, the commitment to the reform process demonstrated by the government and the achievements made to date are heartening.

REFERENCES

1. Summary of Cambodia governance assessment. In *Key Governance Issues in Cambodia, Lao PDR, Thailand and Viet Nam*; Wescott, C., Ed.; ADB: Manila, 2001; 5–19. http://www.adb.org/Documents/Books/Key_Governance_Issues/default.asp?p=govpub (accessed August 2003).
2. ADB and Cambodia Development Resources Institute. *Cambodia: Enhancing Governance for Sustainable Development*; ADB: Manila, 2000. http://www.adb.org/Documents/Books/Cambodia_Enhancing_Governance/default.asp?p=govpub (accessed August 2003).
3. Chandler, D.P. *The Tragedy of Cambodian History*; Yale University Press: New Haven, CT, 1991.
4. World Bank; ADB; IMF. *Cambodia, Enhancing Service Delivery Through Improved Resource Allocation and Institutional Reform: Integrated Fiduciary Assessment and Public Expenditure Review*. in press.

Capital Budgeting

Jane Beckett-Camarata

Department of Political Science, Kent State University, Kent, Ohio, U.S.A.

INTRODUCTION

Capital budgeting is the process of determining need and acquiring, constructing, improving, or purchasing capital assets. For our purposes, a capital asset is any large resource (e.g., land or buildings) that has benefits that extend more than 1 fiscal year. Capital assets are budgeted for in a capital budget. The capital budget is an integrated physical and financial plan for improvements. The capital budget is also the authorization for funding and acquisition of capital assets and must be approved by the legislative body.^[1] The capital budget plays a critical role in most public organizations since the life of capital assets usually extends beyond 1 year. When used in conjunction with the strategic plan, master plan, capital improvements plan (CIP), and debt management plan, the capital budget provides a way of systematically and efficiently acquiring and funding capital expenditures. Capital budgets are used extensively by state and local governments. The federal government currently uses elements of the capital budgeting process and it too is moving toward using a capital budget.

THE CAPITAL BUDGET PLANNING AND DECISION PROCESS

State and local governments are recognizing the importance of planning for capital improvements and tying capital improvements to the overall goals of the community through strategic planning and capital budgeting. The connection between strategic planning, capital improvements planning, and capital budgeting is important because such a connection has a potential positive influence on economic development and offers a way of increasing fiscal capacity. Inadequate planning can result in unnecessary expenditures of public funds.

The key parts of capital budget planning and decision-making process are

- Strategic plan
- Master plan
- CIP
- Debt management plan

The capital budget planning and decision process model is outlined in [Fig. 1](#).

Strategic Plan

For those state and local governments who have and use a strategic plan, the goals, objectives, and strategies outlined in the plan indicate how jurisdictions intend to fulfill their mission and address critical issues including capital needs. The strategic plan should be the basis for identifying capital requirements. The master plan should flow naturally from the strategic plan.

Master Plan

Most state and local governments are required to have a master plan. A master plan delineates current and future land use, and guides land and physical plant use. Master plans may contain, for example, site and utility drawings.

CIP

The CIP is a catalog of capital projects selected for funding for the next 5 to 10 years. Capital projects have an impact on future operating budgets. Each CIP should contain estimates for operating and maintaining the capital improvement after it is completed. The cost estimates include any expenses in connection with maintaining the improvement as well as costs of additional activities or new programs that will be using the new or expanded land/space.

The CIP identifies the needs of the state or local government and ranks the projects in priority order. The primary difference between the capital budget and the CIP is that the capital budget is a legal document that appropriates funding for specific projects. The CIP is important for capital planning for any local government, even if there are not a lot of large-scale projects.^[2] In addition to shaping thinking about future capital assets needs, the CIP compels decision makers to review, compare, and assign priorities to projects. If the capital and operating budgets are combined, the first year of the CIP is then included as the capital portion of the annual budget.

The CIP is a “rolling list” of capital projects. The first year of the CIP becomes the capital budget. Each year the CIP is updated (second year of the plan and a new year is added), and revenue and expenditure

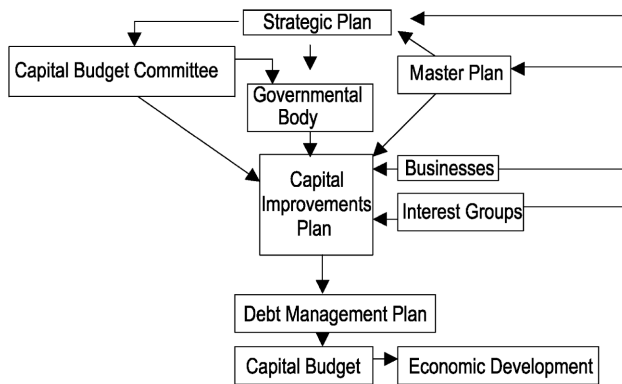


Fig. 1 The capital budget planning and decision process model.

estimates are revised for the remaining period. An effective CIP consists of both an administrative process to identify the location, scale, and timing of needed capital projects and a fiscal plan to provide for funding those projects. Producing a CIP is more efficient than conducting a new search each year for capital projects or capital equipment purchases.

If a state or local government has an existing capital budget, why does it also need a CIP? The capital budget represents the first year of the CIP. The capital budget is a legal document that authorizes specific projects during the following fiscal period. The capital budget lists only the projects and equipment to be carried out or purchased during the fiscal year. A CIP lists projects and equipment purchases that are expected and scheduled over a period of 5 years or more. CIP requires state/local governments to look ahead, see what their capital needs are, and plan for future projects. What are the population trends? Will more schools, prisons, or roads be necessary? How will such projects be funded? Will funding the projects require a bond referendum? Large and expensive projects such as roof replacement may need to be completed in segments, to hold down annual costs. The out years of the CIP are not binding and are therefore subject to change.

The typical CIP is a multi year projection that forecasts spending for all anticipated CIPs. Buying land for parks, airports, retention ponds, and public buildings is much less expensive when the population is small, and the land can be held in a relatively undeveloped state until needed. Long-term planning for water and sewer lines can also be helpful. For example, if projecting ahead discloses that future population estimates exceed current water production capacity, planning may be needed to expand and protect the watershed, build aqueducts, drill wells, or lay pipe to reach surface water. Such projects may require more than a 5-year planning range.

Many capital proposals come from the departments, for example, public works (street improvements),

utilities (water, wastewater, sanitation), and planning (as part of the master plan). Citizen complaints are a second source of requests for capital spending, although these may be funneled through departments or council members. Community-based interest groups can also be an important source of proposals for projects. Proposals also come from members of the business community and elected officials. Normally, the cost of these proposals far exceeds available financing.

Decision making on capital assets that go into the CIP is not simple and can be aided greatly if there are clear, detailed answers to the following questions:

1. How many citizens will be harmed or inconvenienced if the project is not constructed? Which citizens?
2. Adequacy of existing facilities to meet current and projected program demand.
3. Will the project add to the property value of the area, thus increasing the value of state/city property and receipts from property taxes? How much of an increase will be realized?
4. What services will be initiated, expanded, or improved because of this project?
5. What is the impact on those services?
6. Will the project increase the efficiency of performance? How much and where? What cost savings will result? Will the project reduce the cost of performance for a particular service? How much and where?
7. Will the project provide a service required for economic growth and development of the municipality/state?
8. Is the estimated cost of the improvement within the city's/state's ability to pay?
9. How the project fits the approved strategic plan and master plan?

Ultimately, priorities have to be decided and decisions have to be made. Sometimes analytical techniques such as cost-benefit analysis are useful, but judgment is also used. Some communities use a point system to establish priorities. One scale used is:

1. Urgent (highest priority)
2. Essential
3. Necessary
4. Desirable
5. Acceptable
6. Deferrable (lowest priority)^[3]

WHO USES CAPITAL BUDGETS?

Many state and local governments separate capital expenditures from operating expenditures. Eighty-four

percent of state governments separate capital expenditures from operating expenditures,^[4] 40% of counties^[5] separate capital expenditures from operating expenditures, and 71% of cities^[6] separate capital expenditures from operating expenditures. The federal government does not have a “full-scale” capital budgeting process. Funding for most federal capital assets is provided in annual appropriations under discretionary caps.

State and local governments use capital budgets because it improves decision efficiency since capital assets can be financed either by revenue raised currently (taxes, charges, grants, etc.) or by borrowing on the promise to repay from future revenues. If operating and capital expenditures are combined, capital assets appear more expensive relative to operating expenditures, even though they will be used over a number of years. Capital budgets are financed, but not necessarily balanced; operating budgets are balanced.^[7]

WHO IS INVOLVED IN THE CAPITAL BUDGET PROCESS?

Because capital expenditures represent a considerable financial outlay, who participates in the process is important to the success of the capital budgeting process. The main participants in the capital budgeting process are the government unit's (state or municipality) departments, a Capital Improvement Committee, a budget and/or finance department, legislative bodies, individual politicians, citizens, interest groups, and planning commissions.

The most traditional approach to formally soliciting citizen input into the capital budgeting process has been to form a capital budget planning committee. In the past, the committees have been made up of large, politically important interest groups in the municipality. Elected officials are becoming more sensitive to the need to increase the diversity of these committees for both equity and economic reasons.^[8] Some governments have also moved outside merely soliciting input from one citizen committee to broader representation, such as formally soliciting suggestions for capital improvement priorities from several citizen advisory boards representing different planning areas.^[9] Providing citizen input into the capital budgeting process may potentially provide the foundation for public support if a bond referendum is necessary.

The traditional informal approach is that which takes place through informal political bargaining in the capital budgeting process. Politics are an integral part of the capital budget process because elected officials must be perceived as being responsive to the electorate.

There are high power stakes in project selection. For example, the location of a new highway and its major intersections can produce windfalls for property

owners whose land borders the project. Professional staff can examine campaign promises for political priorities relevant to the capital budget.^[1] For example, the mayoral campaign may have promised downtown renewal, council members may have promised downtown renewal, and council members may have promised to revitalize the neighborhoods or to improve the physical plant of a school. Appointed officials usually have some priorities, such as adequate funding of maintenance projects and compliance with state and federal mandates.

INVENTORY OF ASSETS

Decisions to add a public facility or make extensive repairs can be made on the basis of understanding the needs of the community or the state and the resources available. The best approach to identifying the resources is to inventory the existing public facilities. The next step is to catalog the proposed public facilities. The key data in the capital register include:

1. Location of the proposed facility
2. Year of construction
3. Cost priority
4. Project description
5. Financing schedule
6. Prior or sunk costs
7. Projection of future-related funding requests
8. Operating costs
9. Savings in operating and maintenance costs

DEFERRED MAINTENANCE

In addition to cataloging the existing capital assets, they also must be evaluated to project repair, renewal, or replacement needs that will preserve their ability to support the mission or activities that they were assigned to serve.^[10] Condition assessment is critical to a capital budgeting program. To perform such an assessment requires both a useful database of existing conditions and the implications of such conditions for future performance.^[11] Condition assessment, cost estimating, and other asset management approaches are important individual strategies within a capital maintenance program. This is especially true where the highest priority is not acquisition of new assets but maintenance of old ones. Maintenance of capital assets is becoming an increasingly important subject because many public organizations must manage what they have and keep assets in service long after their “useful life” has expired. The challenge is to avoid deferral of necessary expenditures for maintenance, referred to as “deferred maintenance.” Deferred

maintenance can lead to asset failure that in turn can lead to replacement and extensive rehabilitation needs that are much more expensive than would have been the case if the assets had been kept in regular working condition. Many public organizations have poorly developed methods of measuring and costing deferred maintenance and its potential affects, and condition assessment is inconsistent. Because condition assessment is a prerequisite for deferred maintenance reporting, failure in condition assessment can lead to deferred maintenance.^[12]

COST ANALYSIS

Most state and local governments have limited resources, as well as legal political restrictions on their ability to raise revenue. For these reasons, a limited number of capital items can be acquired during any one period of time. Priorities have to be made to fund those that are considered to be the most important. Some combination of economic analysis, such as cost-benefit analysis, citizen input, and the political agenda, are factors that determine importance. A combination of quantitative, qualitative, and political is preferable. The key is to match available revenue sources with a wide-ranging set of expenditure options and to match expenditure options to revenue availability.

Cost-benefit analysis can be used to compare capital improvements. Not all proposed capital expenditures are associated with their cost relative to their benefits. Many decisions are made on other bases. State and local governments would not spend the level of resources they currently do on such things as hazardous waste disposal sites if it were not required by the federal government. Also, some decisions are made on the basis of equity rather than cost. For example, providing mass transit in low-use areas may not make economic sense. If these areas are also populated by low-income citizens, however, eliminating such facilities simply because they are not economical is often rejected by government agencies responsible for such facilities on the basis of equity.

FUNDING SOURCE IDENTIFICATION

Only a limited pool of resources is available to fund all the requested projects. The challenge is to identify the full range of resources available, both current and borrowed, and to determine how these funds can be used to meet the needs with the highest priorities. The resources usually considered are:

- Pay-as-you go financing—costs are paid directly from current income, such as taxes, fees and user charges, and interest earnings

- Grants from other governments
- Debt financing—tapping external funds through the issuance of debt in the capital market
- Public/private ventures, including privatization

Most governments find that current revenues are inadequate to fund all the projects in their capital plan. For this reason, a well-structured CIP should include a review of all financial resources that are available to meet current and future capital needs. The mix of financing methods that will be used to fund a CIP is likely to be one of the most important decisions made in developing and executing a CIP. In the case of debt financing, amortization rates and repayment schedules are determined by the parameters established by the mix of financing tools used. Each method of financing has advantages and disadvantages, and each is viewed in the context of its applicability and its effect on the government's overall spending levels and fiscal health.

DEBT MANAGEMENT PLAN

Rating agencies look favorably on state and local governments that assess affordability in managing debt. Debt capacity recognizes that state and local governments have limits to capacity to issue debt at a given credit level. As a general policy, a local government should decide whether to pursue a pay-as-you-go or a pay-as-you-use strategy. The pay-as-you-go approach relies on current revenues to pay for capital improvements; the pay-as-you-use approach relies on general obligation debt to pay for capital outlays. Using pay-as-you go financing, current taxpayers bear the burden of financing improvements and future residents reap the benefits. As long as the repayment of the debt does not exceed the life of the asset, a pay-as-you-use strategy may be more equitable than pay-as-you-go. However, debt funding of capital projects does have some obstacles. State law may require a voter referendum to obtain citizen approval for issuing bonds, and that approval may be difficult to obtain. Some local governments may face legal limits on the amount of debt that can be backed by the full faith and credit of the local government.

With the pay-as-you-go approach, intergenerational equity is violated. This means that those who benefit from the use of the capital asset are not necessarily the ones who finance the asset. For example, if taxpayers in a community use money from current revenues to build a swimming pool, the beneficiaries of the swimming pool may not be the same as those who put their tax dollars toward building it because some, for example, may move away. This is inequitable because they helped build a facility they have never had a chance to use.

The pay-as-you-use approach provides for intergenerational equity because it helps spread the costs of the capital asset over a number of years, which makes acquisition or replacement more feasible in the majority of cases. It avoids great fluctuations in expenditures and revenues. It also provides for intergenerational equity by allowing those paying taxes or user fees for a particular capital expenditure to have access to that asset.

The issue with the pay-as-you-use approach is that financing a capital asset might exceed the useful life of that asset. In other words, an asset should not be financed for 10 years if the asset will need to be replaced in 5 years. That type of borrowing can place the organization in a financially difficult position of using current revenues to pay for capital assets that no longer exist.

The debt management plan should deal with issues such as:

- What type of bond repayment guarantee to offer investors
- How much borrowing to allow
- What to do if actual expenditures are higher or lower than the funding set aside for the projects

To a certain extent, selecting funding sources for capital projects is a matter of piecing together the funds that are available in any one year. For example, a project may be funded from a mix of current revenues, state or federal grants, bond proceeds, intergovernmental contributions (e.g., from a county or from other communities benefiting from the improvement), reimbursements, and builder contributions. Improvements to utility services may be financed through grants, the sale of revenue bonds, and service charges levied on users.

Whatever the mix of funding sources for capital projects, there are underlying policy issues that must be dealt with—preferably through an explicit debt policy statement. A policy statement protects the operating budget from excessive debt service and sets a limit on the amount (and type) of borrowing.

THE CAPITAL BUDGET

The capital budgeting cycle consists of capital budget planning, development, council/legislative budget review, and capital budget execution. The operating and capital budgets may be prepared separately during the fiscal year, or they may be prepared concurrently. During the capital improvements planning stage, CIPs are proposed. Next the capital improvement requests are prioritized based on the strategic plan goals (if a strategic plan is used) and availability of funds. During

the capital budget development stage, departments submit their capital requests. The budget office refines each project budget and scope to determine if they are appropriate. This phase concludes with the highest-priority projects being included in the capital budget.

The beginning stages in the capital budget process become the link between the strategic plan and the CIP. With capital budgets, an ongoing problem exists, as with all public decisions concerning resource allocations, that is, how to fund capital priorities. When planning capital improvements, the question is “on what basis shall capital asset decisions be made to promote long-term (capital) activity A instead of short-term (operating) activity B?” The most difficult decision faced by policy makers is choosing which capital assets to acquire or replace. State and local governments usually have finite resources, and have legal and political restrictions on their ability to raise revenue. For these reasons, a limited number of capital items can be acquired during any time period. Priorities must be set to fund projects that are seen as most important. No matter how the various projects are ranked, rated, or prioritized, the funding of capital items is dependent on adequate revenue. An important difference between the capital budget and the operating budget is the method of financing. The process of developing a separate capital budget helps public decision makers to choose between funding specific individual capital assets or projects while staying within a total resource threshold.

What are some of the justification for including items in the capital budget?

- Essential public purpose
- Long useful life
- Infrequent and expensive

Unique Characteristics of Capital Budgets

Local and state governments have two types of budgets: capital and operating. The difference between the budgets is that the operating budget funds recurring activities, whereas the capital budget funds large non-recurring expenditures for capital assets. Capital budgets are different from operating budgets in two ways. First, capital and operating financing sources are different. Capital budget items are normally paid out of one-time, earmarked sources such as debt proceeds and grants. Segregating capital expenditures helps to ensure that revenues budgeted for capital items are indeed spent for those purposes.

Second, capital and operating budgets decision-making processes are different. Preparing a capital budget involves using information obtained from the CIP. One way to help the capital budget decision-making

process is to include a definition of a capital improvement/asset in the capital budget policy statement. Everything in the operating budget occurs within a fiscal year; capital projects, however usually take years to plan and implement. For example, road repairs may be defined as operating or as capital, or road repair projects intended to last less than a few years may be defined as capital. Whatever the definition, it should be used consistently, and the legal and policy implications of any deviation from this definition should be explored before a change is implemented. Unlike an operating appropriation, which is usually limited to 2 years' duration, a capital expenditure appropriation routinely may be retained, carried over from year to year, and spent until the project is completed.

Capital projects ordinarily require extensive planning reviews and site preparation. Any major change in a capital project can only be done at considerable expense. However, an amendment or revision in an operating budget can normally be during the current year. Capital projects usually have to be in sequence. For example, a street-resurfacing project may be preceded by replacement of water and sewer lines. Capital spending tends to be uneven because capital projects are often grouped in clusters. Operating expenditures tend to be relatively even throughout the year. Finally, capital projects normally require change orders and require closer monitoring of project performance, whereas actual operating expenditures are assessed against expected expenditure outlays (Table 1).

There is usually some discretion in determining what expenses relate to the operating budget and capital budget.^[2] For example, if a police squad car is leased, is it part of the operating budget? One method

for handling ambiguities relating to categorizing capital assets is to divide them into two types: ongoing departmental capital and large capital projects. Ongoing departmental capital includes furniture, office equipment such as computers and photocopiers, and minor remodeling. For departmental capital, the criterion of lasting a year or more and costing more than some minimum number of dollars is appropriate. In the case of motor vehicles, it is simpler to categorize all expenditures for passenger vehicles as either capital or operating, rather than to distinguish between automobiles that will last more than 1 year and those that will last less. For large capital projects, whatever labor and supplies are used can be considered part of the capital costs. The result may not be perfect, but it is clear and useful. The important point is to develop policies appropriate to the circumstances and then apply these policies consistently to avoid the potential of a lower bond rating or even to a qualified audit opinion.

The capital budget appropriates funds for the projects in the first year of the CIP and recommends the authorization of necessary bond issues. This legislative authority does not extend to the remaining years of the CIP, thereby necessitating an annual update of the plan.

After the proposed capital budget expenditures are approved, appropriated funds are allotted for expenditures as the capital project progresses through the various stages. Usually, funds are allotted so that preliminary site drawing work can begin. Awards are made from submitted bids, and the facility is constructed and/or lands purchased. To access the remaining appropriated funds, formal written justification needs to be submitted. In addition, capital

Table 1 Capital budget decision categories

Acquisition consists of obtaining any interest in real property, including improvements of any kind on the acquired land, except certain utility easements. All acquisitions are subject to the capital improvements process.

New construction is a single undertaking involving construction of one or more facilities. It includes 1) construction of or site work for a new facility; 2) any addition, expansion, or extension to a structure that adds to its overall exterior dimensions; and 3) complete replacement of a facility. If a new construction project meets one or more of the following criteria, it is subject to the capital budget process:

- It creates additional building space of 5000 square feet or greater.
- It has a total project cost of \$250,000 or greater.
- It is acquired through a lease with options to purchase or any other alternative financing approach.

Improvements are a complete and usable change to an existing facility or structure. Improvements are a complete and usable change to an existing facility or structure. Improvements include 1) alteration or conversion of interior space and other physical characteristics; 2) renovation to a facility or its infrastructure; 3) restoration of a facility or structure; and 4) major repairs to restore a facility or system. If an improvement costs \$500,000 or more, it is subject to the capital process.

Equipment is a tangible resource of a permanent or long-term nature used in an operation or activity. No precise criteria exist to help to determine whether equipment is an operating or capital expense.

Source: From Ref.^[13].

appropriation balances are normally reappropriated at the end of each fiscal year until the project is completed.

CONCLUSION

The capital budget plays a critical role in most public organizations because the life of capital assets usually extends beyond 1 year. In capital budgeting, planning and careful action is essential. When used in conjunction with the strategic plan, master plan, CIP, and debt management plan, the capital budget provides a way of systematically and efficiently acquiring and funding capital expenditures. State and local governments are beginning to integrate strategic planning, master planning, capital improvement planning, and debt management. Capital budgets are used extensively by state and local governments because they improve decision making as capital assets can be financed either by revenue raised currently (taxes, charges, grants, etc.) or by borrowing on the promise to repay from future revenues. Decisions to add a capital asset such as a public facility or to make extensive repairs to an existing facility should be made on the basis of understanding the needs of the community and the resources available. The capital improvement planning process is becoming less political. Some local governments have moved away from simply soliciting input from one citizen committee to broader citizen representation by formally soliciting suggestions for capital improvement priorities from several citizen advisory boards representing different planning areas.

REFERENCES

1. Bland, R.L.; Rubin, I.S. *Budgeting: A Guide for Local Government*; International City/County Management Association: Washington, 1997.
2. Bland, R.; Nunn, S. The impact of capital spending on municipal operating budgets. *Public Budg. Finance* **1992**.
3. Lynch, T. *Public Budgeting in America*; Prentice Hall: Englewood Cliffs, 1995.
4. Hush, L.W.; Peroff, C. The variety of state capital budgets: a survey. *Public Budg. Finance* **1988**, 8, Summer.
5. Halachmi, A.; Sekwat, A. Strategic capital budgeting and planning: prospects at the county level. *Public Budg. Financ. Manag.* **1997**, 8 (4), Winter.
6. Forrester, J.P. Municipal capital budgeting: an examination. *Public Budg. Finance* **1993**, Summer.
7. Mikesell, J. *Fiscal Administration: Analysis and Application*; Harcourt Brace: Forth Worth, 1999.
8. MacManus, S.A. Democratizing the capital budget planning and project selection process at the local level: assets and liabilities. *Public Budg. Finance* **1996**, 8 (3).
9. Griffin, M.; Hester, J. Freenets: cities open the electronic door. *Gov. Finance Rev.* **1990**, 6.
10. Rugless, J. Condition assessment survey/capital asset management. *Facilit. Eng. J.* **1993**, 21 (3).
11. Gordon, C. The fables and foibles of federal capital budgeting. *Public Budg. Finance* **1998**, 18 (3), Fall.
12. U.S. General Accounting Office. *Deferred Maintenance Report: Challenges to Implementation*; 1998; GAO/AIMD-98-42.
13. Darr, D. The benefits of long-range capital planning—The Virginia experience. *Public Budg. Finance* **1998**, 18 (3), Fall.

Capital Markets

Kenneth A. Kriz

*School of Public Administration, University of Nebraska at Omaha,
Omaha, Nebraska, U.S.A.*

Peter J. Kiedrowski

Institutional Investments, Wells Fargo, Minneapolis, Minnesota, U.S.A.

INTRODUCTION

When people need to purchase a car, a house, or other things of value, they usually turn to a bank to borrow the money necessary to complete the purchase. The money is capital—financial resources available to purchase something of value. One can say that the market for residential mortgage capital, for example, consists of (mostly) individual borrowers and bank or finance company lenders. The borrowers need the capital to finance the purchase of homes.

The term capital markets refers to the informal meeting of buyers and sellers of capital. We use the term informal, because most often, there is no particular meeting place for buyers and sellers (in contrast, the New York Stock Exchange is a formal market for the buying and selling of stocks listed on the Exchange). More likely, the parties meet through an intermediary such as a broker or dealer of municipal securities. Through the exchange of securities, such as bonds for cash, municipalities are able to raise funds to invest in infrastructure, to carry out economic development activities, or to carry out any number of functions related to the exercise of government authority.

Municipal capital markets are large in a financial sense. At mid-year 2001, the municipal market comprised \$1.6 trillion of debt outstanding.^[1] More than \$240 billion in long-term municipal securities were sold in 2000.^[2] They are also large in an economic sense. Many of the important facets of American life—schooling, transportation, recreation—would not be the same without the influence of the capital markets, both as a provider of capital and as a source of financial discipline for issuers.

In this article, we will sketch the dimensions of the municipal capital markets. We will first define the demand side of the market, detailing who is buying capital, why they buy it, and what they exchange for the purchase of capital. Then we turn our attention to the supply side and describe who provides capital in the municipal market.

THE DEMAND SIDE OF THE MUNICIPAL CAPITAL MARKETS

The primary purchasers of municipal capital are state and local governments and their affiliated agencies and authorities. These organizations buy capital by selling municipal securities, most often in the form of bonds. A bond is considered a debt instrument—a loan for a defined period of time at a specified interest rate. In exchange for capital, the government organization issues a certificate, or bond, that states the interest rate to be paid and when the loaned funds are to be returned (maturity date).

State Governments

State governments borrow to build infrastructure where the benefits of that infrastructure accrue to residents of the state as a whole. As of 1997, state governments had almost half a trillion dollars (\$454.5 billion) of debt outstanding.^[3] On a per capita basis, that is an average of \$1697 in state debt outstanding. However, there is tremendous variation in the per capita amount of state debt outstanding. Dividing the nation into nine somewhat arbitrary regions (and excluding Alaska, Hawaii, and the District of Columbia), Northeastern states have an average of \$4386 in debt outstanding, and Middle Atlantic states have an average of \$2779. Without those regions included, per capita debt outstanding drops to an average of \$1290. The majority of states fall near this lower average. Southwestern and Southeastern states have average per capita debt burdens that are much lower than in other regions.

Nearly three-quarters of state debt outstanding (73.7%) in 1997 was “non-guaranteed” debt, while the other one-quarter was “full faith and credit.” Full faith and credit debt is issued with the full backing of the issuing jurisdiction. That jurisdiction pledges that it will use all resources at its disposal to pay back bondholders. This implies that the government will raise taxes if necessary to repay the holders of debt.

Non-guaranteed debt lacks such a pledge. An example would be a city that issues debt to finance a new stadium. If the city issues full faith and credit debt (most often called general obligation or GO debt), regardless of the financial outcomes of the stadium, citizens of the city are responsible to repay the debt, usually through payment of property taxes to the city. By contrast, if the city issues non-guaranteed debt and the stadium is not profitable, citizens are not held responsible for repaying bondholders.

There is some variation among states in the use of non-guaranteed versus full faith and credit debt. Nine states (Idaho, Indiana, Iowa, Kansas, Kentucky, Nebraska, North Dakota, South Dakota, and Wyoming) have no state issued full faith and credit debt outstanding, and one (Colorado) has full faith and credit debt representing less than 1% of their total outstanding debt. At the other end of the spectrum, four states (Connecticut, Georgia, Nevada, and Washington) have more than 60% of their outstanding debt in the form of full faith and credit debt. With some exceptions, the heaviest users of non-guaranteed debt are Midwest and Great Plains states, while the heaviest users of full faith and credit debt are in New England, the Southeast, and the Pacific Northwest.

The largest category of debt outstanding for states is industrial development revenue bonds (IDRBs). These non-guaranteed bonds account for 42.6% of all state debt outstanding (they also account for 49% of debt issued by states in 1997). State and local governments issue IDRBs to finance improvements of private businesses. The business then makes some form of regular payment to the governments who pass them through to bondholders. The difference between what state and local governments pay in interest on debt outstanding and what the private business would pay is a subsidy to the business; the subsidy is offered in order to encourage capital investment in an area. Again, there is substantial variation in the use of IDRBs. IDRBs represent more than three-quarters of the debt outstanding of seven states (Alaska, Idaho, Montana, Nebraska, North Dakota, South Dakota, and Wyoming). Kansas has no IDRBs outstanding, and in Arizona and Texas, IDRBs account for less than 10% of debt outstanding. With the exception of Kansas, the heaviest users of IDRBs are Great Plains, Great Lakes, and Rocky Mountain states. Southeast and Southwest states use relatively less IDRBs.

The second largest category of debt outstanding for states is debt incurred for higher education purposes. It accounts for 11.3% of state debt outstanding. Florida has the largest per capita debt outstanding for higher education—\$541 or 49.5% of total debt outstanding. Other states with large per capita debt outstanding for higher education are Alabama,

Indiana, and New York. Georgia and Wisconsin, by contrast, have no outstanding debt for higher education purposes.

The Relationship Between State and Local Borrowing

When discussing local government debt issuance, it is most useful to divide the local sector into different types of issuers, such as school districts or special districts. This is because each type of issuer has distinct reasons for borrowing. However, before we move to different types of issuers, we make a general observation about local government borrowing as a whole. The amount of borrowing that local governments make seems to be inversely related to the amount that the state borrows (formally, a regression of local government debt outstanding on state government debt outstanding produces a negative and significant coefficient).

There are at least two explanations for this phenomenon, one structural and the other market based. The structural explanation holds that in some states, borrowing for certain activities is effectively reserved for one level of government. One example of this is IDRBs. New Hampshire has the second highest per capita amount of state IDRB debt outstanding, while at the same time having no local IDRB debt outstanding. Massachusetts has the fifth highest level of state IDRB debt outstanding and the 48th lowest level of local IDRB debt. In these states, borrowing to subsidize private business development has (explicitly or perhaps only implicitly) been designated a state function. On the other end of the continuum of state/local responsibility, the examples are Kansas with zero state IDRB debt has the sixth highest local IDRB debt outstanding, and Florida, with the 47th highest level of state IDRB debt and the ninth highest level of local IDRB use. In these states, local governments bear the burden of financing economic development through borrowing. So, there seems to be some evidence supporting a structural explanation of the relationship between state and local borrowing (furthermore, a regression of local IDRB debt outstanding on state IDRB debt outstanding also yields a negative and significant coefficient).

The market-based explanation is based on the notion of market segmentation. Instead of one large market for municipal capital, there are actually multiple segmented markets, at least one for each state and potentially more for different types of securities. The interest rate that an issuer must offer in order to sell bonds is based not only on prevailing interest rates and on the default risk of an issue, but also on

the amount of demand for municipal capital.^a In states where the state government issues debt quite often, the increased demand may bid up interest rates. This forces the local issuer to sell bonds at higher interest costs or to forego borrowing. While it would be nearly impossible to test this explanation directly, the market segmentation hypothesis has held up quite well in empirical tests.

Local Governments

Local governments are large issuers of municipal bonds. As of 1997, local governments had \$750 billion in debt outstanding. As mentioned above, there are several types of local government issuers, each with a unique set of needs for municipal capital. We will consider four types of issuers: counties, cities (along with townships with borrowing authority similar to cities), special districts, and school districts.

There is large variation in the amount and type of debt issued by U.S. counties. In total, counties had \$163 billion in debt outstanding, or about 21% of the local total. The average per capita county debt outstanding for states was just over \$610. But the standard deviation of per capita county debt is almost as large as the mean. The range of county debt is from essentially no debt outstanding (in Vermont, the figure is less than one cent per capita) to over \$2000 per person in two states (in Maryland and Nevada). In eight states (Maryland and Nevada, as well as Florida, Kansas, Kentucky, Mississippi, North Carolina, and Tennessee), county debt outstanding is equal to or greater than state debt. However, in most of the remaining 42 states, state debt far outpaces county debt. County debt issuance tends to be highest in the Middle Atlantic, Southeast, Southwest, and Western states. The Midwest ranks far below the other regions, and there is little county debt outstanding in New England states.

IDRBs are by far the largest reason for counties borrowing. Counties had \$62.4 billion in IDRBs outstanding in 1997, or about 38% of their total debt outstanding. In six states (Arkansas, Kentucky, North Dakota, Oklahoma, West Virginia, and Wyoming), IDRBs comprise over 85% of total county debt outstanding. In per capita terms, the largest state concentrations of county IDRB use are in the Middle Atlantic, Mideast (Ohio, Kentucky, West Virginia), Southeast, Midwest, and Upper Great Plains (North Dakota, Wyoming) states. States in the Pacific Northwest and Great Lakes regions have relatively little

county IDRB debt outstanding. Four states in New England (Maine, Vermont, New Hampshire, and Massachusetts) have no county IDRB debt outstanding. The next largest amount of county debt outstanding (about \$8.3 billion) is for elementary and secondary (K–12) education. Most of the debt outstanding in this category is concentrated in four states (Maryland, North Carolina, Tennessee, and Virginia).

Cities are by far the largest issuers of local debt. As of 1997, cities and townships had \$302.2 billion in debt outstanding. On a per capita basis, that amounts to about \$1100 per person. Again, there is some variation in city borrowing but not as much as with counties. Alaska, New York, Minnesota, and Colorado cities all have more than \$2000 in per capita city debt outstanding, while Idaho, Maryland, Mississippi, South Carolina, and West Virginia each have less than \$400 per capita. The heaviest city borrowers are located in the upper Midwest, lower Great Plains, and Southwest regions.

The largest reason for city borrowing is to provide utility services to residents. Cities had \$64.7 billion in outstanding utility debt, or about 21% of total city debt outstanding. Cities in the Southwest and West Coast regions tend to have more debt outstanding than in other regions. Most of the utility debt outstanding is for water and electric utilities. Within the general category of utilities, water utilities are 69% of the debt outstanding, electric utilities comprise 29% of the debt outstanding, and gas utilities the remaining 2%. The Southwest, Rocky Mountain, and Southeast regions have the most debt outstanding for water utilities, while the West Coast, Southwest, lower Midwest (Iowa, Nebraska, Missouri, and Arkansas), and Mideast (Kentucky and Tennessee) regions have the most electric utility borrowing. Most utility debt (81%) is non-guaranteed. City water utilities are most often run on a fee-for-service basis. The fees provide a dedicated revenue stream that can be leveraged with bond issuance to provide increased water system capacity.

The other major area of borrowing by cities is once again for industrial development. But, there is marked variation of city borrowing in this area. In 18 states, less than 10% of debt outstanding is IDRBs. However, in 10 states, more than one-third of borrowing is for industrial development. Also, there appears to be no geographic pattern to city use of IDRBs. The states with the heaviest use are spread throughout each geographic region.

There is surprising variation in the amount that school districts borrow to fund capital investment in K–12 education. School districts across the United States had \$105 billion in outstanding debt in 1997, or about \$393 per capita. School districts in 11 states had more than \$500 per capita in outstanding debt, led by Nevada, Minnesota, and Pennsylvania with

^aFor more on segmented markets, see Hendershott, Patric H.; Kidwell, David S. The impact of relative security supplies: a test with data from a regional tax-exempt bond market. *Journal of Money, Credit and Banking* 1978, 10 (3), 337–347.

more than \$900 per capita. Conversely, 10 states had outstanding school district debt of less than \$200 per capita, with five states (Connecticut, Indiana, Massachusetts, Rhode Island, Tennessee, and Virginia) having less than \$100 per capita outstanding. However, these figures are somewhat misleading. In these states, other government types borrow to fund capital investment in K–12 education. In Indiana, school districts are considered separate special districts. In Connecticut, Massachusetts, and Rhode Island, cities and townships borrow for this purpose. Finally, in Tennessee and Virginia, counties carry the bulk of K–12 debt. Once the borrowing by all types of government for K–12 education is taken into account, only one state, North Dakota, has less than \$200 per capita in outstanding K–12 debt.

The final category of borrower that we consider here is the special district. Special districts have more debt outstanding in aggregate than do school districts—\$179 billion in 1997. The purposes for which special districts issue debt are diverse. The largest category of debt outstanding is for electric utilities. Just over \$51 billion of special district debt for this purpose was outstanding in 1997. Once again, variation is high in electric utility debt. Seventeen states have no special district electric utility debt outstanding, and 25 have less than \$10 per capita outstanding. However, nine states have more than \$300 per capita outstanding, and three states (Nebraska, Utah, and Washington) top \$1000 per capita. Another large category of special district debt outstanding is for water utilities. The heaviest issuers of special district water utility debt are in the Southwest and Rocky Mountain regions.

In sum, the demand side of municipal capital markets shows a wide variety of issuers issuing debt for divergent purposes. These issuers may be in competition with each other for scarce municipal capital. We next turn to the supply side of the market for municipal capital.

THE SUPPLY SIDE OF MUNICIPAL CAPITAL MARKETS

Who provides capital to state and local governments? The suppliers (buyers of municipal securities) have changed over time. In 1980, commercial banks were the largest buyers of municipal securities, holding 37% of outstanding debt; individual investors through direct holdings, mutual funds, and fiduciary accounts held 34%; property and casualty insurance companies held 22%; and other organizations held 7%. By 2000, the individual investor became the largest owner of municipal securities. Individuals owned 71% of municipal debt outstanding in 2000, either directly, through mutual funds (which grew from 2% in 1980 to 31% in 2000), or in fiduciary accounts. Property and

casualty insurance companies owned another 12% in 2000. Commercial banks fell to 7%, and other organizations rose to 10%.

Individuals

Individuals buy municipal securities most often because the interest paid on the securities is in general not included in the calculation of income on federal income tax returns. The size of these “tax preferences,” and how they impact various buyers of municipal securities, is often the driving force behind the supply of financial capital. The Sixteenth Amendment to the U.S. Constitution (which created the provision for the federal income tax) does not authorize tax exemption of municipal securities directly; rather Congress must provide it and budget for its effects annually. The original authorizing federal income tax legislation contained a specific exemption for interest on municipal securities. There have been various attempts over the years to tax municipal securities interest, although there have been none in recent years, and the constitutionality of the proposals has been hotly contested. In 1986, South Carolina contested the right of Congress to require that municipal bonds be in registered or book entry form to be exempt from federal income taxes (Congress had done away with the right to hold bond certificates, with their negotiable bearer coupons, due to concerns about the sale of coupons outside of the normal coupon interest redemption). The U.S. Supreme Court ruled against the state of South Carolina on that issue, and also ruled that Congress had the right to tax municipal bond interest if it chose to do so. Congress could also change major provisions of the federal income tax (for example, create a flat-rate structure) that would diminish or eliminate the tax preferences of municipal securities.^[4] In most states, the interest from municipal securities issued within the state is also exempt from state and local taxes. In addition, interest income derived from securities of the territories of the United States is free from federal, state, and local income taxation in all 50 states.

An example of the tax reduction for individuals helps to illustrate the advantages of municipal securities over taxable debt securities. Assume a couple with an income of \$155,000 (which puts them in the 36% federal income tax bracket) files a joint return. They have \$30,000 to invest and are considering two possible investments: a tax-exempt municipal bond yielding 5% and a taxable corporate bond yielding 7.5%. If the couple invested in the municipal bond, they would annually earn \$1500 and pay no federal taxes on that interest. The taxable bond would provide the couple \$2250 annually, but they would have to pay \$810 in federal income taxes on that interest every

year, so their after-tax earnings would be \$1440 (or \$60 less than the municipal bond). The tax-exempt bond would be an even better investment in a state that had an income tax and exempted municipal bond interest from taxation. Market participants often speak of the “taxable-equivalent yield,” which is the interest rate an investor would have to earn on a taxable bond to equal a municipal bond with a specific interest rate. (In our above example, the couple would have a taxable-equivalent yield of 7.8125% for a 5% municipal bond, taking into account only federal tax consequences of the investment decision.)

There are other tax provisions that are important to individuals as they consider municipal securities. One of these is the taxation of capital gains. While the interest on municipal securities is exempt from the federal income tax, any gains or losses realized from selling municipal securities are treated as taxable securities sales and are subject to the capital gains tax, either short or long (depending on whether held for longer than 12 months before sale). Capital losses may be applied to capital gains to offset the capital gains tax, however.

Another provision for individuals is that certain types of municipal bonds are taxable, and certain types are subject to the federal alternative minimum tax (AMT is a set of provisions that restricts the amount of tax-free income that can be derived by an investor). Private activity municipal securities (such as IDRBs) and certain refunding bonds (which in combination account for over 10% of all municipal bonds issued) are included in the calculation of the AMT. A further complicating provision is that for some individuals, a greater portion of their social security benefits can be taxed if they have tax-exempt interest. There are also special tax provisions for buying original issue discount bonds (bonds sold at prices far below their face value, for instance, a \$5000 face value bond sold at \$2000), bonds at a premium (bonds selling above face value), and bonds with a market discount (bonds sold below face value but not by as much as original issue discount bonds). Individuals may not borrow to buy municipal bonds, and annual municipal interest must be reported on their federal income tax returns.

Property and Casualty Insurance Companies

Property and casualty insurance companies also own municipal securities in order to take advantage of tax preferences. These companies need tax-exempt income as their business fluctuates. In years of poor insurance underwriting results, they have less need for municipal securities with their tax exemption in their insurance portfolio, because they can offset their net income with the underwriting losses. In good underwriting years, they need the municipal income from their insurance

portfolio to be tax exempt to maximize after-tax profit. Property and casualty insurance companies tend to favor long-term bonds and revenue bonds in order to increase the annual interest. In 1999, 24% of the assets outstanding of these insurance companies were in municipal securities.

Commercial Banks

Commercial banks own municipal securities for tax exemption as well. As noted above, their share of municipal capital supply dropped over 20 percentage points from 1980 to 2000. The Tax Reform Act of 1986 removed many of the tax incentives for commercial banks, and the ones remaining were difficult for large commercial banks to utilize.

Commercial banks still utilize municipal securities for two purposes. First, they can buy bank-qualified municipal securities into their bank portfolio, and the carrying cost is tax deductible. Bank-qualified municipal securities are general obligation bonds for public purposes of jurisdictions (or revenue bonds issued for non-profit organizations within their boundaries) that offer no more than \$10 million dollars in securities annually. Small banks utilize these bonds more than the large banks, because of their size. Second, banks will actually make loans to governmental units. This helps fulfill their requirements under the Community Reinvestment Act. Also, government loans are more attractive in credit markets, where there are few opportunities to extend credit with reasonable risk to businesses. Commercial banks tend to favor short-and medium-term securities that better match the maturities of their bank portfolios. In 1999, only 2% of commercial banks' assets were in municipal securities.

Other Organizations

The other organizations purchasing municipal securities are small companies, large pension funds, and hedge funds. Small companies that are owned by an individual may choose to use municipal securities for their tax-exempt nature. The case with large pension and hedge funds is quite different. These organizations actually monitor the municipal markets for discontinuities between rate patterns on municipal bonds and taxable treasury securities. They believe that the municipal market will tend to the norm, and they hope to capture gains on temporary discrepancies in prices from the actual value of the bond (this technique of exploiting temporary mispricing between different types of bonds is called arbitrage). Municipal futures and options can be used for leveraged-hedging (providing some insurance against falling bond or stock prices) and speculation. Futures are securities that

pay a holder the difference between the actual price of the bond or bond index and some future price. Options are rights to buy or sell a bond or an index at a specific price at some future date.

Other Factors Important to the Buyers of Municipal Bonds

For all municipal securities investors, the provisions of each individual bond will also determine their attractiveness beyond their tax preference. A security's stated interest rate, call features (calls are options that the bond issuer puts in the bond indenture that allow the issuer to redeem the bond at a given point in time at a specified price), source of repayment, credit rating, liquidity, maturity, general economic conditions, securities, market conditions, and the perceived credit quality and sector of the jurisdiction influence demand.

CONCLUSION

In short, many institutions and individuals comprise the municipal securities market: Governments and

associated special districts demand municipal capital by issuing municipal debt. The buyers of that debt, including individuals by themselves and through mutual funds, insurance companies, banks, and other organizations, purchase the bonds for tax preferences and as part of their overall investment strategy. Purchasing the bonds provides capital to governments for economic development, education, basic public services, and for other capital-intensive purposes.

REFERENCES

1. Board of Governors of the Federal Reserve System. *Trends in the Holdings of Municipal Securities*; <http://www.bondmarkets.com/Research/munios.shtml> (accessed 20 August 2001).
2. Thomson Financial Services. *Short-and Long-Term Municipal Issuance*; <http://www.bondmarkets.com/Research/issuance.shtml> (accessed 20 August 2001).
3. U.S. Census Bureau. *State and Local Government Finances: 1997 Census of Government*; <http://www.census.gov/govs/www/estimate97.html> (accessed 18 July 2001).
4. *South Carolina v. Baker*. 485 US 505.

Capital Programming and Budgeting: Comparative Local Government Perspectives

George M. Guess

*Open Society Institute, Local Government and Public Service Reform Initiative,
Budapest, Hungary*

Costel D. Todor

Ministry of Justice, Bucharest, Romania

INTRODUCTION

One of the advantages of decentralized, autonomous local governments is that they should be able to plan and finance their own infrastructure and capital needs. To do this, they need capital budgeting processes that rigorously evaluate projects creating assets (those that meet useful life standards of 1–2 years, and cost threshold minimums) such as roads, water and sewer systems, solid waste, sanitation facilities, and projects for local economic development. Capital planning and budgeting “institutions” are the rules of the game critical to the success of local government programs. Program results typically depend upon the construction and maintenance of infrastructure. Institutional requirements include the efficiency of organizational processes, the efficacy of regulatory rules and systems, clear definition of management roles, and linkage of financial management processes and rules to service and program results. Such institutions are all critical to planning, financing, and implementing local capital programs.^[1] Based on training and technical assistance experiences overseas, particularly in Central and Eastern Europe, this article details 12 essential steps for the development of capital planning and budgeting systems.

CRITERIA AND STEPS TO STRENGTHEN LOCAL PUBLIC INVESTMENT INSTITUTIONS

Best international practices suggest that within a sound institutional framework there are 12 major steps that need to be followed in capital programming and budgeting (Fig. 1).

It is often tempting to think that additional funds can be used to avoid solving institutional problems that affect capital investment. The fallacy is that countries and cities can spend their way out of them. Almost 30 years ago, Caiden and Wildavsky^[2] pointed to the phenomenon of institutional “poverty” that

created revenue uncertainty and hindered fiscal management of projects. This was termed lack of “positive redundancy.” Absence of positively redundant institutions to identify and remedy problems can exist in the form of weak accounting and audit controls, lack of check and balance systems for regulatory oversight, and opaque and corrupt procurement practices. Such institutional weaknesses affecting capital investments can be found in the United States as well as in Honduras or Pakistan.

Step 1: Determine Organizational Structure

While there are several models available to determine the organizational structure, the one selected should ensure that the capital program and budget are based only on projects submitted by operating departments that meet economic [e.g., benefit–cost (B/C) ratios] and broader social criteria (e.g., water system project deemed essential to prevent risks to public health). The city will work out for itself the respective roles and responsibilities between the mayor’s office and the city council. Within the adopted structure, responsibility will be assigned to a unit (e.g., public works or budget office) for coordination of the entire 12-step process to ensure capital program implementation.^[3]

Step 2: Establish Capital Policies

The purposes of establishing clear written capital policies up front are: to facilitate planning and objective setting, to establish parameters for identification, ranking, and financing of projects, and to provide staff direction.^[4] These are advisory rules that citizens can use to hold officials accountable at the ballot box. They are not laws—which is very important for officials in rule-oriented civil law countries to understand (i.e., most transitional and developing countries). They are not laws to be circumvented or evaded that might be used as a pretext to send officials to jail.

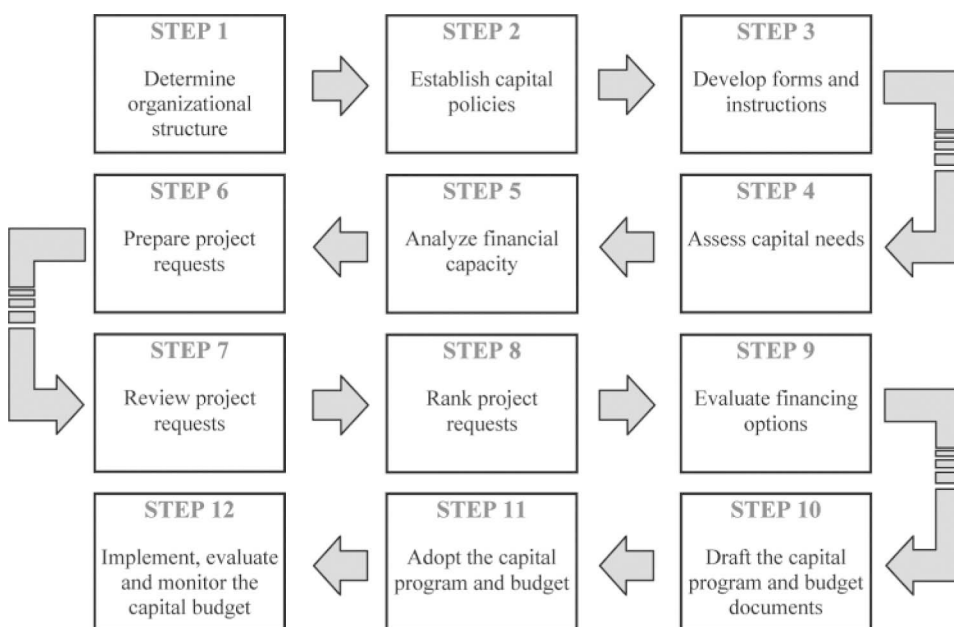


Fig. 1 Steps in capital programming and budgeting.

The policies serve the objective of fiscal accountability and should serve to resolve important issues. For example, the city might be working from an overbroad definition of “capital” in its budget accounting system. This could allow deliberate misclassification of current expenditures as capital items to balance the annual budget and provide political cover for spendthrift local officials. A new policy might distinguish major maintenance/rehabilitation (a capital expense) from minor and routine maintenance (a current expense) by a monetary figure (e.g., minimum of \$10,000/task is a capital expense). Thus, as a policy matter, a city could use a dual definition of capital that: 1) requires a minimum cost; and 2) creates a non-recurring asset such as land, buildings, and equipment with a useful life of 1 or 2 years. This would prevent financing of current expenses through the capital budget with debt in order to hide fiscal deficits.

Step 3: Develop Forms and Instructions

In the third step, cities must develop appropriate budget calendars, forms, and instructions. In many transitional and developing countries, these requirements are fixed by the central government.^a Officials need to know that they can change some features, especially the calendar, and provide additional fiscal and performance results information in annexes and

on other forms. Local officials should be rewarded for generating additional decision-relevant information by having their overall reporting burdens reduced—through elimination of redundant forms. Central governments have been slow to reduce reporting burdens, which acts as a disincentive to producing modern performance data useful for management decisions. There are several criteria here. First, the calendar should ensure that capital programming ends prior to the beginning of the operating budget cycle.^b This allows more accurate forecasts of maintenance and debt service requirements before the operating budget is approved. Using a multiyear framework with a separate capital programming period facilitates integrated preparation of the capital plan and budget. This makes fiscal decisions more rational even if institutional responsibilities are separate—and even if the budgets themselves are published separately.^c Second, initial forms should provide basic data on proposed projects, such as need, costs, and conformity to local plans. This is not the formal appraisal stage; they are project profiles—rough wish lists by departments. The capital programming staff should provide pilot testing of any new forms and give training on form preparation. Successful capital programs, such as in the City of Milwaukee, have developed a quid pro quo mechanism that allows elimination of redundant forms in exchange for

^aFor Romania, see Ref.^[5].

^bIn Romania, the budget calendar is fixed as per Ref.^[5]. The new law provides for completion of capital programming before the operating budget cycle begins.

^cFor Romania, see Refs.^[5,6].

required generation of new data by line managers and acceptance of responsibility for project performance.^[7]

Step 4: Assess Capital Needs

Local governments with effective capital budgeting systems will have systems indicating when capital replacement is justified or not to satisfy citizen needs. Officials face a spate of competing project profiles from departments each year. Which ones should be put in the program and later financed by the budget? The first question to be answered is: Which ones meet both technical and social need criteria? Effective cities will have two databases in order to determine whether to: 1) repair existing facilities; 2) replace them with new capital investment; or 3) abandon the existing facility and forget replacement. First, cities will have a regularly updated technical inventory of the age and condition of all its capital facilities. These data should be regularly generated by engineering crews from surveys of pavement, road, school, hospital, and sewer conditions compared to industry standards. Second, cities will perform regular citizen surveys of facility condition and performance. This would record and analyze complaints on service interruption. The two databases often produce different results. Engineers would want to repair bridges and buildings until the designated replacement age; citizens might be tired of patchwork repairs and want a new facility to prevent safety problems. Local officials need both kinds of data and transparent criteria on which to base their decisions, i.e., keep repairing (preventive maintenance) or schedule for replacement.^d

Step 5: Analyze Financial Capacity

Cities also need capacity to finance projects that compete for budget shares with other services and programs. Cities in transitional and developing countries pay for their shares of most capital projects via cash financing. The bulk of capital large-project financing is still grants from the central government or external donor, e.g., European Commission Poland-Hungary Aid for Reconstruction of the

Economy (EC PHARE) or the European Bank for Reconstruction and Development (EBRD). As external sources dry up, cities will need to replace these funds with own-source revenues. One of the best sources of local financing for projects in Organization for Economic Co-operation and Development (OECD) or more advanced countries is credit. Larger cities in developing and transitional countries and medium-sized cities that demonstrate sound fiscal management also have borrowed from banks or issued debt in the form of bonds. Rating agencies, such as Moody's, regularly evaluate country sovereign bonds and use roughly the same criteria for sub-national issuers. The general categories are: 1) demographic health—age of population and workforce, and local and regional migration patterns; 2) financial condition—budget and revenue trend analysis, per capita debt as percentage of total revenues, debt service as percentage of total budget, and future debt capacity; 3) economic condition—per capita income, unemployment rates, and industrial trends; and 4) management performance—on-time approval of budgets, tax collection rates, and staff turnover.^[9] Evidence that cities are generating valid and reliable data in these categories and using them for management purposes invariably leads to higher creditworthiness ratings. Improved fiscal condition facilitates local infrastructure financing. Properly borrowed funds can reduce financing costs (since payments no longer have to be made by the end of the fiscal year) and permit reallocation of savings to other capital project and current service needs.

The stronger the financial condition as determined by these criteria, the easier it is for cities to borrow in the market for capital needs. Some countries, such as Romania, have growing local bond markets. More than 25 cities have issued and repaid their bond debts. At the same time, smaller and medium-sized cities have lower debt repayment capacities. Lending by banks to cities with little credit history and less financial management capacity is risky for both banks (which also have little experience in market-based lending) and the central government (since it could increase contingent liability and weaken its credit rating for sovereign bonds). Local officials are often unfamiliar with the mechanics of negotiating cheaper loans on the basis of good creditworthiness. For many years, they had been used to receiving subsidized credit from banks that they did not have to pay back on time (especially by loss-making city enterprises). They were also used to receiving capital grants from the central government for which they had only been responsible for covering maintenance costs. The problem is that local governments rarely covered maintenance needs.^e Cities have also been used to engaging freely in odd governmental accounting practices that hindered fiscal

^dMany cities lack inventories of existing assets or any systematic way of assessing facilities' conditions. By contrast, Milwaukee uses such mechanisms as its Pavement Condition Index (PCI), which allows it to set "outcomes" or results objectives, such as preservation of the quality of life with a 6.3 PCI rating. The 6.3 score represents a weighted average of technical/citizen survey results. Cities in transitional and developing countries would like to have more technical condition data but are reserved about citizen inputs in many cases. Using weighted averages can allay some of these concerns. See Ref.^[8].

transparency. For example, Japanese road corporations still count interest paid on loans to state banks as assets.^[11]

Step 6: Prepare Project Requests

Systems and procedures are required to evaluate project requests. Public works officials are often engineers or planners, especially in Central-Eastern Europe, meaning they like numbers (especially large construction budgets that keep everyone employed). Unfortunately, they have been rarely forced by local fiscal institutions to consider the opportunity costs of projects competing for scarce funds. The view by engineers in line ministries is that all capital projects should be funded immediately. In transitional countries, engineers have often been trained under the Soviet system and believe the numbers provide a scientific aura that will persuade European funding sources such as EC PHARE or EBRD to fund their projects. They do not consider that such donors have to choose between many worthy projects in many countries. Local planners and engineers have not yet learned that engineering statistics or B/C calculations alone rarely provide sufficient justification for project funding. Local officials in newly developing cities need to go through seven steps, none of which are purely technical:

1. Effective capital planning and budgeting institutions will define project objectives based on actionable problem definitions. If they allow the problem to be defined by the consulting engineer, this could eliminate feasible alternative solutions and increase local budget costs later. For example, while the City of Venice is sinking and the sea is rising, the consulting engineer defined the problem as the absence of expensive floodgates that the city should build and profit from. This should have raised warning flags in the city finance department.^[12]
2. Fiscal institutions will require rigorous analysis of alternatives via full review (including non-capital alternatives, e.g., enforce the leash laws

instead of building a new dog pound) and cost effectiveness—where the objective is given, such as better health care, and analysts find the most efficient ways to achieve this.

3. Project benefits will be estimated accurately. In contrast with the old Soviet norm-driven system, which used measures of physical capacity (e.g., beds per capita) to compute demand, Western analysts use market value and consumer surplus measures.^[13] Local officials will recognize that benefits estimates are not facts. Local officials also will adopt criteria that distinguish indirect from direct benefits. Including indirect benefits as direct benefits inflates the value of benefits and distorts overall project analysis (e.g., agricultural marketing and income benefits to farmers 800 miles away from a proposed barge canal).
4. Fiscal institutions will estimate project costs accurately to compare with benefits. Here, analysts need to measure total costs (direct and indirect) and opportunity costs (costs of alternatives not selected).
5. Before project costs are compared to benefits and placed in an overall program, local institutions will establish prior decision rules by which individual capital projects will be selected for the capital program. There are four major criteria: 1) the payback period—consisting of projections of net results with undiscounted costs and benefits; 2) B/C ratios—which weight the estimated costs and benefits for differences in timing and the time value of money; 3) net present value (NPV)—the monetary value or present worth of the project with discounted values; and 4) internal rate of return—the discount rate at which the B/C ratio is 1 and the NPV is 0. Most of these rules are useful for distinguishing large projects that have longer term construction schedules. They are less helpful in selection of smaller projects that are all needed by the community, e.g., water pipes, sewerage treatment, clinics, schools etc. Formal analysis of projects via one or more of these rules should be one of the criteria used to determine the final capital program.^[14] Thus, all project proposals should be subjected to formal economic analysis.
6. The next step in project analysis is to compare undiscounted costs and benefits (the payback method noted above). This produces preliminary estimates that are usually not sufficient for decision-making.
7. The last project analysis step compares discounted costs and benefits. City fiscal institutions will carefully review the discount rate used

¹¹The irrational split between fiscal authority and program management responsibility has been a major disincentive to capital maintenance in transitional and developing countries. Central governments tend to provide capital funds if the municipal units pay for maintenance of schools, hospitals, etc. In the United States, the same poorly design arrangement for intergovernmental capital financing can be identified in transit grants to cover buses and rail cars. When local governments fail to pay for maintenance, they can use new grant funds to replace the cars and buses. For further discussion of this problem see Ref.^[10].

by analysts. This is important since a low rate underestimates the value of current consumption or investment and favors long-term, large projects. The question is: What criteria should local officials use to select the discount rate? There are two major methods or criteria. The first, called opportunity cost of capital (OCC), focuses on profits, and the second, the social rate of time preference (SRTP), considers the welfare of future generations. The former methods normally produce a higher discount rate, meaning that opportunity costs of this public investment are higher and future benefits really need to outweigh the costs of alternative fund use to be approved (e.g., very high demand energy projects). A lower rate, often derived from the SRTP method, results in approval of projects that may not be profitable at all but will produce social benefits (e.g., water purification systems).^[15] Even in OECD countries, analysts use politicized or lower discount rate to generate higher B/C ratios. For example, a billion dollar rail project in one U.S. city used a 6% discount rate when the rate of inflation was 20%. Coupled with inflated ridership forecasts, the project was finally built and predictably benefited few people.^f The important point is that institutions must be designed to respond to powerful pressures by project advocates to distort figures in their favor.

Step 7: Review Project Requests

Cities with effective capital planning and budgeting systems have professional bodies to conduct at least four basic types of project review according to empirical data and professional standards: 1) administrative—all required forms, approvals, citizen surveys, environmental impact statements, and other compliance requirements have been met; 2) planning—fiscal and demographic impact studies have been performed, i.e., How much will supporting local services cost for this project? Is the project consistent with physical land use, sectoral, and city strategic capital improvement plans (CIPs)?; 3) financial—Have all capital costs been included in the proposal? Are costs and benefits properly estimated and measured? Has the fiscal impact on local tax rates been estimated?; and 4) engineering/architectural—Have

alternative approaches such as continuation of repairs been considered? Is the project construction schedule realistic? It is important that these issues be reviewed and checked off by appropriately qualified institutions before financing plans are made.^[16]

Step 8: Rank Project Requests

Next, cities rank the full list of approved projects, only some of which will be financed with limited funds available. The process of project appraisal and approval only produces a list of cost-beneficial projects. This is not the final capital program. Nor is it the capital budget. The list does not compare projects in different sectors with each other or indicate objectively which ones should be approved. Using B/C ratios, for example, one cannot compare the benefits of educated students and productive farmers with those of hospital patients receiving quality health care. Methods are used to approximate the comparative benefits of different projects for dissimilar groups. The ranking system used by many OECD cities adapts economic methods of interpersonal utility benefit models and marginal productivity curves to administrative practice. Consistent with this adapted method, to develop a capital program, the city needs to develop a rational system of transparent project ranking. Before projects are prepared and reviewed, the ranking system must be agreed upon by multiple stakeholders, including citizens groups, community watchdog groups, producers, and consumers.^g

Cities with effective capital systems make certain that local stakeholders agree on a weighted ranking system that incorporates multiple values and criteria. This may be the most difficult part of the entire capital planning and budgeting process. Personal, partisan, and institutional problems often combine here to work against modernization of the local capital budgeting and planning processes in developing and transitional countries. Some OECD cities use single standard systems, typically focusing on B/C ratios. The problem is how to use such ratios to compare the worth of projects in different sectors such as health and transportation. Many OECD cities use value categories, such as urgent (danger to health), desirable (to replace obsolete facilities), or deferrable (duplicates other projects). But on what basis can we achieve consensus on these categories among many community groups? Defining agreeable criteria for such categories as “desirable” projects could waste substantial amounts of time and political

^fMuch has also been written about the self-serving project analysis and review of the Dade County Metrorail system done in the 1980s. Low discount rates, high ridership projections, and lowball cost estimates by omission of costly items such as pedestrian bridges and sound panels produced a feasible-looking project for rubber-stamp approval.

^gSee Refs.^[14b,17].

capital. Thus, the multiple criteria approach used by most cities in advanced countries allows balancing of competing needs and efficiency-welfare values, such as economic worth, existing occupancy or user rates as a measure of need, fiscal and economic impact, health, safety effects, and environmental costs/benefits. Values and priorities are made explicit and quantified. City officials typically propose their own criteria (e.g., level of public supporting services required; availability of cost-sharing, and forecasted demand), weights for each criterion (i.e., facilities with poor existing conditions should receive higher weights than those in better condition), and scores from which to determine project rankings.^[14b] The framework has a strategic value in allowing more transparent cutbacks of projects in times of revenue shortfalls—or budget additions in time of buoyant revenues. While agreement on criteria and weights can be a lengthy and contentious process even before projects are inserted for overall scoring, many cities in advanced countries use this approach.

Step 9: Evaluate Financing Options

Cities with effective capital planning and budgeting systems develop financing packages consisting of multiple sources of project finance. The ultimate mix of grants-loans will be affected by the fiscal condition of the city, noted in Step 5. Cities typically develop mixes of current pay-as-you-go and debt funds to pay for their projects. Poorer cities such as Milwaukee rely on self-supporting debt (enterprise funds, special assessments, and tax increment districts) and cash-tax levies (mainly for small works such as neighborhood alleys and street lights) along with central transfers to finance about 45% of their infrastructure needs.^[18] The rest is covered by tax-levy supported debt, meaning general obligation bonds. To issue bonds, the city has to maintain a solid credit rating based on demonstration of sound financial condition. In the case of Milwaukee, financial officials were able to counter low per capita income and sagging economic growth data with evidence of strong financial management controls and predictable budgets. This produced a relatively high credit rating by Moody's and allowed them to issue debt at lower interest rates and yields (lower financing costs) than if they had received a lower rating.

Cities can also combine current tax revenues with creative approaches such as leasing or use of build-operate-transfer (BOT) contracts to obtain both capital financing and operating expertise.^h Depending on their relative fiscal condition in the hierarchy of municipal finance, cities in developing and transitional countries also: 1) borrow from municipal development banks at subsidized rates; 2) borrow from

suppliers/vendors at usually high rates; or 3) issue bonds or borrow from commercial banks if they are in the category of advanced cities. More advanced cities (meaning those with experience and skills in financing—not necessarily the wealthiest cities) use innovative financing packages that include such vehicles as tax increment financing districts (TIFs or TIDs). TIDs are city economic development districts that pay back the principal and interest on debt issued for new facilities and improvements within their boundaries from the expected increase in property values within those districts. Also called “betterment districts,” Milwaukee has established over 50 of these since the 1970s.ⁱ In short, city officials need to improve skills and systems for generation and analysis of fiscal data. This will allow them to compare financing options in order to minimize burdens on their budgets while maximizing the number of projects they can finance.

Step 10: Draft Capital Program and Budget Documents

Cities must draft and disseminate the capital program and budget in transparent budget documents. The goal should be to explain how the 5 to 6 year capital program links to its first year of financing through the capital budget and to indicate what results can be expected. This cannot be accomplished by presenting reams of accounting figures and budget laws. Despite the limited public audience for such documents (which is an implicit purpose of this approach), many transitional and developing country cities still do this. For transparency, it is essential that the city adopt a budget structure and format that simplifies and explains fiscal information to the reader. Public finance officials, especially accountants, seem to have a difficult time doing this. As a result, budgets are greeted with suspicion in many cities by community stakeholders and the media. Lack of transparency—presentation of figures and their meaning in plain language—is an important reason for the often ferocious conflicts between city councils and executive/mayor offices over financial issues. Skills are needed to combine fiscal data and narrative explanation. In many cities, too many data are available, which encourages officials to bombard the public with fiscal data that have not been clearly translated into meaningful information. This can stifle interest in and debate on budget matters and decrease public sector accountability.

By contrast, some cities in transitional and developing societies have developed capital improvement

^hSee Refs.^[18b,19].

ⁱRef.^[18a]. See also Ref.^[20].

programs and budgets consistent with the highest OECD standards.^j

As a leading U.S. example, the annual Milwaukee budget provides basic annual data in a user-friendly format that can easily be copied elsewhere.

The presentation consists of two parts:

1. The budget is prefaced by a transmittal letter from the mayor to the council. Some cities also publish the approved final budget with a letter from the chair of the council finance committee. It is important in these letters to describe revenue and expenditure trends, capital policies, factors considered important in setting priorities, percentage of projects by category, financing sources, changes from past year (e.g., projects dropped), unmet needs (e.g., deferred maintenance), and results of citizen surveys. The finance office should summarize here the information noted in the previous nine steps for use by the media and the public.^[18a]
2. The Milwaukee budget provides highlights and summary schedules of each project by department, purpose, special fund, priority, future costs in debt service and maintenance, and financing source. The Milwaukee budget consolidates the revenues and expenditures of all units of city government, including special authorities, extra-budgetary funds, and city enterprises. It is important to link current and capital budgets together (e.g., break out personnel costs, debt service, and maintenance related to capital expenditures).
3. The City of Milwaukee uses a unique approach that combines current and capital expenditures in relation to defined “outcomes” or programmed results measures. The integration is logical in that particular capital facilities, such as hospitals, are supposed to provide quality patient days. Since this outcome is a combination of current and capital expenditures—functioning equipment, available medicines, and trained personnel—it is important to measure capital expenditures with a results-oriented framework similar to this. For example, the Milwaukee water department is required to report on and respond to the “percentage of main water breaks back in service within 24 hours.”^[18a] The budget indicates that the department could not produce this outcome in fiscal year 2003, which called attention to an aging capital stock

(pipes) and the questionable efficacy of its maintenance program. The outcomes budget format allows the city department to be held accountable by elected officials and the public; it also allows the department to use its information to request more funds for capital and current items.

Step 11: Adopt Capital Program and Budget

The city then adopts the capital program and budget.^[22] City governing bodies need to have time for review of the proposed program and first-year capital budget. Time is required for inputs via citizen surveys, citizen review on capital program committees, and public hearings. After review, the governing body follows one of two courses: 1) adoption of the capital program and 1-year capital budget; or 2) adoption of the capital budget only. In the second case, some city governments fear being locked into a multiyear program of expenditures. The way out is to adopt policies that indicate that the capital program is a rolling plan that each year (in some cases every six months) will be rolled forward to take account of changes in both financial projections (e.g., revenue shortfalls) and the project portfolio (deletions, modifications, additions). In this way, the government is free to modify the program based on regular reconciliations of plans and budgets.

Step 12: Implement, Evaluate, and Monitor Capital Budget

Following approval, cities will monitor and evaluate the progress of their capital budgets. Sound project management and financial control is essential for the best project results. This is difficult to achieve even in advanced countries, and cities must constantly review their capital implementation institutions to ensure proper functioning. For example, the City of Milwaukee faced regular cost overruns and conflicts over change orders that were driving up project costs. The capital budget noted that authority and responsibility for project design/construction and oversight/accountability were divided between the Public Works Department (DPW) and the Budget Management Department (BMD), which had led to major communication and control problems. As indicated in its budget, Milwaukee recently adopted a new approach that centralized normative control of capital account management in the BMD (requiring individual departments to report regular fiscal and physical progress of projects), centralizing project design/construction management in DPW, and decentralizing expenditure authority to the individual departments to take

^jSee for example Ref.^[21], which includes a multiyear projection of borrowing and debt service requirements through 2008.

responsibility for projects in their sectoral areas. Thus, dual project oversight is provided by DPW–BMD.^[18a]

SEQUENCING CAPITAL PLANNING AND BUDGETING REFORMS—CONCLUSIONS

It should be clear from reviewing the 12 steps that the most difficult question is how to design them and to sequence their installation. Because of resource and absorptive capacity constraints in most countries, it is impossible to do everything at once. Based on overseas experience, implementation of a capital programming and budgeting system should likely follow this four-step sequence:

1. The city should fix the organizational framework. Cities need to clarify planning and implementation responsibilities and eliminate negative institutional redundancies, such as excessive approval steps that produce corruption opportunities and drive up project costs. Cities should perform reviews of core transactions in capital planning, such as project approvals and contracting, to ensure that these processes are performed as efficiently as possible.
2. Professional management skills need to be improved. Managers should have the skills (as well as the authority) to analyze project proposals and develop weighting and scoring matrices from which to develop overall capital programs.
3. Capital plans and budgets should be integrated with the current budget process in a multiyear framework. This allows the finance department to oversee planning of the capital program and to ensure that financing is available for future years to cover maintenance and debt service requirements.^[23]
4. Cities need to improve their practices of municipal budgeting and financing. This means more than just installation of computerized accrual accounting and integrated procurement and budgeting systems. Officials need to know how to operate and modify such systems to meet their future needs. They also need to know how to apply innovative options for project financing, such as leasing, tax increment districts, and note/bond relationships with commercial banks.

REFERENCES

1. Burki, S.J.; Perry, G. *Beyond the Washington Consensus: Institutions Matter*; World Bank: Washington, DC, 1998.
2. Caiden, N.; Wildavsky, A. *Planning and Budgeting in Poor Countries*; John Wiley: New York, 1974.
3. (a) Bland, R.L.; Rubin, I.S. Planning and budgeting for capital improvements. In *Budgeting: A Guide for Local Government*; Bland, R.L., Rubin, I.S., Eds.; ICMA: Washington, DC, 1997; 167–196; (b) Forrester, J.P. Municipal capital budgeting: an examination. *Public Budg. Finance* **1993**, *13* (2), 85–103.
4. Matzner, J.J., Ed.; *Capital Projects: New Strategies for Planning Management, and Finance*; ICMA: Washington, DC, 1989.
5. Romanian Cabinet. The Emergency Government Ordinance No. 45/2003 of Local Public Finances, article 59. The Official Gazette No. 431, 2003; June 19.
6. The Federation of Local Authorities from Romania (FALR). *Guide for Applying Government Ordinance No. 45/2003*; 2004.
7. Presentation by Anne Kinney, Director of Budget, City of Milwaukee, at Government Finance Officer's Association headquarters, Chicago, IL, 1997.
8. Fisher, G.W. *Financing Local Improvements by Special Assessments*; Municipal Finance Officers Association: Chicago, 1974.
9. (a) Kleine, R.; Kloha, P.; Weissert, C.S. Monitoring local government fiscal health: Michigan's new 10-point scale of fiscal distress. *Gov. Finance Rev.* **2003**, *19* (3), 18–23; (b) *Guide to Moody's Ratings, Rating Process, and Rating Practice*; Moody's Investors Service: New York, 1997; (c) Prunty, J.; Jacob, K. Top 10 ways to improve or maintain a municipal credit rating. *Standard & Poor's Public Finance*, 2002; February 4.
10. Bird, R.M.; Ebel, R.D.; Wallich, C., Eds.; *Decentralization of the Socialist State: Intergovernmental Finance in Transition Economies*; World Bank: Washington, DC, 1995.
11. Japan: highways to hell. *The Economist* **2003**, August 30, 29.
12. Flood barriers: saving Venice. *The Economist* **2003**, September 27, 81.
13. (a) Ammons, D.N. Chapter 13: The time value of money: opportunity costs, discounting, compounding, future value, and present value; Chapter 22: Making choices systematically—from audit targets to capital projects. In *Tools for Decision-Making: A Practical Guide for Local Government*; CQ Press: Washington, DC, 2002; (b) Guess, G.M. *Public Investment Programming and Budgeting in Western Countries: Comparison with Kyrgyzstan*. Unpublished Report. Asian Development Bank: Jalal-Abad, Kyrgyzstan, 1997.
14. (a) Lyden, F.; Lindenberg, M. *Public Budgeting in Theory and Practice*; Longman: New York, 1980; (b) Calia, R. *Priority-Setting Models for Public Budgeting*; GFOA: Chicago, 2001.
15. Lyden, F.; Lindenberg, M. *Public Budgeting in Theory and Practice*; Longman: New York, 1980; Vol. 192.
16. Guess, G.M. Role conflict in capital project implementation: the case of Dade County Metrorail. *Public Admin. Rev.* **1985**, *45* (5), 576–585.
17. Vogt, A.J. *Capital Budgeting and Finance: A Guide for Local Governments*; International City/County Management Association: Washington, DC, 2004; Chapter 4.
18. (a) City of Milwaukee. *2003 Plan and Budget Summary*; Milwaukee, WI, 2003; (b) Snyder, T.P.; Stegman, M.A.

Financing the Public Costs of Growth: Using Development Fees, Exactions, and Special Districts to Finance Infrastructure; Department of City and Regional Planning and the School of Business Administration, University North Carolina, for the U.S. Department of Housing and Urban Development: Chapel Hill, NC, 1985.

19. (a) Petersen, J. *Creative Capital Financing for State and Local Governments*; GFOA: Chicago, 1983; (b) *A Guide to Municipal Leasing*; Vogt, A.J., Cole, L.A., Eds.; GFOA: Chicago, 1983.
20. (a) Lawhon, L.L. Development impact fee use by local governments. In *The Municipal Year Book 200*; ICMA: Washington, DC, 2003 (Chapter A3); (b) Snyder, T. *Paying for Growth: Using Development Fees to Finance Infrastructure*; Urban Land Institute: Washington, DC, 1986.
21. Primaria Municipiului Medias. In *Programul Multianual de Investitii: 2005–2009*; Medias: Romania, 2004.
22. Axelrod, D. *Modern Government Budgeting*, 2nd Ed.; St. Martin's: New York, 1995.
23. Schiavo-Campo, S.; Tommasi, D. *Managing Government Expenditure*; Asian Development Bank: Manila, 1999.

Capital Purchases

Wendell C. Lawther

Department of Public Administration, University of Central Florida, Orlando, Florida, U.S.A.

John Adler

*Department of Transportation, Arizona Department of Administration,
Phoenix, Arizona, U.S.A.*

INTRODUCTION

Capital equipment is usually defined as any equipment with a useful life of at least three years.^[1] In the public sector, it is usually funded by the capital budget. In many ways, capital purchases differ from purchase of services and other relationships defined by contracts between public sector agencies and private vendors.

These differences often produce controversial issues that the purchasing staff must face in buying this equipment, often coming into conflict with the agency that requests the equipment. First, there is conflict between those who wish equipment with the latest technological advances and those who argue in favor of the equipment with the lowest price. Although this equipment may be more expensive and the use of the new technology cannot be fully justified by the using agency, the attractiveness of owning the fastest personal computer in anticipation of future uses is very appealing.

Second, ignoring the issue of productivity may result in purchasing inappropriate equipment. Part of the analysis in deciding what equipment to purchase depends upon how the purchased equipment impacts productivity. To achieve this information, outputs of the equipment must be identified and analyzed. The number of copies per minute for a copy machine, as well as the number of copies made by a using agency should both be identified. In this manner, a unit cost—cost per unit of output for a given time period—is calculated. The purchase of more expensive equipment may not be justified if the unit cost is much higher than that of less expensive equipment. This may be the case if the number of outputs and the resulting productivity of using the equipment is low.

Third, life cycle costing (LCC) assesses the total cost of owning, operating, maintaining, and disposing of the equipment. As early in the acquisition process as possible, LCC should be performed. The lowest initial cost equipment model may not be the wisest purchase because repair and maintenance costs may be higher than a more expensive model.

The increased technological advances for equipment such as personal computers and customer service systems based on on-line access have significantly altered

the capital purchasing process. As writing specifications for such equipment may be difficult, as in-house purchasing staff may not have sufficient knowledge to do so, obtaining the equipment relies more heavily on processes such as Invitation to Negotiate (ITN) in which the purchasing process is very different from the acceptance of the lowest cost item.

THE CAPITAL PURCHASING CYCLE

An overview of the procedures and methods used in purchasing capital equipment will help to illustrate the unique aspects of capital purchasing. Because the process repeats itself as old equipment is replaced, it is most accurately viewed as cyclical in nature.

Need Determination

Various factors contribute to determining the need for capital equipment. Changing service or workload may require additional equipment to maintain acceptable levels of productivity. Advances in technology may make the currently operating equipment obsolete because it cannot maintain new standards of responsiveness.

Also, accompanying a replacement policy and set of procedures are decision rules chosen by an agency that identifies the need for new equipment to be purchased (or leased). For example, when the cost of repair reaches 60% of the value of the equipment, the agency may determine that replacement is necessary. Police vehicles may be replaced after 80,000 miles of usage.

Planning and Budgeting

Depending on the requirements of the budgeting process of the government, capital equipment needs must be identified one or two years into the future. In many cases these needs must be prioritized, justification provided by the using agency, and funds identified to support the purchase. The impact of the capital purchase on the operating budget is usually part of the projections that need to be made.

Specification Development

A specification is a description of the physical and/or functional characteristics of the equipment to be purchased. It may include the requirements to be met by the equipment and the procedures used to assess whether these requirements have been met.^[2] There are many aspects to developing specifications that provide challenges to the procurement official. First, the specifications should not be restrictive so that competition among potential suppliers is not dampened. Second, the use of brand names may be too restrictive and should be avoided unless it can be justified by the using agency. The purchasing official must guard against the possibility that the agency has been persuaded by sales personnel that their brand is superior.

The issue of standard specifications vs. customized specifications must be resolved in many instances before purchasing can proceed. The using agency may request more customized equipment, stating that “it’s our money.” Unless a specific need can be identified, it is the role of the purchasing official to explore with agency personnel using standardized specifications as much as possible. The more standardization that is possible, the more “off-the-shelf” software and hardware, for example, can be purchased, usually at a lower cost to the government.

In some CE areas, such as the purchase of information technology (IT) and related telecommunications products, the fast-changing technology may make it difficult to identify and write accurate descriptions of what hardware and software is needed to meet a customer service need. In the field of advanced traveler information systems, for example, governments have had to state the goals of the system they wish to acquire without detailed specifications identified. It is expected that with this ITN or similar purchasing process, that specifications will be identified as a result of a negotiation among purchasing and agency personnel and potential suppliers.

Sourcing

This part of the process involves identifying and maintaining a list of those vendors or suppliers that produce or can obtain the required equipment and would respond to an invitation to bid (ITB), request for quotation (RFQ), or request for proposal (RFP) provided by the government. To maintain competition, the list should be updated as much as possible by adding new vendors. Purchasing staff can obtain these from agency personnel who are more likely to be in contact with relevant vendors. Also, those suppliers who have failed to respond to an ITB or who have provided

unsatisfactory performance should be removed from the list.^[3]

In many cases bidders are required to pre-qualify. This process not only is more efficient for the purchasing department, as vendor qualifications do not have to be reviewed after every bid award. It also facilitates the bid review and award process by assuring that all vendors considered have the capability to provide the equipment as required.

Source Selection

The process of selecting which equipment to buy is different for capital equipment compared to purchasing services or other supplies. No matter what source selection process is chosen, there is likely to be more negotiation with capital purchases. Other items are consumed relatively quickly after purchase, while capital equipment may be with the agency for years. Even if standard specifications can be chosen for the equipment, and an ITB, RFQ, or RFP is used, there may be the expectation by the supplier that a counteroffer from the government or agency will be forthcoming, and that negotiation will resolve differences.^[1]

If the equipment is customized, or complex enough so that specifications cannot be written in great detail, or part of a larger system, e.g., information technology, a RFP or invitation to negotiate may be the source selection process chosen. These are multistep processes that may involve extensive negotiation with more than one vendor. Usually, there is a proposal review process that involves creating a bid rating scale, establishing a team of raters from purchasing, the using agency, and from other governments or agencies, and prioritizing the bids received.

At the highest degree of complexity, when the agency goal or problem is clear/well-defined but the equipment and the means to accomplish that goal are unclear, the agency may use the source selection process to assist it in choosing the means. The Commonwealth of Kentucky, for example, has established a pre-qualification process for obtaining agreements with information technology. As of fall 2000, 15 IT vendors were pre-qualified: 5 for “full-service” while the other 10 fill specific niches. Once private vendors are qualified, then any state agency sends a letter to all vendors outlining the problem that needs to be solved, and inviting vendor to propose solutions. If there is an interest, negotiations begin. The final result is a fixed price contract, with specific deliverables identified.

Contract Administration

This function ideally should be performed by a team of purchasing and agency personnel. The activities

involved are similar to the administration of all contracts, including 1) monitoring the delivery of equipment to ensure that it is provided in a timely fashion according to the dates specified in the contract; 2) resolving problems/complaints if deadlines are not met and/or equipment does not meet specifications; 3) assessing the capability of the supplier to provide the equipment as required; and 4) assessing penalties or other sanctions including termination of the contract.

In the case of capital equipment, however, there must be a process testing the delivered equipment to ensure that it produces the required output. In addition, administering the warranty, including knowing when and under what conditions the warranty applies, is part of contract administration.

Maintenance and Service

The primary decision rules adopted by the agency pertain to the use of routine or preventative maintenance (PM). Without such maintenance, the equipment is repaired when it fails to operate. This is known as failure maintenance (FM).

The obvious benefit of PM is that frequent usage will extend the useful life of the equipment. The assumption here is that the more funds invested in PM, the lower the costs of FM. Failures will be less frequent, and resulting downtime less costly as well. Preventative maintenance costs include the cost of materials, the time/salary of those performing the repairs, the cost of spare parts that may be installed, and the downtime of the equipment while being maintained.

The issue of downtime becomes important in determining the usefulness of PM. In many cases, the cost of downtime is negligible. Alternative equipment may be available. If a personal computer needs PM for one day, email may be accessed via another computer. A laptop computer may substitute for the PC. If a vehicle is “in the shop” for PM, work can be rescheduled or another vehicle may be available. If downtime incurs realistic costs, however, and requires renting replacement equipment, then these costs must be factored into the need for PM.

An opposing view is that PM is not needed. When the equipment fails, it should be repaired at that time. This approach can be adopted under certain conditions. If the assumption expressed above about PM is not accurate: increased PM does not lessen or has little impact on the cost of FM, then PM may be of little value. Preventative maintenance for a personal computer may have little bearing on whether a hard drive crashes.

A second condition is relevant to the value of the equipment when FM is likely to be necessary. If the value depreciates quickly, so that when the equipment breaks down the cost of repair is a high percentage of its value, then replacement instead of repair is a more feasible option.

Replacement Planning

As indicated above, many agencies have a replacement plan or policy in place that will start the purchasing cycle. Decision rules exist that are often implemented based on industry standards and/or past experience. These will trigger replacement of existing equipment. Agency or government committees may be established to monitor and oversee vehicle and equipment replacement.

Disposal

In many respects the capital purchasing cycle ends and begins again with the disposal of the equipment and the planning for its replacement. There are several disposal options available, including 1) trade-in for new equipment; 2) sell to a used-equipment buyer; 3) used for spare parts; and 4) pay to have the equipment removed.

The option chosen depends upon the availability of a market for the used equipment. In some cases, if upgrades have been added to the equipment over its useful life, it may be worth more when disposed than its initial cost.

CONCLUSION

With the need to purchase computer hardware and software, especially in the context of creating information technology systems, the importance of capital purchasing has risen dramatically. The stakes are higher, as the risk of failure increases with the complexity of the equipment purchased. Countless numbers of examples are similar to the following instance.

Florida state legislators plan to audit a child welfare computer system that is seven years behind schedule and more than \$200 M over budget. The system, designed to replace paper files and several separate computer systems that track abused and neglected children in Florida, was first approved by the legislature in 1990. It was originally supposed to be completed in 1998 at a cost of \$32 M. It is now scheduled to be completed in 2005 and cost nearly \$232 M.^[4]

The greater need for a partnership approach to capital purchases is evident from this example.^[5] Public sector purchasing and agency personnel need to partner

with private vendors so that the requisite knowledge needed to accomplish the goals is shared by all. Problems need to be identified early in the implementation process, and solutions sought with the input of all concerned to avoid waste, delays, and cost overruns.

REFERENCES

1. Newman, R.G.; Simkins, R.J. *Capital Equipment Buying Handbook*; American Management Association: New York, 1998.
2. The National Institute of Governmental Purchasing, Inc. In *Dictionary of Purchasing Terms*; Herndon, VA, 1996.
3. Council of State Governments. In *State and Local Governmental Purchasing*; Lexington, KY, 1988.
4. Kestin, S. "Legislators to Conduct Audit as Computer System Goes \$200 Million Over Budget" *South Florida Sun-Sentinel.Com News*; January 7, 2002. <http://www.sun-sentinel.com/news/local/florida/sflcomputerjan07.story>.
5. Lawther, W.C. *Contracting for the 21st Century: A Partnership Model*; The Pricewaterhouse Coopers Endowment Business for Government: Arlington, VA, 2002.

Career Development

Teva J. Scheer

University of Colorado at Denver, Highlands Ranch, Colorado, U.S.A.

INTRODUCTION

Career development is the process of setting developmental goals and of identifying training and other developmental opportunities that will achieve those goals. An organization's human resources programs should include a career development system that helps individual employees to meet their own development goals within the context of the organization's goals and training needs. This section explains the components of an organizational career development system, lists the individuals or organizations with responsibility for the career development system, and discusses a variety of career development tools.

WHAT IS CAREER DEVELOPMENT?

In its simplest terms, career development is the process of setting developmental goals and of identifying training and other developmental opportunities that will achieve those goals. Any person can complete these activities individually. However, within the context of a public organization, a career development system exists to meet the knowledge and skill needs of the entire agency. The individual's development needs are still met, but within the larger context of the organization's goals and objectives.

To be effective, an organization's career development system must be integrated with the other components of the human resources (HR) program. The agency's recruitment and selection system is a major source of job analysis information which should be fed into the identification of training needs, both to provide entry-level knowledge and skills and to prepare current employees for higher-level opportunities. The agency's position management system can provide detailed information about occupations, which can be used to develop standard career paths and succession plans. The agency's performance management system can provide information about mastery levels in all occupations; this information should be used to identify gaps between current and optimum levels of performance. And the agency's long-range planning processes can provide information about future skill needs and the growth or erosion of occupations and/or functions.

Career development is a balancing act between two sets of values: those of the organization and those of the individual employees. Since training is an organizational investment, the agency's values "trump" those of the individual employees. For example, if a state park sets an annual goal to insure that all rangers will become efficient in fire suppression techniques, then an individual ranger's personal goal of studying wildlife management will likely need to be deferred. However, the closer the agency's developmental goals are to those of its employees, the more likely that the career development efforts will be effective.

CAREER DEVELOPMENT RESPONSIBILITIES

The responsibility for career development is divided between three groups: the HR organization on behalf of the agency; the managers; and the employees.

The HR organization is responsible for making sure that managers and employees have the information they need about the agency's long-range plans, e.g., staffing projections, performance goals, planned initiatives, and any expected changes in agency mission. Ideally, HR will house a career center, where employees can access career planning literature, obtain personal career counseling, and complete instruments such as skills assessments. Human resources also may serve as the clearinghouse for developmental opportunities available throughout the organization. Most HR organizations also host periodic events for employees, such as internal career fairs and presentations on application procedures, individual development plans, and effective interviewing. Finally, HR is responsible for assisting managers in carrying out their development activities.

Managers are responsible for assisting their employees in the career planning process. They need to communicate regularly with employees about their performance, their developmental needs in their current jobs, and their long-range planning for future jobs. They need to work with their employees to develop individual development plans. They are responsible for helping employees to identify appropriate training classes, for securing the necessary funding, and for relieving employees of their workload during the training period. They are responsible for devising

assignments on the job that will build new skills or enhance current ones, and they are also responsible for finding details outside the current organization in accordance with the employees' individual development plans. Ideally, they will become their employees' career coach and mentor, or they will arrange for another mentor as appropriate (i.e., if an employee's career objective is located elsewhere in the organization).

Employees are responsible for clarifying their own career goals and for helping their managers to identify their current and future training needs. Employees need the cooperation of HR and of their managers to meet their developmental goals, but they need to take primary responsibility for their own careers. If a manager lacks the time or the temperament to be an involved mentor, then employees need to take the initiative to ask for performance feedback, to identify appropriate training opportunities, and to discuss their development plans with the manager.

CAREER DEVELOPMENT TOOLS

Agencies have a variety of tools available to them to help plan and administer their career development programs.^[1]

1. To assist in the organizational planning and goal-setting that should precede individual development efforts, agencies can use work plans, agency planning documents, budgets, auditors' reports, and feedback from outside the organization (e.g., city manager directives, Congressional hearings). If the agency has not conducted an internal skills inventory, either a contractor or the in-house position management staff can conduct a study of existing vs. future skill needs.
2. Once the long-range development plans have been developed, the focus turns to individual development. Available self-assessment tools include career planning workshops, assessment centers, and the use of instruments such as the TriMetrixT System Personal Talent Plus Report, the Hartman Kinsel Profile Report, or the Myers-Briggs Type Indicator. Instruments and

questionnaires can help employees to identify their style preferences, career priorities, performance strengths, and training needs. Many books on career planning contain simple questionnaires or sample developmental plans that managers and employees can adapt to meet their needs.

3. The employee may receive individual counseling from the manager, an assigned mentor, the HR staff, or a professional career counselor, internal or external to the organization. Specialized referrals can include professionals in outplacement and employee assistance (e.g., financial counseling, family issues, or substance abuse, which affect on-the-job performance).

THE CHALLENGE FOR TOMORROW'S ORGANIZATION

Many 21st century organizations are becoming flatter and lowering their spans of control. They are also more likely to contract out for special skills or projects in order to increase their flexibility and reduce their staff costs. Employees will need to keep in mind that career development is no longer limited to advancement to higher-graded positions in the current organization. They will increasingly need to focus on lateral placement or movement outside the current organization. The challenge for organizations with reduced opportunities for advancement will be to show their valued employees how lateral moves can benefit their careers.

REFERENCES

1. Gutteridge, T.; Hutcheson, P.C. Career development. In *The Handbook of Human Resource Development*, 2nd Ed.; Nadler, L., Nadler, Z., Eds.; John Wiley & Sons: New York, 1990; 30–12.

BIBLIOGRAPHY

1. Leibowitz, Z.B.; Farren, C.; Kaye, B.L. *Designing Career Development Systems*; Jossey-Bass Publishers: San Francisco, 1986.

Cash Flow Forecasting: Principles

William R. Voorhees

*Department of Political Science and Public Administration, University of North Florida,
Jacksonville, Florida, U.S.A.*

Jeongwoo Kim

School of Public Affairs, Arizona State University, Phoenix, Arizona, U.S.A.

INTRODUCTION

Management of cash flow is extremely important to the fiscal efficiency of governments. This importance is underscored by the fact that most general-purpose governments in the United States generate their budgets on a cash basis, even though they are accounted for on a modified accrual basis.

At least three different terms are used to describe this process, “cash budgeting,” “cash flow budgeting,” and “cash flow forecasting.” Each of these terms may be used with slightly differing connotations. In the finance field, cash budgeting is considered a part of the master budget, which includes the operational budget and the capital budget; however, in the political science field, cash budgeting is the process of making the cash appropriation for the yearly budget.^[1,2] In the former case, service-level assumptions are assumed and the cash flow is calculated. In the latter case, the process starts with the appropriation and the service levels are a function of the appropriation. This article will assume that policy drives appropriation and that service levels are set requiring the cash flow for operating at a given service level to be calculated. To avoid confusion, the term cash flow forecasting will be utilized throughout this article.

THEORY OF MANAGING CASH FLOWS

Good financial management seeks to optimize the cash flow of receipts and minimize the cash flow of disbursements, thus maximizing surpluses or minimizing deficits. When the timing of cash receipts and disbursements is considered, it is easy to see how budgets for identical gross amounts may not yield the same results in resource utilization. This is due to the time value of money. Budgets that have accelerated disbursements and decelerated receipts will be less efficient than budgets with decelerated disbursements and accelerated receipts. The cash budget is the tool of choice for managing the timing of receipts and disbursements.

Consider the following two cash budgets, both with appropriations and receipts in the amount of \$10 million (see Table 1). In the first case, an inefficient budget is presented, in which all \$10 million of the receipts are collected in the last month of the budget and all of the expenditures are disbursed in the first month of the budget year (12mo earlier). In this example, the community must borrow funds in order to pay the disbursement in the first month, because no current revenues are available. Utilizing a 6% discount rate and converting the budgeted amounts to future values [Eq. (1)] for comparison, it is discovered that the jurisdiction of the first budget actually had a deficit when the amounts are adjusted for the time value of the money.

$$FV = \frac{\$10,000}{\left(1 + \frac{0.06}{12}\right)^{12}} = \$10,617 \quad (1)$$

The second budget, on the other hand, collected all of its revenues during the first month of the budget year and disbursed all of its expenditures in the last month of the budget year, resulting in a surplus of \$617,000. By investing the revenues at the market rate, the second jurisdiction was able to generate an additional \$617,000 in revenue, resulting in a surplus.

Even though both of the budgets had appropriations of \$10 million dollars, the first community received considerably less in value from their appropriations than the second community. Naturally, these are exaggerated examples, but the basic principle of maximizing the time before disbursement of expenditures and minimizing the time in collection of revenues is fundamental for every financial manager. The concept of efficient cash management is, of course, taken from the perspective of the government and not individual citizens or vendors.

In reality, one would expect revenue receipts and expenditure disbursements to occur periodically throughout the budgetary year. For example, the majority income-based taxes are usually collected in

Table 1 Cash budget efficiency and inefficiency

Numbers in thousands	First month of budget (\$)	Last month of budget (\$)	Time value-adjusted budget (\$)
<i>Community 1—inefficient</i>			
Receipts		10,000	10,000
Disbursements	10,000		10,617
Deficit			−617
<i>Community 2—efficient</i>			
Receipts	10,000		10,617
Disbursements		10,000	10,000
Surplus			617

the first part of April. Depending on the jurisdiction, property taxes may be due in a lump sum or in prorated amount over the year. Because the collection of these taxes are not synchronized to the disbursement of the funds for goods and services these taxes are designed to support, additional costs or revenues may be realized. The ability to identify these costs or revenues on an interim basis allows managers to make corrections to fiscal policy decision, in order to optimize resources.

In the case of departments that provide a significant part of their services on a fee for service basis, such as water and waste departments, the cash flow forecast allows these departments to minimize the lag time between service provisioning and collection of the fees for the service. In order to maximize fiscal resources, some departments have even required citizens to pay in advance for services rendered.

An additional benefit of the cash flow forecast accrues when laws require that budgetary allotments be utilized. An allotment is the periodic release of funds to a department for expenditures. For example, if a department has a \$12,000 budget, they may only be allowed to spend \$1000 per month, ensuring that there will be funds for the entire year. Good cash flow forecasting provides a valuable tool for managers to determine if they need special exemptions made for even allotments.

MECHANICS OF CASH FLOW FORECASTING

Cash flow forecasts are based on the likely pattern of receipts and disbursements over a period of time and are used to determine the daily, weekly, monthly, or yearly cash position of the government. Following generally accepted accounting principles for reporting of cash flow statements, this article presents the cash flow forecast in the same format as is found on a governmental financial statement.^[3] This cash flow forecast

is typically composed of six elements: cash flows from operating activities; cash flows from non-capital financing activities; cash flows from capital and related financing activities; cash flows from investing activities; net monthly cash change; and new financing needs.^[4]

Cash Flows from Operating Activities

The formulation of the cash flow forecast is a multifaceted process. This process begins with analyzing the existing patterns of cash inflows (receipts) and outflows (disbursements) from operating activities, based on accounting records. For receipts, revenue sources, amounts, and payment dates are examined to determine when the posted revenue is converted from an accounts receivable into a cash asset. For disbursements, billing amounts, billing dates, and payment dates are likewise examined to determine the conversion of an accounts payable into an actual cash outflow. The historical behavior of the receipts and disbursements is then utilized as the foundation for forecasting future cash in- and outflows.

To demonstrate the construction of the cash flow forecast, consider a water service department (see Table 2). The cash flow forecast of the water service department is begun by estimating the charges for services for each month. The department's revenues are mainly derived from fees and service charges assessed to residential and commercial customers of the city's water and wastewater systems. Based on past experience, the department expects to collect 50% of all charges in the month they occur, 35% within 1 month, and 15% within 2 months. Once the cash collections have been calculated from the forecasted revenue for each month, the total receipts can be calculated. To calculate collections for September, for instance, the department would determine 50% of its current charges ($\$2,115,900 \times 50\% = \$1,057,950$) and add 35% of August charges ($\$3,526,500 \times 35\% = \$1,234,275$) and 15% of July charges ($\$3,056,300 \times 15\% = \$458,445$) for the total collections of \$2,750,670.

The second component of the cash flow forecast is the anticipated cash disbursements. A detailed understanding of the payment schedule for expenditures is of utmost importance in calculating disbursements. For example, supplies may be purchased on accounts payable at the beginning of each month but not paid until 30 days later, in accordance with trade credit terms. Of course, a good cash manager will attempt to delay disbursements as long as possible, by utilizing the optimal terms of available trade credit.

In Table 2, total disbursements for September are \$1,812,000. The total operating activities cash flow for September is calculated by subtracting

Table 2 Example cash flow forecast

Quarterly cash budget for water service department								
	May	June	July	August	September	Total budget	Actual	Variance
<i>Cash flow from operating activity</i>								
Charges for services (\$)	2,300,000	2,950,000	3,056,300	3,526,500	2,115,900	8,698,700	23,505,000	14,806,300
Receipts								
Current month collections	50%		1,528,150	1,763,250	1,057,950	4,349,350	11,754,500	7,405,150
Previous month collections	35%		1,032,500	1,069,705	1,234,275	3,336,480	8,273,000	4,936,520
Second previous month collections	15%		345,000	442,500	458,445	1,245,945	3,573,000	
Total receipts			2,905,650	3,275,455	2,750,670	8,931,775	23,600,500	14,668,725
Disbursements								
Personal services			677,000	717,000	597,000	1,991,000	7,003,000	5,012,000
Materials and Supplies			325,000	375,000	225,000	925,000	2,499,500	1,574,500
Fees and services			1,170,000	1,350,000	810,000	3,330,000	8,999,000	5,669,000
Other			260,000	300,000	180,000	740,000	1,979,000	1,239,000
Total disbursements			2,432,000	2,742,000	1,812,000	6,986,000	20,480,500	13,494,500
Total operating activities cash flow			473,650	533,455	938,670	1,945,775	3,120,000	1,174,225
<i>Cash flows from non-capital financing</i>								
Non-operating revenue			162,470	170,850	145,710	479,030	1,713,000	1,233,970
Transfers in			50,000	0	0	50,000	100,000	50,000
Transfers out			0	540,000	0	540,000	1,080,000	540,000
Total noncapital financing cash flow			212,470	(369,150)	145,710	(10,970)	733,000	743,970
<i>Cash flows from capital and related financing</i>								
Interest bond interest—\$65,000,000 at a rate of 6%			0	0	1,950,000	1,950,000	3,900,000	1,950,000
Total capital financing cash flow			0	0	1,950,000	1,950,000	3,900,000	1,950,000
<i>Cash flows from investing</i>								
Interest on Investments—\$1,500,000 at a rate of 3%			3,750	3,750	3,750	11,250	45,000	33,750
Total cash flow from investing			3,750	3,750	3,750	11,250	45,000	33,750
<i>Net monthly change</i>								
Plus: beginning cash balance			100,000	789,970	958,815	100,000	100,000	0
Total cash flow from operating activities			473,650	533,455	938,670	1,945,775	3,120,000	1,174,225
Total cash flow from noncapital financing			212,470	(369,150)	145,710	(10,970)	733,000	743,970
Total cash flow from capital financing			0	0	1,950,000	1,950,000	3,900,000	1,950,000
Total cash flow from investing			3,750	3,750	3,750	11,250	45,000	33,750

(Continued)

Table 2 Example cash flow forecast (*Continued*)

		Quarterly cash budget for water service department							
		May	June	July	August	September	Total budget	Actual	Variance
Plus: total monthly cash flow balance				689,870	168,055	(861,870)	(3,945)	(2,000)	1,945
Less: interest on short-term notes—at a rate of 2.4%				0	0	0	0	3,400	3,400
Plus: interest on excess cash (previous month balance)—at a rate of 1.2%				100	790	959	1,849	8,900	7,051
Equals: net monthly change				789,970	958,815	97,904	—	—	
Cash flow financing needs (measure of efficiency)			0	0	2,096	2,096	740,000	737,904	
Cash flow note repayment				0	0	0	0	740,000	740,000
Ending cash balance			100,000	789,970	958,815	100,000	97,904	103,500	5,596
Cumulative cash flow borrowing				0	0	2,096	0	0	0
Minimum cash balance		\$100,000							
Interest rate									
Bond proceeds		6%							
Investment		3%							
Short-term notes		2.4%							
Excess cash		1.2%							

total disbursements from total receipts (\$2,750,670 – \$1,812,000 = \$938,670).

The management of cash flow is the practice of accelerating cash receipts and decelerating disbursements. To accelerate the receipt of revenues, the government can utilize techniques such as earlier billing, prompt payment discounts, and late payment penalties to improve revenue flows. Collection acceleration can often be improved by reducing the time between a payment, typically in check form, and the point in which the payment is converted to cash and the government has use of the funds. This is called collection float. Lockbox systems, electronic fund transfers, and accepting credit cards for payment are all proven techniques for reducing collection float.^[5]

Like acceleration of collections, the government can minimize disbursements to optimize cash flows. A variety of ways have been suggested for disbursement management.^[6,7] First, pooling cash balances into a central concentration account improves management by consolidating funds at a single point. Centralizing disbursement also permits more efficient monitoring of payables and float balances. Second, zero-balance accounts (ZBA) are special disbursement accounts having a zero-dollar balance for the

government's expenditures. ZBA simplify the control of disbursements and cash balances and hence reduce the amount of idle cash. Third, expenditure controls can be used to track and brake unnecessary early payments. Finally, employee payment periods should be standard and as long as feasible. Using periodic payment dates can reduce processing cost and increase the predictability of payment amounts and the time availability of money.

Cash Flows from Non-Capital Financing Activities

Cash flows from non-capital financing activities are tracked separately from operating activities. These activities include expected receipts from grants or subsidies other than those specifically restricted for capital purposes or those for specific operating activities and grants or subsidies expected to pay to other governments, funds, or organization, except payments for specific operating activities of the grantor government. To control disbursements, discretionary payments such as grant payments or shared revenues should be delayed as much as reasonable.

Cash Flows from Capital and Related Financing Activities

Cash flows from capital and related financing activities include proceeds of debt attributable to acquisition of capital assets, receipts from the sale of capital assets, and payments to acquire capital assets. The payments of long-term obligations follow a more predictable pattern. In the example, total financed amount is \$65,000,000 at the rate of 6%, and bond proceeds are scheduled to pay twice a fiscal year. First payment of the interests ($\$65,000,000 \times 6\% \times \frac{1}{2} = \$1,950,000$) is disbursed in September. The variance reflects the other amount of bond proceeds that should be paid in the fiscal year.

Cash Flows from Investing

The estimation of cash flow provides a pool for investment of available excess cash. The excess cash can be placed in a sweep account that earns interest on a daily basis but has no other use for the funds. On the other hand, a more active investment would involve investments in securities with specific maturities, to smooth out cash flows. Beyond the special limits placed by federal, state, and local laws, managers operate investment within the constraints of sound financial management, including no default risk, no credit market risk, and no illiquid assets. The water service department made an investment of \$1,500,000 at 3% rate and earns interest of \$3750 a month.

Net Monthly Cash Change

Once the anticipated cash flows have been determined for the fiscal year, net monthly cash change is calculated by adding total monthly cash flow balance and interest on excess cash to the beginning cash balance and subtracting interest on short-term borrowing. In September, the department's beginning cash balance is \$958,815 and interest on the balance is \$959. The department has negative total cash flow balance ($-\$861,870$). Thus, net cash change is \$97,904 ($\$958,815 + \$959 - \$861,870 = \$97,904$) for the month.

New Financing Needs

The cash budget indicates whether short-term financing will be necessary, in what month, and in what amount. Most organizations set minimum cash balance and will borrow funds to maintain that balance if necessary. Repayment of borrowing is also possible in any month with anticipated excess cash. When there

needs a financing, the minimum cash balance is the ending cash balance for the month. Otherwise, ending cash balance is calculated by subtracting the amount of repayment from the net monthly cash change. Finally, the cumulative amount of borrowing is tracked on the last line.

In the example, the department's net cash change in September is \$97,904, which is less than the minimum balance of \$100,000. It indicates that the department needs to borrow the difference ($\$100,000 - \$97,904 = \$2096$). Ending cash balance for the month is \$100,000, and the amount of cumulative borrowing is \$2096.

FINE-TUNING CASH FLOW FORECASTS

Structural components of the cash flow forecast will influence its accuracy. For example, forecasts with longer horizons will generally be less accurate than budgets with shorter horizons. Secondly, when forecasts are made more frequently, they have the advantage of incorporating more recent information. The forecast period will also influence the accuracy of the cash flow budget, with quarterly budgets being more accurate than daily budgets. On a daily basis, it may be difficult to determine if a payment or receipt will be made that day, but, on a monthly basis, there is a greater confidence that a specific payment will be made within a given month. Additionally, if a single payment or receipt should fall outside of the expected payment period, it will have less impact in a monthly forecast than in a daily forecast. However, there exists a trade-off between usefulness and accuracy.

Because of the substantial variation in balances during a quarter, quarterly budgets offer less aid to managers attempting to maximize the use of excess cash and minimizing the use of credit. From a cash management perspective, a monthly, weekly, or even daily cash budget provides better information for maximizing invested cash and minimizing borrowing than a longer forecast period. The length of the forecast period is a function of the needs of the government and the cost of producing the forecast. A daily forecast period requires more resources and may not be appropriate for a small government that has minimal cash to invest. Instead, a smaller government may opt for the less resource-intensive monthly forecast period. On the other hand, a government with daily balances of hundreds of millions of dollars would likely find that a daily forecasting system is necessary to optimize their cash investments. In designing a cash forecasting system, the size of the government is a critical factor in determining forecast needs; however, the size of the government will also influence the resources available to generate a cash flow forecast.^[8] However, it is safe

to say that regardless of size, all well-run governments generate a cash flow forecast of some form.

REFERENCES

1. Petty, W.; Keown, A.; Scott, D. *Basic Financial Management*; Prentice-Hall: Englewood Cliffs, NJ, 1988; 93 pp.
2. Schick, A. Micro-budgetary adaptations to fiscal stress in industrialized democracies. *Public Admin. Rev.* **1988**, *48* (1), 523–533.
3. Kravchuk, R.; Voorhees, W. The new governmental finance reporting model under GASB Statement 34: an emphasis on accountability. *Public Budg. Finance* **2001**, *21* (3), 1–30.
4. Governmental Accounting Standards Board. *Codification of Governmental Accounting and Financial Reporting Standards*; Governmental Accounting Standards Board: Norwalk, CT, 2004; 329–345.
5. Reed, B; Swain, J. *Public Finance Administration*; Sage Publications: Thousand Oaks, CA, 1997; 207–209.
6. Lynch, T. *Public Budgeting in America*; Prentice-Hall: Englewood Cliffs, NJ, 1995; 236–237.
7. Mikesell, J. *Fiscal Administration: Analysis and Applications for the Public Sector*; Wadsworth/Thompson Learning: Belmont, 2003; 592–595.
8. Coe, C. The effects of cash management assistance by states to local governments. *Public Budg. Finance* **1988**, *16* (2), 80–90.

Cash Management for State and Local Governments

Dwight V. Denison

*Robert F. Wagner Graduate School of Public Service, New York University,
New York, New York, U.S.A.*

INTRODUCTION

Cash is the lifeblood of a public organization. Without cash, it becomes difficult, if not impossible, to provide goods and services. Governments without adequate cash cannot buy garbage trucks, maintain bridges, pay teachers' salaries, or operate social programs. On the other hand, cash not actively employed in providing government services should be prudently invested and earning interest revenues.

The day-to-day operations of governments generate and consume cash. Cash receipts arise from tax, fee, grant, and interest collections. Cash disbursements occur with vendor payments, employee salaries, finance charges, and capital expenditures. Cash must be obtained through borrowing during a cash deficit. Excess cash balances provide the opportunity for investment revenue. Comprehensive cash management is comprised of four primary components: cash forecasting, cash mobilization, investment, and bank relations. Each component is vital in fostering good fiscal health.

CASH FORECASTING

Cash management insures that cash inflows are adequate to meet cash outflows. To do so requires an ability to forecast the arrival and amount of cash receipts and disbursements. The timing of cash receipts is classified as certain, forecastable, or unpredictable cash flows. Certain flows are known in advance of when they occur. Examples of certain cash expenditures include contract salaries and interest payments. Forecastable cash flows can be estimated with reasonable accuracy using statistical methods and professional judgment. Examples of forecastable cash flow include sales tax collections and pension fund disbursements. Unpredictable cash flows arise from unanticipated events such as a lawsuit settlement or severe weather damage.

Governments usually forecast the operating cash deficit or surplus on at least a monthly basis through a cash budget. The cash budget tracks cash receipts and disbursements while taking into account the collection cycle of receivables and payment schedules. Monthly cash budgets help to identify financing needs

and to determine investment strategies. Daily cash balances can be monitored on a daily basis to determine opportunities for investing idle cash balances. Adams and Harrison^[1] discuss common techniques for forecasting cash balances.

CASH MOBILIZATION

Cash mobilization is the process of maximizing cash balances by decreasing the time it takes to collect cash owed to the government and by deferring disbursements. Unless efforts are made to mobilize cash, cash balances may be insufficient to meet disbursements. Furthermore, aggressive cash mobilization potentially increases the idle cash balances available for investment.

The first aspect in cash mobilization is to monitor the accounts and taxes receivable to the government. An aging schedule can be used to track delinquent tax payments and help determine appropriate collection efforts. Governments usually promote timely tax payments by imposing charges on delinquent accounts. In cases of non-payment, liens are frequently imposed. As an additional incentive, some governments offer a discount if the payment is received prior to the due date. Governments should regularly monitor the billing and collection processes to find ways to promote timely payments and reduce delinquency.

Larger governments often take steps to reduce collection float. Collection float is the time it takes for the funds to become available in government bank accounts once the payment has been submitted. Collection float is the total time it takes to transmit the payment from the payer to the government (transmission or mail float), plus the time it takes for the government to process the payment (process float), plus the time it takes for the payment to clear in the bank account (clearing float). To illustrate, consider that a citizen remits a property tax payment on June 1. It takes 2 days by mail to reach the revenue department and 1 day for the revenue department to process and deposit the check. Then, the check takes 1 day to clear before the funds become available for the government to spend. The collection float in this case is 4 days. Every day of collection float represents lost interest revenue to the government, for if the cash were available, it could be invested.

While it is impractical to expect that collection float could be completely eliminated, there are several techniques that governments employ to reduce collection float. Payment of taxes is primarily done via check, but some governments have implemented electronic transfers directly from one bank account to another through the Automated Clearing House (ACH). Electronic transfers between accounts reduce the mail and clearing floats. In the prior example, electronic transfer would reduce the collection mail float by at least 3 days. ACH transfers are recommended for transferring large payments between governments and are also convenient for smaller reoccurring payments.^[2]

Lockbox services are provided by banks to expedite the deposit of remitted checks, and if strategically located geographically, will also reduce mail float. In a basic lockbox service, the bank empties a P.O. box, deposits checks to the appropriate account, and forwards the documentation to the government. Many state and local governments significantly reduce collection float through the use of lockboxes.

The second aspect of cash mobilization requires that governments pay current obligations without incurring unintended fees or charges. Generally, obligations are not prepaid unless there are compensating incentives or discounts; otherwise, the cash could remain invested and generating interest. Vendors usually impose late charges if a payment is not received by the due date. Governments should avoid paying late charges unnecessarily but may consider late payment as an alternative to borrowing to meet a temporary cash deficit.

Governments may utilize float to detain cash outflows and increase idle cash balances. Disbursement float is the collection float in reverse. Governments can increase disbursement float by increasing mail float or clearing float. For example, there are generally 2 to 3 days of disbursement float on employees' wage checks before the balance posts to the government's bank account. In practice, excessive actions to extend disbursement float may result in late payments and jeopardize relationships with suppliers. Actions to extend clearing float, such as remote disbursement, are opposed by the Federal Reserve System.

Electronic transfers (ACH) have several benefits for disbursements. Electronic transfers are immediate, meaning the cash can be held until just prior to the due date. Electronic transfers can reduce paperwork and administration and therefore may be more cost effective than printed check disbursements. Moreover, supplier relationships are enhanced, because ACH transfers reduce their collection float.^[2]

The third aspect of cash mobilization is concentrating cash into pools that make it easier to invest. Banks offer a number of services for concentrating cash into pools that preserve the legal restrictions for accounting on the separate funds. Two common

services are the sweep account and the zero balance account. In a sweep account, the bank transfers the balance at the end of the day from each individual account to a central account, where the money can be invested overnight. The next morning, the appropriate balance is returned to the originating account. The zero balance account operates on the reverse principle. Individual account balances are zero, with the combined balances being held in a central account. Once a draft is presented to an individual account, the draft is honored, and the appropriate amount is transferred from the central account to bring the account balance back to zero.

Cash pooling makes forecasting daily cash needs easier, because pooled cash flows become less volatile and are more predictable than the cash flows of each component fund. As a result, concentrating cash balances also benefits investment aspects of cash management.

INVESTMENT

Cash not actively employed in providing government services should be invested and earning interest revenues. Investment revenues are increasingly important to governments with restricted tax revenues. Because higher rates of return are generally compensation for assuming greater risk in the portfolio, the public has been increasingly concerned about the regulation of government investment practices. Public concern in the oversight of government investment practices is warranted, because loss of tax dollars, potential default on debt, or even bankruptcy, can result from irresponsible investment.

The basic investment objectives considered when investing idle cash balances include liquidity, security, and return. Liquidity refers to the ability to transform the investment into cash. Liquidity is an important priority when investments may need to be converted to cash to meet unanticipated cash disbursements. Security refers to the credit and market risk associated with the investment. All investments carry some degree of risk, but certain classifications of investments are more vulnerable to loss. For example, securities issued by the U.S. Treasury have a low risk of default and are therefore called "risk-free," while non-investment-grade corporate bonds have a much higher probability of default. The return or yield from investments is the interest revenue that provides additional cash into government coffers. Investment returns may include direct interest payments or gains that result from securities purchased at a discount. Investment returns also include inflows of dividend payments and capital gains from mutual funds and stocks.

The investment objectives of liquidity, security, and yield typically conflict. In general, investment alternatives with higher risk generate higher yields. Under normal market conditions, higher yields are also associated with investments of longer maturity, as illustrated in the typical yield curve. Thus, aggressive maximization of yields often requires the sacrifice of security and liquidity.

Governments should establish investment goals through a written investment policy.^[3] Each government prioritizes the basic investment objectives based on cash flow characteristics, organizational goals, and various external factors, including legal, regulatory, and economic environments. Money market securities vary with respect to liquidity, security, and yields, and therefore, the securities held in a portfolio are screened to ensure compliance with the investment policy and to make sure investment objectives are met. A listing and description of common money market securities is presented in [Table 1](#).

Permissible Assets

Security of principal is crucial for government organizations, because investment default results in the loss of public funds and invokes political and fiscal repercussions. Consequently, all states constrain to some degree the investment of idle cash to permissible investment assets that meet a minimum standard of security. In a recent survey, all but five of the states reported administrative statutes influencing the permissible assets in which the state treasurer could invest public funds. Furthermore, 14 of the states also had constitutional restrictions in addition to state statutes regulating permissible assets.^[4]

Municipalities are also subject to legal restrictions when investing idle cash. Based on a survey of cities with over 200,000 in population, state statutes are the source of restrictions in 76% of the responding cities, and 6% of the municipalities are subject to state constitutional restrictions on permissible assets. In addition, municipal law dictates permissible assets for 41% of the surveyed cities.^[5]

It is important to emphasize that the basic investment objectives of liquidity, security, and yield are constrained by permissible assets. As presented in [Table 1](#), 94% or more of the states permit the purchase of repurchase agreements and U.S. Treasury and agency obligations. Other investments commonly allowed are certificates of deposits issued by state banks and commercial paper. Repos, U.S. Treasury and agency obligations, and CDs comprise the core of permissible assets. Other assets like banker's acceptances and commercial paper are now more permissible than 15 years ago.^[6]

Arbitrage

Investment of tax-exempt bond proceeds is regulated by the Internal Revenue Service (IRS). The IRS is concerned most with arbitrage of tax-exempt bond issues. In this context, arbitrage occurs when proceeds of tax-exempt bond issues are invested to generate investment revenue from securities bearing higher taxable yields. The difference between the investment revenue and the interest payment on the bond is a cash subsidy that comes at the expense of the federal government. The rules defining arbitrage can be tedious, but generally, arbitrage is not a concern if the proceeds are used in a timely manner to construct buildings or purchase other capital assets.^[7] Minor arbitrage infractions require cash payments to the IRS equal to the subsidy. Egregious arbitrage threatens the tax-exempt status of the bond issue.

Investment Strategy

The investment objectives of security, liquidity, and yield are accomplished through the investment strategy. The investment strategy employed by the portfolio management may vary among the different types of cash balances. The nature of the idle cash balances invested by state governments can be classified into three types: temporary surplus, long-term surplus, and pooled surplus. A temporary surplus results from the lag between revenue collections and expenditure disbursements in the general, special-revenue, capital, and enterprise funds. A temporary cash surplus will be expended within the year, and therefore, liquidity is crucial. Long-term cash balances held by the state for a year or longer are often associated with the trust, pension, or debt service funds. A long-term investment horizon provides more flexibility in the investment strategy to increase yields without compromising security. Pooled surplus results from the pooling of cash balances from separate funds or even multiple governments. The characteristics of pooled surpluses are different in nature than the comprising temporary or long-term funds, because pooling can increase the stability and predictability of the cash flows.

Cash-flow characteristics also have a great deal of influence on the strategies utilized in managing the investment of idle cash balances. Government funds with more certain cash flows may be managed more aggressively to enhance yield without unduly compromising liquidity and security. On the other hand, funds with unpredictable cash flows require a more liquid strategy to ensure the availability of funds on demand.

Investment strategies may also be influenced by the expertise of the state's investment staff and the ability to consolidate cash balances and manage risk.

Table 1 Description and permissibility of select money market securities

Security	Description	Percent of states where permissible
Bankers acceptances (BA)	BA is a negotiable draft payable at a specified future date where a bank guarantees payment at maturity. BAs are used to make payment easier in the movement of goods, particularly international trade. The credit risk of a BA depends on the accepting bank. BAs are highly liquid and are bought and sold at a discount from par. (\$25,000–\$1,000,000 denomination)	78%
Commercial paper (CP)	CPs are unsecured short-term obligations (1 to 270 days) of a corporation usually sold at a discount from the par value. There is a strong secondary market, and the credit risk depends on the creditworthiness of the issuing corporation. (usually \$100,000 denomination)	90%
State investment pools	This is a money market fund operated by a state government for municipal and county investors. Pools provide good liquidity and diversification at substantially lower minimum investment requirements.	60%
Negotiable and non-negotiable certificates of deposit (CD)	Financial institutions issue certificates of deposit for maturities generally up to 1 year but may have maturities up to 5 years. The liquid (negotiable) aspects of the CD depend on the reputation of the issuing bank. Negotiable CDs are normally denominated in \$100,000 per round lot with interest paid semiannually or at maturity. Non-negotiable CDs can be purchased at much lower denominations, but there is no secondary market. Some states also require that CDs be obtained from banks within the state.	In state 88%; National 48%
Repurchase agreements (repo)	Repos are a collateralized transaction, where a securities dealer agrees to sell securities (usually U.S. Treasury or agency obligations) to the investor and then repurchase the securities at a predetermined price on some future date. The investor receives a fixed rate of interest as compensation for the exchange. Repos require \$100,000 to \$1 million to initiate and are negotiated between dealer and investor and are therefore difficult to sell on a secondary market. However, the maturities are overnight to 90 days and, therefore, are highly liquid. The risk of repos depends on the counter party and the market value of the collateral.	98%
U.S. treasury bills, notes, and bonds	These are highly liquid securities that are direct obligations of the U.S. Government, and therefore, are considered credit risk-free. T-bills have short maturities and are purchased at a discount. Notes and bonds are longer-term securities that generally pay semiannual interest and are vulnerable to market risk if sold before maturity.	96%
U.S. agency bonds	These are low-risk securities issued by U.S. Agency corporations (Fannie Mae, Freddie Mac, Sallie Mae, Ginnie Mae). They are issued in denominations of \$1000 and \$5000 with maturities usually more than a year, and therefore, they are vulnerable to market risk.	94%
State/local government obligations	State and local governments issue general obligation securities that are backed by the full faith and credit of the government and therefore have a low risk of default. Securities can be short-term notes or bonds with longer maturities. Revenues bonds have more credit risk, as the underlying collateral is a pledged revenue source. Interest on these securities is generally exempt from federal income tax and may be exempt from the state income tax.	70%

(Continued)

Table 1 Description and permissibility of select money market securities (*Continued*)

Security	Description	Percent of states where permissible
Corporate notes/bonds	These offer fixed income security that obligates the corporation to make scheduled payments to the bondholder. They vary in maturity and denomination and are vulnerable to credit and market risks.	68%
Corporate stocks	These are an equity instrument that constitutes an ownership share in a corporation and therefore may exhibit high financial risk. Stock share prices fluctuate widely with the market. There exists a strong secondary market for the stock of many publicly traded firms, but they are considered too risky by most organizations for short-term investment.	Domestic 32%; Foreign 22%
Derivatives	Derivatives are financial securities, where the market value depends on an underlying market indicator. For example, STRIPS are the principal portion of Treasury Bonds sold at a deep discount (zero coupon bond) that are extremely vulnerable to market risk.	14%

Permissibility percentages are calculated from NAST.^[14,15]

Investment strategies are generally classified as active or passive, depending on assumptions about market efficiency. An efficient market is one in which a security's price reflects all public information. This is important, because in an efficient market, it would be futile to devote time searching for underpriced bargains or avoiding overpriced lemons. A passive investment strategy assumes market efficiency. Therefore, there is no effort to play the market in order to enhance returns. The goal of a passive investment strategy is to maximize coupon income through a buy and hold approach. On the other hand, an active investment strategy assumes the portfolio can outperform the market by searching for undervalued securities and incorporating expectations about future market conditions. For example, a portfolio manager may purchase a 5-year treasury note with the expectation that interest rates are declining, and the note can be sold in 30 days at a capital gain. For more on investment strategies, see Ramsey and Hackbart.^[8]

Prudent Person

Many states and municipalities are implementing a prudent person rule. The prudent person rule emphasizes preservation of capital and discourages speculative transactions by requiring that the government's funds be invested with the same discretion and care that a reasonably well-informed person would use to manage their own affairs.^[9] The advantage of the prudent person rule is that the portfolio manager is held legally accountable for speculative investing, without the governing board being overly concerned that the investment policy may not delineate the permissibility of every new financial market innovation or

derivative. Still, the prudent person rule does not excuse a legislative body or governing board of its oversight responsibilities.

Local Government Investment Pools

Local government investment pools are increasingly popular for investing idle cash from state and local governments. Investment pools generally operate as a mutual fund, with investors purchasing a share of the portfolio. Total portfolio returns are distributed to the investors proportional to the number of shares. The advantage of investment pools is that the governments can obtain greater diversification at lower investment minimums. There are 30 local government investment funds operated through states for a combined portfolio worth more than \$143 billion, with more than 16,000 participating governments nationwide.^[10]

Outsourcing

Governments are increasingly outsourcing investment functions; nearly 65% of the cities responding to a 1999 survey utilized an external investment manager or consultant. While municipalities can delegate investment selection, the municipality remains responsible for the safety and liquidity of governmental funds. This does not mean, however, that the investment advisor does not have due diligence in adherence to investment policies and acting prudently. The Government Finance Officers Association (GFOA) has produced guidelines on the contracting of services that underscore the issues of investment priorities and permissible assets.^[11]

Table 2 Types of cash management services offered through banks

<i>Collection services</i>	<i>Disbursement services</i>
• Electronic (ACH) collections • Wire transfers • Preauthorized checks • Lockbox services • Armored car services • Night deposit • Coin counting services	• Electronic (ACH) payments • Wire transfers • Concentration and zero-balance accounts • Controlled disbursement accounts • Account reconciliation • Check imaging and digital archiving
<i>Credit services</i>	<i>Investment services</i>
• Revolving line of credit • Letter of credit • Short-term notes • Leasing • Community and economic development lending	• Certificates of deposit • Money market instruments • Custodial arrangements for repurchase agreements • Overnight sweep accounts

BANK RELATIONS

Bank relationships are vital to effective cash management. While banks provide core services like checking and savings accounts, banks are also expanding the range of cash management services available. Table 2 lists a range of services that banks provide to help the cash mobilization and investment efforts of governments. For an in-depth description of specific bank services, see Harrell^[12] and Redlin.^[13]

Banks do not provide services for free. Banks recoup the costs of these services through fees or compensating balances. The cost of compensating balances to a government is the foregone interest revenue. The decision to utilize banking services should consider the costs and benefits of the arrangement, keeping in mind that the charges for specific services often vary from bank to bank. As with any outsourcing, contracts should detail the services provided and specify fees and other compensation.

CONCLUSION

Cash is the lifeblood of a public organization, and cash management ensures the vitality of the organization. The four primary components of cash management are cash forecasting, cash mobilization, investment, and bank relations. Cash management must be adapted to meet the needs of a specific public organization. For some organizations, the fundamentals of cash management discussed here may be a sufficient foundation in forming effective cash management processes. Most, however, will need to explore the cash

management components in more detail through the references and additional study. As long as governments continue to bolster non-tax revenues and feel the pressure to improve fiscal efficiency, cash management will play a vital role in public administration.

REFERENCES

1. Adam, P.S.; Harrison, W.A. *Essentials of Cash Management*, 6th Ed.; Treasury Management Association: Bethesda, MD, 1998; Chapter 11.
2. McAndrews, J. The automated clearinghouse system: Moving toward electronic payment. *Bus. Rev.* **1994**, 15–23 (July/August).
3. Government Finance Officers Association. *Recommended Practices for State and Local Governments*; Government Finance Officers Association: Chicago, IL, 2001; 44, May.
4. National Association of State Treasurers. *State Treasury Activities and Functions*; National Association of State Treasurers: Lexington, KY, 2001; 44, Table 20.
5. Denison, D.V. How conservative are municipal investment practices in large cities. *Munic. Finance J.* **2002**, 23 (1).
6. Denison, D.V. *States' Investment of Idle Cash Balances*; Association for Budgeting and Financial Management National Conference: Washington, DC, 1998, November.
7. Golembiewski, P.; Bornholdt, G.; Jones, T. *Allocation and Accounting Regulations For Arbitrage Bonds*; http://www.irs.ustreas.gov/plain/bus_info/eo/topics.html (referenced June 2001).
8. Ramsey, J.R.; Hackbart, M.M. Public funds management: current practices and future trends. In *The Handbook of Municipal Bonds and Public Finance*; Lamb, R., Leigland, J., Rappaport, S., Eds.; New York Public Institute of Finance: New York, 1993.
9. Miller, G.; Larson, M.C.; Zorn, W.P. *Investing Public Funds*, 2nd Ed.; Government Finance Officers Association: Chicago, IL, 1998; 44 pp.
10. National Association of State Treasurers. *State Treasury Activities and Functions*; National Association of State Treasurers: Lexington, KY, 2001; 76 pp.
11. Government Finance Officers Association. *Recommended Practices for State and Local Governments*; Government Finance Officers Association: Chicago, IL, May 2001; 59–61.
12. Harrell, R. *Banking Relations*; Government Finance Officers Association: Chicago, IL, 1986.
13. Redlin, B.M. Cash management technology for small governments. *Gov. Finance Rev.* **1999**, 15 (4), 55–57.
14. National Association of State Treasurers. *State Treasury Activities and Functions*; National Association of State Treasurers: Lexington, KY, 2001; 59, Table 28.
15. National Association of State Treasurers. *State Treasury Activities and Functions*; National Association of State Treasurers: Lexington, KY, 2001; 78, Table 38.

Chinese Public Administration: History and Current Trends

Yijia Jing

Department of Public Administration, Fudan University, Shanghai, P.R. China

INTRODUCTION

Chinese public administration (PA) has been undergoing a significant transformation since late 1970s. Despite its kaleidoscopic developments, the major underlying impetuses are twofold: the historical heritage and the modernization momentum. This chapter lays out a miniature of Chinese PA by a brief review of the administrative history of China and a generalization of the recent tendencies in Chinese PA. The analysis makes it explicit that Chinese PA is experiencing a rapid paradigm-shifting process toward modernization.

A BIRD'S-EYE VIEW OF CHINESE ADMINISTRATIVE HISTORY

China has a very long administrative history, spanning thousands of years. According to the nature of administration and its relation to politics, such a history has three major stages of development.

Royal Administration (pre-1912 Period)

Kingly administrative history of China began in Xia Dynasty in the 22nd century B.C. In 221 B.C., *Qinshihuang* united China and established a centralized and complex administrative system. Until the end of Qing Dynasty (1644–1912), subsequent feudal empires created glorious ancient civilization. The ancient royal administration of China was characterized by several features. First, it had a highly centralized structure of power. The emperor was an absolute ruler with legislative, executive, judicial, and spiritual powers. Local officials were appointed by the emperor. Second, a formal examination system, *Kejuzhi*, was created since Sui Dynasty (581–618) and became a major channel for the government to recruit talented and educated literati, some of whom were from the lower class. Orthodoxy education and examination were strictly based on Confucianism and its later ramifications. Third, morality was deemed as essential for good governance. Officials were evaluated regarding their loyalty to the emperor, their understanding of the Confucian philosophy, and their dedication to the popular well-being. Fourth, despite the generalist trend

and the focus on moral tenets, sophisticated administrative institutions and a delicate system of division of labor were developed to administer the society. Ancient bureaucrats were specialists called *li*, different to *guan*, officials who were selected through state examinations, assigned political leadership, and hardly equipped with administrative competence. *Li* were the standing forces of the bureaus and possessed professional skills, capabilities, and institutional knowledge.^[1]

As Wilson argues that “the desire to keep government a monopoly made the monopolists interested in discovering the least irritating means of governing,” the carefully designed Chinese ancient administrative systems performed well in many aspects such as finance, public works, and relief.^[2] Without steam engine and telegraph, it could govern a population over 200 million since the late 18th century.^[3] Yet, the highly personal system lacked a built-in mechanism to sustain development and prevent social inequality, corruption, and tyranny. The ancient Chinese history roughly exhibited a 200-year cycle of new dynasty, prosperity, decline, peasant war, and new dynasty.

State Administration (1912–1978 Period)

The Qing Dynasty was ended by Republic of China (ROC, 1912–1949) in 1912 through violent revolution. ROC was replaced by People's Republic of China (PRC, 1949–) in 1949 in the same way. Between 1912 and 1978, it was a period of modern state building and nation building. Administration mainly served the authoritative state, with a Leninist party as an instrument of political struggle.

1912–1949 period

The early stage of ROC was a period of warlord battles and disintegration. The Kuomintang Party (KMT) founded the Nationalist regime in Canton in 1925 and established the State Council as the highest executive branch. It was a highly integrated party–state system and was maintained after 1928, when KMT reunited China. In 1929, the *Five-Yuan System*, including the *Executive Yuan*, the *Judicial Yuan*, the *Legislative Yuan*, the *Examination Yuan*, and the *Control Yuan*, was established according to Sun Yat-Sen's *Five-Power* theory. Although this system was designed

under the idea of the check and balance of powers, it was finally subject to the control of Central Political Committee of KMT.^[4] Due to the Japanese invasion (1931–1945), the two civil wars (1927–1937 and 1945–1949) between the KMT and the Chinese Communist Party (CCP), and the actual warlord control over many provinces, the administration of ROC mainly served wars, although it unsuccessfully endeavored to introduce many modern institutions, for example the Comptroller-General Office with independent authority of budgeting.^[4]

1949–1978 period

National security and social transformation were CCP's two primary missions during the first three decades of PRC. Imitating the USSR socialist regime, the society was organized as a "state syndicate," which substituted the market by centrally made plans and directions in distributing and redistributing resources. First, administrative power was vertically concentrated to the central government and horizontally concentrated to party leaders. The party had a similar organizational structure to that of the government and held the decision-making power. Second, state-owned enterprises (SOEs) almost monopolized industry and commerce. Fast capital accumulation of heavy industry was pursued by using the *Residence Registration System* to limit migration of peasants to urban areas and by enforcing significant price differences between agricultural and industrial products. Third, the government was omnipotent and penetrated almost all social spheres. Due to the high level of integration between the party, the state, the economy, and the society, the expansion of PA system lacked basic institutional constraints. A vicious cycle of administrative "streamlining–swelling–streamlining again–swelling again" emerged.^[5] The administrative system was driven by communist ideology and focused on class struggle, which culminated during the cultural revolution (1966–1976). Despite the dual emphasis on "red and professional," morality mandate, namely the loyalty to the party and its ideals, was the overwhelming criterion in recruiting and evaluating a cadre.

Public Administration (post-1978 Period)

After the third Plenary of the 11th CCP's National Conference in 1978, CCP recognized peace and development as the world's dominant themes and adopted the reform and open policy. China's administrative system gradually incorporated modern values of PA and exposed itself to popular control. China's PA system experienced a major transformation, as will be analyzed in the following.

CURRENT TRENDS IN CHINESE PA

Reidentification of Role and Missions

Pre-1978 regime was extremely syncretized and deprived the society of its capacity of homeostasis. In the post-1978 era, Chinese PA has been substantially relocated and rationalized by functional differentiation. First, economic development and market reform became PA's major tasks. The government gradually separated itself from economic operation through deregulation and privatization. Plans and direct intervention into the economy were largely reduced. SOEs were either sold or incorporated. From 1978 till 2004, the share of SOEs and state-holding companies in national industrial output decreased from 77% to 35%.^[6] China has been actively seeking the acknowledgement of its market economy status in international community. Second, the government has retreated successfully from many social spheres and shed its functions. The rural communes collapsed and the urban working unit system quit their social services.^[7] Grassroots governance was, to some extent, shifted to self-government organizations, including *Villagers' Committee* and *Urban Residents' Committee*. Third, not so significantly, there was a trend to separate the party from the government (*dang zheng fen kai*). Reformers in 1980s attempted to limit the party role to political leadership and enlarge the authority of professional administrators in personnel management. Yet, since 1989, the official ideology regressed and such efforts were reversed.^[8] Fourth, intergovernmental relations were streamlined. Jurisdictions of central and local governments were adjusted, according to economic and administrative principles. Decentralization was a major trend since 1980s.

Adjustment of Administrative Values

Post-1978 Chinese PA, to some extent, internalized modern administrative values that reshaped its behavior and dynamics.

Democratic values

Democratization of China's PA system has two aspects. First, the legislature, People's Congress, has strengthened its legislative and supervisory functions in recent years and imposed on governments increasing burden of legal and managerial accountability. Second, democratic values such as transparency and participation were incorporated. While a *Government Information Act* is not available yet, citizen's right to know has been taken care of incrementally, for example by requiring governments at and beyond county level to have websites and publish specified information regularly. In 2003, the Head

of Ministry of Health and the Mayor of Beijing were deposed partly due to their failure to disclose the information of the infection of SARS to the public. Policy making was made more accessible. Without having binding decisions, public hearings, including administrative penalty hearing, price hearing, and legislation hearing, were initiated since late 1990s.^[9] Public opinion is increasingly taken into decision-making process.

Legal values

Traditional Chinese PA was not a system of rule of law, but of man. The post-1978 period highlighted a gradual adaptation to modern legal standards. First, protection of individual rights improved. The 24th Constitutional Amendment passed in 2004 added to the 33rd Article that “the state respects and protects human rights.” Besides the constitutional enhancement, the practice of PA also demonstrated a “credible commitment” of property rights protection.^[10] Second, administrative laws prospered and shifted the focus from administrative empowerment to equal rights protection. In 1990, the *Administrative Litigation Law* was enacted, which ensured the equal status of non-governmental actors in suing against administrative agencies. Third, procedural due process was strengthened. Besides procedural laws in special administrative areas, for example 1996 *Administrative Penalty Law* and 2004 *Administrative Permission Law*, a comprehensive *Administrative Procedure Law* has been under official consideration for almost two decades. Like its U.S. counterpart, the 1946 *Administrative Procedure Act*, this proposed law introduces legal procedures and promotes values such as transparency, accountability, participation, and limitation on intrusiveness. Yet, law enforcement was seriously hindered by two facts. First, despite the legal authorities of People’s Congress in creating, budgeting, and reviewing administrative agencies, the government is de facto accountable to the party. Second, the judicial system was also non-independent. The 2006 *Corruption Perception Index* ranked China as the 70th in its level of incorruption among the 163 countries and regions, identical to that of India, Brazil, and Mexico.^[11]

Managerial values

Performance was emphasized to serve pragmatic goals of economic development. Since 1980s, PA was professionalized and rationalized, to some extent, following the Weberian ideal-type bureaucracy. Recruitment, training, and career track were normalized. Position responsibility system was introduced. In 1988, the Ministry of Personnel was founded. In 1989, State Civil Servant Examination began to replace the old labor allocation system in staffing non-leading positions. In 1993, the *Provisional Regulations for State Civil*

Servants were promulgated, replacing the cadre system with a civil service. In 1994, State Administrative School was founded for administrative research and education. Yet, neutral competence, the core of civil service, was not established. Further, while civil servants in general got younger, better educated, and better trained, there was a biased favor of economic technocrats, with a background of engineering, SOE management, or economic planning.^[5] In late 1990s, “reinventing government” bandwagon got popular in China. Among the two themes of New Public Management, managerialism achieved little in China, while public choice, in the form of privatization, gained enormous success.^[12]

Administrative Empowerment

Besides functional differentiation and value adjustment, organizational restructuring was taken to rejuvenate the administrative system and enhance administrative capacity.

Adjustment of the central–local relation

The central government has taken all efforts to motivate, coordinate, and control the local governments. Despite a general trend of decentralization, central government managed to strengthen itself. In early 1980s, central and local governments redelineated their scope of authority. “The subnational governments have primary authority over the economy within their jurisdictions,” while “the national government has the authority to police the common market and to ensure the mobility of goods and factors across subgovernment jurisdictions.”^[13] A parallel decentralized fiscal system, plus the lack of taxation capacity, led to a weak state and a weak central government in collecting and distributing economic resources.^[14] Local state corporatism, as well as localism, rose in China due to the direct link between local economic growth, local fiscal revenues, and cadres’ evaluation and compensation.^[15] As a response, in 1994, the central government introduced the tax-assignment system to “replace the old, discretion-based system of revenue-sharing with a new rule-based system of revenue-sharing.”^[16] The proportion of budgetary revenues in national GDP rose continuously from the lowest point of 10.7% in 1995 to 19.3% in 2004. By spending 27.7% of fiscal expenditures but collecting 54.9% of the fiscal revenues in 2004, the central government obtained the upper hand in influencing local decision making.^[6] Compatible with the transfer of responsibilities to localities, “two-level-down management” replaced “one-level-down management” in personnel appointment after 1984. Local officials were delegated full authority in appointing their direct subordinates.^[17]

Reinventing government

The efforts of Chinese PA to actively reshape and adapt to the government–business and state–society relations can be covered under the “reinventing government” umbrella. Since early 1980s, many governmental functions were decentralized, shed, or transformed to serve the market reforms. Administrative directions and plans gave way to legal and economic means. Governments were expected to be rule makers, adjudicators, and public goods suppliers, rather than direct market participants. Administrative agencies were reorganized, according to the general idea to incorporate their productive functions and maintain their regulatory functions. Industrial ministries or bureaus such as the Electronic Industry Ministry or Coal Ministry were either revoked or demoted. Accordingly, there was enlarging boundary blurring in providing public services. Increasing fiscal revenues, loss of direct operational capacity, growing market and society, and the agreed-upon role of the government in promoting domestic demands created a significant incentive of governments in co-opting external resources by economic means. Public–private partnership developed quickly, and a simple estimation shows that from 2001 to 2004, the aggregate level of outsourcing was between 20% and 30%.^[18] Major outsourcing went to business organizations, creating a major source of corruption. In 2006, China launched a special campaign to strike business bribes. Non-profit organizations are getting more important in providing social and human services. These organizations are often small and affiliated to governments. Another important aspect is the adjustment of the means and scope of government’s social intervention. After the collapse of the rural commune system and the decline of urban working unit system, the omnipresent state penetration declined. Mass mobilization and movement were no longer useful in motivating and monitoring local officials. State retreat and the weak civil society created the vacuum or aperture of governance. Spontaneous grassroots mass organizations, such as *Villagers’ Committee*, were officially recognized, promoted, led, and financed by the government. Non-governmental organizations, such as Homeowner’s Association, developed very fast, and their collective private governance received favorable treatments from the government. Corporate social responsibility was also promoted by the government.

CONCLUSIONS

This article reviews the royal, state and public administration stages of Chinese administrative history, which essentially corresponded to China’s political development.

Functional differentiation, value adjustment, and organizational restructuring were major drives that placed Chinese PA on an irreversible track of modernization since late 20th century. Due to many unique factors, such as the huge size of the population and the special historical path, such a modernization process has to be long and full of Chinese characteristics.

REFERENCES

1. Zhang, C. Public administration in China. In *Public Administration in China*; Mills, M., Nagel, S., Eds.; Greenwood Press: Westport, CT, 1993; 3–19.
2. Wilson, W. The study of administration. *Polit. Sci. Q.* **1887**, 2 (2), 197–222.
3. Durand, J. The population statistics of China, A.D. 2–1953. *Popul. Stud.* **1960**, 13 (3), 209–256.
4. Chien, T. War-time government in China. *Am. Polit. Sci. Rev.* **1942**, 36 (5), 850–872.
5. Liu, M. *Administrative Reform in China and Its Impact on the Policy-making Process and Economic Development after Mao: Reinventing Chinese Government*; Edwin Mellen Press: Lewiston, NY, 2001.
6. State Statistical Bureau. *Chinese Statistical Yearbook 2005*; China Statistical Press: Beijing, 2005.
7. Guan, X. China’s social policy: reform and development in the context of marketization and globalization. *Soc. Policy Admin.* **2000**, 34 (1), 115–130.
8. Tao-chiu, L.; Chan, H. Reforming China’s cadre management system: two views of a civil service. *Asian Surv.* **1996**, 36 (8), 772–786.
9. Yang, K.; Lauer, H. Assessing China’s public price hearings: symbolic aspects. *Int. J. Public Admin.* **2003**, 26 (5), 497–524.
10. Perotti, E. Credible privatization. *Am. Econ. Rev.* **1995**, 85 (4), 847–859.
11. Transparency International (TI). *2006 Corruption Perceptions Index*; <http://www.transparency.org>.
12. Hood, C. A new public management for all seasons? *Public Admin.* **1991**, 69 (1), 3–19.
13. Montinola, G.; Qian, Y.; Weingast, B. Federalism, Chinese style: the political basis for economic success in China. *World Polit.* **1995**, 48 (1), 50–81.
14. Wang, S.; Hu, A. Zhong Guo Zheng Fu Ji Qu Neng Li De Xia Jiang Ji Qi Hou Guo. The declining extractive capacity of Chinese governments and its consequences. *21 Century* **1994**, (1), 5–14.
15. Oi, J. Fiscal reform and the economic foundations of local state corporatism in China. *World Polit.* **1992**, 45 (1), 99–126.
16. Wang, S. China’s 1994 fiscal reform: an initial assessment. *Asian Surv.* **1997**, 37 (9), 801–817.
17. O’Brien, K.; Li, L. Selective policy implementation in rural China. *Comp. Polit.* **1999**, 31 (2), 167–186.
18. Jing, Y. Zhong Guo Gong Gong Fu Wu Wai Bu Gou Mai Shui Ping Shi Zheng Fen Xi. Analysis of the aggregate outsourcing of public services in China. *Manage. World* **2007**, (2), 37–43.

Citizen Participation in Local Government

Anthony J. Nownes

Department of Political Science, University of Tennessee, Knoxville, Tennessee, U.S.A.

INTRODUCTION

Over the years, scholars have devoted considerable attention to the topic of citizen participation in local government and politics. In this brief article, I will summarize what these scholars have learned about this important topic. I will begin with a cautionary note on the inherent difficulty of understanding citizen participation in local government. From here, I will examine the most well-known form of citizen participation in local politics—voting. Next, I will discuss the linkage institutions through which citizens participate in local politics—political parties and interest groups. Penultimately, I will examine citizen contacting. Finally, I will say a few words about other forms of citizen participation in local government and politics.

UNDERSTANDING CITIZEN PARTICIPATION IN LOCAL GOVERNMENT

Arriving at a comprehensive understanding of citizen participation in local government is tricky at best. First, there is the problem of variety. In the United States, there are many types of local governments, including authorities, cities, counties, villages, municipalities, districts, towns and townships, and metropolitan governments. Because these types of governments differ from one another, learning about citizen participation in one may tell us little about citizen participation in any of the others. The problem of variety is compounded by the fact that *within* specific types of local government, there is also considerable variety. For example, city governments generally come in three types—mayor-council, commission, and council-manager. Because different institutional arrangements make for different opportunities and incentives for citizen participation, this variety makes it difficult fully to comprehend and understand citizen participation in local government.

The task of understanding citizen participation in local government is made all the more difficult by the fact that citizen participation is itself a multifarious phenomenon. There are literally dozens of forms of political participation, including voting, working on political campaigns, contributing money to candidates for public office, contacting public officials, protesting

or demonstrating, joining political organizations, and informally discussing politics with others. A full understanding of citizen participation would require a full understanding of a large variety of forms of participation.

In short, if there is one thing we know about citizen participation in local government and politics, it is that we are far from a comprehensive understanding of the phenomenon. This lack of knowledge, as they say, “comes with the territory.” As scholars and students then, we cannot possibly hope fully to comprehend citizen participation in local government in its entirety. Instead, we simply must look to gain a broad understanding of the phenomenon—one that gives us a general idea of what is going on in the over 80,000 local governments in the United States.

VOTING, ELECTIONS, AND CAMPAIGNS

Voting is the most obvious form of political participation in the United States. Voting is also the simplest and most straightforward method by which citizens make their wishes, demands, and opinions known to government officials. The broad topic of voting can be broken into two specific subtopics: voter turnout and voter choice. In this section, I will discuss each briefly. I will also discuss direct democracy.

Voter Turnout (Who Votes?)

The question of who votes in local elections invites an obvious answer: *hardly anybody*. Although turnout rates vary widely, turnout in local elections is typically below 30%. Turnout levels of below 20% are not unusual. There are many reasons for low voter turnout in local elections. Three reasons in particular stand out. First, large majorities of citizens simply are not interested in local politics. Apparently, issues such as zoning and solid waste disposal are not of interest to many voters (though they probably should be). Second, many local elections are non-partisan. Overall, approximately two-thirds of all local elections are nonpartisan, which means that candidates do not appear on the ballot with a party affiliation. Studies show that partisan election campaigns generate far more citizen interest than non-partisan campaigns.

Third, local elections are often held in “off years” or “off times”—that is, when no high-level government officials (for example, the president and the governor) are being chosen. This tends to depress interest and turnout in local elections. Research on who actually votes in local elections has produced very few surprises. As is the case in elections at the state and national levels, in local elections affluent and well-educated citizens tend to vote at much higher rates than not-so-affluent and not-so-well-educated citizens.

One topic that has received a great deal of research attention over the years is how local government structure affects voter turnout. The evidence is somewhat ambiguous, but four findings are widely accepted. First, cities with council-manager governments have lower levels of turnout than cities with mayor-council governments. Second, cities in which mayors and/or council members serve full time and have staff experience relatively high levels of turnout. Third, local governments that directly elect the executive have higher levels of turnout than those that do not. Finally, voter turnout tends to be higher in places with significant social (e.g., racial) cleavages.

Voter Choice (Who Wins?)

As for who wins local elections, a few patterns emerge. First, incumbents generally win. Like incumbents running for state and federal office, local incumbents tend to have higher name recognition than challengers and thus tend to attract more media attention and money than challengers. Second, in many places, race plays a substantial role in determining electoral outcomes. For example, there is evidence that voters, especially non-white voters, tend to support candidates of the same race. (There is also evidence, however, that overtly racial electoral campaigns are not particularly effective.) Third, in low-profile local elections (such as those for Clerk of Deeds or County Supervisor), name recognition is crucially important, so candidates with long histories of civic involvement and extensive contacts in the business world and local associations tend to fare best. Finally, the last 15–20 years or so have been extraordinarily fruitful for non-white candidates. For example, approximately one-third of all large cities (cities with more than 200,000 people) now have nonwhite mayors, and there are well over 3000 Hispanic or Latino local elected officials serving in the United States.

Direct Democracy

Before we leave the topic of voting, it is important to note that in many localities citizens have the right directly to make policy by voting on initiatives

and/or referenda. Generally, there are hundreds of local initiatives and referenda on the ballot every election year. There is some evidence that particularly controversial initiatives or referenda can draw relatively large numbers of voters to the polls. As for which side prevails in direct democratic elections, there is a great deal of controversy surrounding this question. On the one hand, some argue that “money talks” and the side that spends the most has a huge advantage. On the other hand, some scholars maintain that money is hardly ever the determining factor in a direct democratic election. For now, this remains an open question.

LINKAGE INSTITUTIONS

Of course, citizens who are interested in local government can do more than just vote. Another way they can participate in local politics is by becoming active in linkage institutions—political parties and interest groups.

Interest Groups

Many years ago, local government expert Paul Peterson wrote that local politics is largely “groupless.” He may well have been correct at the time. But today, groups are everywhere in local politics.

The dominance of business

Interest groups, which are broadly defined as organizations that seek to influence government decisions, are decidedly active in local politics. Research has shown that several types of organizations regularly and actively seek to affect the decisions of local governments of all kinds. First, there are business firms. Virtually all comprehensive treatments of local government decision-making note that business elites—the people who own and run business enterprises—are key players in local politics and often have close relationships with local government officials. Not surprisingly, some types of businesses are more active in local politics than others. For example, builders, construction companies, and real-estate developers are very active, primarily because much of what they do is regulated by local governments. Banks are also quite active. Second, there are business associations. Chambers of Commerce and junior Chambers of Commerce, for example, are active in localities of all shapes and sizes, and their leaders and activists often have a great deal of clout with government officials. Civic clubs such as the Rotary Club, the Lions Club, and Kiwanis, which consist largely of local businesspeople, are also quite active.

Non-Business organizations

Businesses and business organizations are not the only types of interest groups active in local politics. In recent years, neighborhood groups and “not in my backyard” (NIMBY) groups have become quite prolific and powerful in local politics. These groups are generally interested in quality-of-life issues surrounding zoning and planning and often form to oppose new projects that they think will decrease property values or increase traffic, pollution, noise, or crime. In some places, especially large cities, labor unions are also powerful interest group actors. Many local government employees, including police officers, fire fighters, and teachers, are union members, and their unions are often powerful actors in local politics. Municipal unions are most visible when budget time rolls around and their salaries and benefits are determined.

Finally, a variety of other types of organizations, including churches, local environmental groups, and the local affiliates of nationally active groups, are also active in local government when the issues they care about reach the local governmental agenda.

How much power do interest groups wield?

The fact that interest groups are active at the local level begs the following question: How influential are interest groups in local politics? The answer appears to be “very powerful.” Studies conducted in the 1940s, 1950s, and 1960s, for example, showed that business interests—individual business firms and business associations—were dominant players in many localities. In recent years, a more balanced view has won out, as scholars have shown that while business interests are almost always “at the table” when local government decisions are made, they are not all-powerful. Research conducted over the past two decades has shown, for example, that labor unions exert substantial influence over the taxing and spending policies of many local governments. Moreover, research on neighborhood groups and NIMBY groups has demonstrated that groups of citizens are often very influential players in local politics, especially land-use politics.

Political parties

There is not a great deal of research on political parties at the local level. This is the case primarily because large numbers of localities have non-partisan elections. Moreover, people who study political parties tend to focus on parties in national (and, to a lesser extent, state) politics. Nonetheless, political parties are important players in local politics. The high point of local

party strength came in the early 1900s, when literally hundreds of localities were ruled by political party “machines” that nominated candidates for office and virtually bribed voters to support them with tangible material incentives such as government jobs, enhanced services, and favors. By mid-century, most party machines had lost strength, and today the party machines of old are gone.

At the other end of the spectrum from party machines are local party organizations that do virtually nothing. Research shows that in many places local party organizations are either inactive or understaffed. This does not mean, however, that local party organizations are completely moribund. Today, perhaps as many as 50–60% of local party organizations are active, engaging in a large variety of activities including recruiting candidates, raising money, campaigning for candidates, contributing money to candidates, and conducting get-out-the-vote drives.

As for the influence of local party organizations, it is difficult to determine. But the extant research suggests a few conclusions. First, in many places local party organizations play a vital role in the recruitment of candidates for local, state, and even federal office. Second, in the last decade or so, local party organizations have become increasingly adept at fund-raising (as have state and national party organizations) and thus are invaluable to many candidates. Finally, even in localities with non-partisan elections, local party organizations play a crucial role in providing cues to voters and mobilizing citizens to vote.

CITIZEN CONTACTING

Citizen contacting—that is, the practice of citizens contacting acting government officials with either requests for services or complaints—is a common form of citizen participation in local government. Moreover, because most government services are provided by local governments rather than state governments or the national government, it has been studied extensively by scholars of local politics. The research on citizen contacting has produced several findings of note. First, levels of contacting are surprisingly high. In fact, in a typical year, far more citizens contact local government officials than vote. Second, homeowners and citizens with minor children engage in more contacting than renters and citizens without minor children. Apparently, people who own homes and/or have children tend to feel that they have a large stake in their communities. Third, citizens who pay attention to local politics engage in contacting much more than citizens who do not. Fourth, citizens with high levels of political efficacy engage in more contacting than citizens with low levels of efficacy. Fifth,

socioeconomic status is not strongly related to levels of contacting. In fact, there is some evidence that unlike most forms of political participation, contacting is done as much by citizens from relatively low socioeconomic strata as by citizens from relatively high socioeconomic strata. Sixth, citizens increasingly contact local government officials via the Internet, as more and more local governments stake out an on-line presence. Finally, citizens who engage in citizen contacting generally express relatively high levels of satisfaction with local government. In other words, they are generally pleased with the response they receive when they contact government officials.

OTHER FORMS OF PARTICIPATION

There are several other forms of citizen participation that I have not covered here. Perhaps the most common is participating in public hearings (which are common at the local level, especially on land-use matters). This form of participation can be effective, especially if citizen participation is extensive. Among less common forms of citizen participation are serving on local government boards and commissions, participating in surveys conducted by government officials, protesting and demonstrating, running for public office, campaigning for candidates, contributing money to candidates' campaigns, and attempting to influence appointment to public office.

CONCLUSION

There is still a lot we do not know about citizen participation in local government. In fact, it is fair to say that this topic is ripe for further research attention. Nonetheless, we do know a few things about citizen participation in local government and politics. First, participation in local government is neither common nor unusual. On the one hand, large majorities of citizens fail to turn out for local elections. On the other hand, large numbers of people engage in citizen contacting. Second, political parties and interest groups are important players in local politics. Although no one thoroughly has taken account of the power of parties and interest groups in local politics, it is safe to say that they are influential players in many places. Third, people can make a difference in local politics. The successes of both neighborhood and NIMBY groups in stopping controversial developments and the astounding increase in the number of non-white local elected officials attest to the power of citizen participation.

BIBLIOGRAPHY

1. Abney, G.; Lauth, T.P. Interest group influence in city policy-making: the views of administrators. *West. Polit. Q.* **1985**, 38 (1), 148–161.
2. Berry, J.M.; Portney, K.E.; Thomson, K. *The Rebirth of Urban Democracy*; The Brookings Institution: Washington, DC, 1993.
3. Brown, M.C.; Halaby, C.N. Machine politics in America, 1870–1945. *J. Interdiscip. Hist.* **1987**, 17 (3), 587–612.
4. Cronin, T.E. *Direct Democracy: The Politics of Initiative, Referendum, and Recall*; Harvard University Press: Cambridge, MA, 1989.
5. Dahl, R.A. *Who Governs? Democracy and Power in an American City*; Yale University Press: New Haven, CT, 1961.
6. Elkins, D.R. The structure and context of the urban growth coalition: the view from the Chamber of Commerce. *Policy Stud. J.* **1995**, 23 (4), 583–600.
7. Ferman, B. *Challenging the Growth Machine: Neighborhood Politics in Chicago and Pittsburgh*; University Press of Kansas: Lawrence, KS, 1996.
8. Fleischmann, A. Participation in local politics. In *Handbook of Research on Urban Politics and Policy in the United States*; Vogel, R.K., Ed.; Greenwood Press: Westport, CT, 1997; 146–158.
9. Gibson, J.L.; Cotter, C.P.; Bibby, J.F.; Huckshorn, R.J. Whither the local parties? a cross-sectional and longitudinal analysis of the strength of party organizations. *Am. J. Pol. Sci.* **1985**, 29 (1), 139–160.
10. Hajnal, Z.L.; Lewis, P.G. Municipal institutions and voter turnout in local elections. *Urban Aff. Rev.* **2003**, 38 (5), 645–668.
11. Hirlinger, M.W. Citizen-initiated contacting of local government officials: a multivariate explanation. *J. Polit.* **1992**, 54 (2), 553–564.
12. Kaufmann, K.M. *The Urban Voter: Group Conflict and Mayoral Voting Behavior in American Cities*; The University of Michigan Press: Ann Arbor, MI, 2004.
13. Krebs, T.B. The determinants of candidates' vote share and the advantages of incumbency in city council elections. *Am. J. Pol. Sci.* **1998**, 42 (3), 921–935.
14. Logan, J.R.; Molotch, H. *Urban Fortunes: The Political Economy of Place*; University of California Press: Berkeley, CA, 1987.
15. Logan, J.R.; Rabrenovic, G. Neighborhood associations: their issues, their allies, and their opponents. *Urban Aff. Q.* **1990**, 26 (1), 68–94.
16. Peterson, P.E. *City Limits*; The University of Chicago Press: Chicago, 1981.
17. Schaffner, B.F.; Streb, M.; Wright, G. Teams without uniforms: the non-partisan ballot in state and local elections. *Polit. Res. Q.* **2001**, 54 (1), 7–30.
18. Smith, K.B.; Greenblatt, A.; Buntin, J. *Governing States and Localities*; CQ Press: Washington, DC, 2005.
19. Stone, C.N. *Regime Politics: Governing Atlanta, 1946–1988*; University Press of Kansas: Lawrence, KS, 1989.
20. Wood, C. Voter turnout in city elections. *Urban Aff. Rev.* **2002**, 38 (2), 209–231.

Citizen Participation in the Budget Process: Exit, Voice, and Loyalty

Carol Ebdon

*Department of Public Administration, University of Nebraska at Omaha,
Omaha, Nebraska, U.S.A.*

INTRODUCTION

Various methods enhance citizen participation in the budget process. Even though empirical evidence is sparse, participation appears to be beneficial when it educates citizens in the complexities and costs of government services, and may help to reduce distrust of government. There is still little evidence, partly due to methodological difficulties, that participation actually changes budget outcomes, although several cases and anecdotal stories support the notion that there is some effect.

In spite of the perceived benefits of participation, opportunities are still not prevalent. This is particularly true early in the budget process, when public opinion would conceivably have the greatest impact on decision making. To move beyond education to true two-way dialog regarding the budget is not easy, due to structural and cost barriers, difficulties in gaining input from numerous representative citizens, and conflicts in the roles and attitudes of both administrators and elected officials.

PARTICIPATION MECHANISMS

Researchers looking at general participation mechanisms have found that it is most beneficial when it occurs early in the process and when it is two-way deliberative communication, rather than simply one-way information sharing.^[1] In the case of the budget process, then, this requires more than the ubiquitous public hearing that is usually held at the end of the process, just before final adoption by the legislative body. Citizens also have a voice in direct voter referenda, which are often required for governments to increase taxes or to issue general obligation debt. Again, though, referenda usually occur at the end of the process, and voters may not be well informed about the potential consequences of their actions.

Citizen surveys can be used early in the budget process to determine service satisfaction levels, needs, and

priorities. For example, the city of Auburn, Alabama, used surveys prior to budget development; citizens reviewed a list of city services and were asked to which they would give the highest priority in the upcoming budget.^[2] Surveys can show officials citizen satisfaction levels, responses may represent views of the public at large, and if used consistently over time can show important trends in opinions. However, appropriate question design is important, and surveys may not reflect the intensity of a respondent's opinion.^[3] In addition, cost is a factor, and the responses may not be well thought out.^[4]

Public meetings may also be used to discuss budget issues during or prior to budget development. These forums provide citizens the opportunity to voice their opinions and can provide for two-way communication early in the process. However, attendance is often low and may not represent the community as a whole, and participants may have insufficient knowledge for effective input.^[3] Some communities use public meetings in an intensive way for a broad "community visioning" process to plan for the future or allow citizens to work through options related to major issues.^[5] A Minnesota school district superintendent used a public deliberation process to cut \$2.4 million from the budget after significant state aid reductions. The process lasted 4 months, beginning with recommendations from teachers, administrators, staff, and students, followed by reviews and cost analyses. Over 2000 citizens then attended public meetings held throughout the district to discuss the alternatives. The superintendent finally reviewed 4000 recommendations; her final choices were approved unanimously by the board of education.^[6]

Citizen advisory committees are another participatory mechanism applied to the budget process. These may be used during budget preparation, as a review of the proposed budget, or may deal with specific programs, such as the Capital Improvement Program, Community Development Block Grant funds, or social service agency funding. Advisory committees allow members to develop expertise in an area, so that input may be broader and more in-depth than that received

through surveys or public meetings. However, these mechanisms can require more effort and time on the part of both city administrators and participants, and may not be representative of the community.^[3] A study of advisory committees in New Jersey municipalities found that committees were perceived as effective when the appointment process was democratic, and the committee had clearly stated goals and objectives.^[7]

Some communities use several types of participatory mechanisms in conjunction with each other. Eugene, Oregon, in the face of severe fiscal stress, developed a deliberative process, with a series of input methods, including public meetings, a budget-balancing exercise, and citizen surveys with varying levels of service cost information provided.^[4] Hillsborough County, Florida, combines a series of public hearings during budget consideration with community meetings throughout the year, advisory committees, and surveys.^[8] The benefits of one method can offset disadvantages of another.

FREQUENCY OF USAGE

To what extent is input sought in the budget process? Participation in budgeting is apparently relatively low in local governments, which is where most surveys have been conducted. One study found extensive use of public hearings and budget summary documents across various local government types, but little usage of other methods; larger jurisdictions (over 100,000) were more likely to use these tools than smaller ones.^[9] A 1996 International City/County Management Association (ICMA) survey of managers in council-manager cities also found larger cities more likely to use participation, but the overall levels were not high. Only 18% of respondents hold community meetings for budget development input, and 32% receive formal recommendations from citizen groups during city council consideration of the budget. The most-used methods in these cities were those that simply provide information to the public, rather than seeking input.^[10]

A study based on interviews with 28 midwestern city budget directors found that only 21% have used participation mechanisms during development of the budget proposal. Approximately one-third use input from surveys or strategic planning sessions that are not directly budget related, whereas another one-third allow input for allocation of specific funds, such as the Community Development Block Grant (CDBG). Many of these cities focus their efforts on improving communication of the proposed budget to the public, rather than encouraging citizen input.^[11]

A citizen participation survey of chief administrative officers in cities with populations over 50,000 found that 96.9% hold public hearings, 81.6% use citizen advisory boards, 52.9% use citizen surveys, and

56.5% use citizen focus groups. When asked about participation in specific functions, 46.2% reported active involvement by citizens in the budgeting function (as compared with 93.9% for zoning and planning, 87.1% for parks and recreation, and 72.5% for policing and public safety). Again, larger cities reported more participation.^[12]

Overall, then, there is some movement toward increasing citizen input in the budget process, but it is stronger in larger cities than smaller jurisdictions. Also, much of the participation effort remains at the end of the process, rather than during budget development, and with attempts to improve communication and education of citizens.

EFFECTS OF PARTICIPATION

Measuring the usage of participatory mechanisms is easier than determining the actual effects of citizen input. Empirical evidence, beyond anecdotes and individual perceptions of effectiveness, is limited. Survey respondents believe that citizen input affects budget decisions; in the earlier mentioned study of budget directors, 77% of respondents agreed that input definitely influenced decisions.^[11] Citizen survey priorities also appeared to affect capital spending decisions in Auburn, Alabama.^[2] The Eugene, Oregon, city council acted on the budget strategies identified through their deliberative process.^[4] A case study of participation in two cities, however, found little evidence of decision-making effects.^[13] We cannot conclusively state the effects of participation on budget decisions, though, without knowing what would have happened in the absence of citizen input.

Decision making is not the only possible benefit from participation. "For administrators, those benefits can include better information on service needs, more services for the dollar, and better feelings about their work. For citizens, the gains can include services better suited to their needs, a more accessible and responsive public bureaucracy, and more positive feelings about themselves and government."^[3]

Much of the work in this area stresses the advantages of educating the public in budget complexities and providing people an opportunity to be heard.^[11] Information may affect citizen perceptions. For example, the city of Dayton, Ohio, uses community boards to improve neighborhoods; with their support, the city has not lost a tax election in 20 years.^[14] In studies of the Eugene case, the amount of budgetary information provided to citizens, as well as the citizen's use of services, affected their revealed preferences.^[4]

This increased openness may increase public trust in and support for government. Wang found that managers perceived the use of general participatory

mechanisms to be associated with increased public trust in government, although this was not the case for specific functions such as budgeting.^[12] In addition, input can provide officials with information on where they need to better market and communicate their proposals. For example, in the Auburn surveys, the areas considered to be high priorities to city council were not seen this way by the public, so the council improved publicity and communication regarding the importance of these projects.^[2]

OBSTACLES

The literature centers around three primary sources of obstacles to citizen participation. The first is the public itself. There are various “publics,” not one homogeneous whole; some people want to be active, for different reasons, whereas others do not, and still others could possibly be reached with some effort.^[5] Lack of knowledge and time, apathy, and perceptions that their opinions are unwanted often lead to low and non-representative participation levels.^[11]

Second, structuring budget participation to be meaningful and effective is difficult. As noted earlier, different methods have distinct advantages and disadvantages; public hearings that are accessible to everyone often are less useful for two-way communications and for educating citizens than are advisory committees or prioritization exercises with smaller groups of people. Timing is also an issue. Opportunities for budget input are often placed at the end of the process, although it would presumably have more of a chance to affect decisions if it occurred earlier. Finally, some mechanisms are very costly, requiring considerable effort on the part of both citizens and public officials.^[6]

Third, public administrators and elected officials may resist increased citizen input. Expert professional administrators may feel better able to make decisions on complex issues.^[3] Research has found that finance officers are interested in participation, but that they would rather have it be initiated by elected officials than administrators.^[5] Some elected officials, though, believe that they already have ample ways to find out what citizens want and believe that formalized group mechanisms are unnecessary given the increased access that citizens have to individual elected officials through e-mail, etc.^[15]

EXIT, VOICE, AND LOYALTY

Albert O. Hirschman, in his book *Exit, Voice, and Loyalty*,^[16] suggested that members of an organization (or customers of a firm) can use some combination of

these three mechanisms to respond to declines in product/service quality. In some cases, this decline may arouse primarily voice from the members/customers, while the organization may actually be more sensitive to exit (or vice versa), so will not take appropriate actions to halt the decline. Therefore, institutional mechanisms are important to have greater congruency between the member's response and the organization's sensitivity. In the current case, then, giving citizens greater opportunities to express their “voice” by participation in the budget process can help to provide valuable signals to governments regarding organizational performance, which may both reduce the level of outmigration from the jurisdiction and allow the government to “fix” the problems. If people believe that voice will be effective, they may postpone exit, especially if they have loyalty to the organization.

As noted earlier, serious concerns in this area revolve around the frequently low levels of participation and the potential that participants are not representative of the community. In some sense, Hirschman suggested, that may be positive rather than negative. It is better to have a mix of “alert” and “inert” customers/organizational members. In this way, sufficient numbers of citizens use voice to let the organization know that there are quality concerns, but the quiet public gives the organization time to address the issues before massive exit. Also, with limited use of voice, citizens retain “reserves” of resources to be used in extreme cases when they are most needed. In addition, in some cases, the “captive customers” for whom exit is the most difficult may actually have the greatest incentive to use voice to get the organization to pay attention.

Hirschman also discussed potential difficulties with the voice option. Voice can be overdone, to the point where protests may actually hinder rather than help organizational operations. In addition, the most quality-conscious constituents, rather than the “captive customers” may be more likely to use voice rather than exit in the short term. This can lead to class cleavages and more organizational sensitivity to the elite. If, however, the organization listens too much to the “captive customer” with loud voice and ignores individuals who are exiting as a result of dissatisfaction, the organization may face greater problems in the long run. The issue of representation, then, is an important one.

Hirschman's work sheds valuable light on the relationship between quality and price, and the potential response of customers/citizens to quality decline. However, determining the quality of government services is complex. The correlation between an individual's perception of government performance and their willingness to pay for services has been found to be strong for many people. However, there are also groups of individuals who are either willing to pay even though they believe performance to be low, or

vice versa.^[17] In addition to the value of citizens using voice to signal dissatisfaction, governmental officials may be able to use their own voice to provide more information to citizens regarding the true cost of services, and can use performance data to convince some that government performance is better than commonly perceived. Hence, the efforts by many jurisdictions to increase opportunities to educate citizens in budget issues.

The greatest challenge in applying Hirschman's thesis is that he concludes that it is difficult, if not impossible, to find a stable optimal mix of exit and voice. "Once members have a slight preference for, say, voice over exit, a cumulative movement sets in which makes exit look ever less attractive and more inconceivable. As a result, voice will be increasingly relied on by members at a time when management is working hard to make itself less vulnerable to it."^[16] For this reason, organizations may have to go through cycles with exit and voice alternating as dominant mechanisms. This suggests the difficulty in institutionalizing budget participation mechanisms that will be optimally effective in the long run, because officials will become less sensitive to voice over time and may eventually simply pay lip service to citizen demands.

CONCLUSION

The citizen participation literature has focused on the voice option, without specifically addressing the role of citizen exit from the community. To what extent do citizens leave a community due to dissatisfaction with service quality or taxation levels? Does citizen participation actually reduce or delay the degree of exit? These are questions that require future research.

Efforts to enhance citizen input will undoubtedly continue to gain momentum, and we will hopefully accrue more hard evidence of the effects of participation. Mechanisms can be designed to provide congruence between citizens' need for voice and government officials' sensitivity to this voice. However, we also need to heed Hirschman's warning that officials will not be sensitive to voice forever, and the pendulum may need to shift toward a greater sensitivity to citizen exit from the community at some point.

REFERENCES

1. King, C.S.; Feltey, K.; Susel, B.O. The question of participation: toward authentic public participation in public administration. *Public Adm. Rev.* **1998**, *58* (4), 317–326.
2. Watson, D.; Juster, R.; Johnson, G. Institutionalized use of citizen surveys in the budgetary and policy-making processes: a small city case study. *Public Adm. Rev.* **1991**, *51* (3), 232–239.
3. Thomas, J.C. *Public Participation in Public Decisions*; Jossey-Bass Publishers: San Francisco, 1995.
4. Simonsen, W.; Robbins, M.D. *Citizen Participation in Resource Allocation*; Westview Press: Boulder, CO, 2000.
5. Miller, G.J.; Evers, L. Budgeting structures and citizen participation. *J. Pub. Budg. Account. Financ. Manag.* **2002**, *14* (2), 233–272.
6. Roberts, N. Public deliberation: an alternative approach to crafting policy and setting direction. *Public Adm. Rev.* **1997**, *57* (2), 124–132.
7. Callahan, K. The utilization and effectiveness of citizen advisory committees in the budget process of local governments. *J. Pub. Budg. Account. Financ. Manag.* **2002**, *14* (2), 259–319.
8. Johnson, E.R. Recommended budget practices: incorporating stakeholders into the process. *Gov. Finance Rev.* **1998**, *14* (4), 15–18.
9. O'Toole, D.E.; Marshall, J.; Grewe, T. Current local government budgeting practices. *Gov. Finance Rev.* **1996**, *12* (6), 25–29.
10. Ebdon, C. The relationship between citizen involvement in the budget process and city structure and culture. *Public Prod. Manage. Rev.* **2000**, *23* (3), 383–393.
11. Ebdon, C. Beyond the public hearing: citizen participation in the local government budget process. *J. Pub. Budg. Account. Financ. Manag.* **2002**, *14* (2), 273–394.
12. Wang, X. Assessing public participation in U.S. cities. *Public Perform. Manage. Rev.* **2001**, *24* (4), 322–336.
13. Ebdon, C.; Franklin, A. Searching for a role for citizens in the budget process. **2001**, Working paper.
14. Gurwitt, R. A government that runs on citizen power. *Governing* **1992**, *6*, 48–54 (December).
15. Franklin, A.; Ebdon, C. Aligning priorities in local budgeting processes. *J. Pub. Account. Financ. Manag.*, forthcoming.
16. Hirschman, A.O. *Exit, Voice, and Loyalty*; Harvard University Press: Cambridge, MA, 1970; 125 pp.
17. Glaser, M.A.; Hildreth, W.B. Service delivery satisfaction and willingness to pay taxes. *Public Prod. Manage. Rev.* **1999**, *23* (1), 48–67.

City Government

Jon C. Teaford

Department of History, Purdue University, West Lafayette, Indiana, U.S.A.

INTRODUCTION

City government has adapted markedly to the changing demands of exploding urban populations. Throughout U.S. history there has been a strong deference to grassroots authority, and this has ensured local governments a major role in public policy developments. In the nineteenth century, city leaders offered a growing list of services to urban dwellers, while facing complaints about corruption. Reform-minded citizens of the late nineteenth and early twentieth centuries revised the structures of local rule and invented new schemes of municipal government. Suburbanization in the twentieth century produced a fragmentation of local authority, as metropolitan residents migrated to myriad suburban municipalities, and thereby avoided the taxes and social problems of older central cities.

HISTORY

Since the first European settlement of North America, there have been city governments. The earliest municipal corporations were patterned after European models. During the Middle Ages and early modern era, European cities acquired charters from the crown that granted them a degree of self-government and certain immunities from royal interference, as well as valuable commercial privileges. These European municipal corporations largely existed to promote and regulate trade, and to protect the interests of artisans and merchants. Similarly, the first North American city governments operated under royal charters that authorized them to regulate local trade activities and bestowed certain lucrative privileges on the municipal corporations. The corporation of Albany, New York, enjoyed a monopoly on the fur trade, and New York City was granted ownership of all tidal lands surrounding Manhattan Island and profited from a monopoly on local ferry service.^[1]

During the eighteenth century, city functions expanded as fire protection, street lighting, and the regulation of such nuisances as stray livestock assumed new importance. Moreover, the American Revolution shifted responsibility for the granting of charters from the crown to the state legislatures. During the nineteenth century, the legislatures granted virtually any

petition for a municipal charter and eventually adopted general incorporation laws to free themselves from the burden of enacting hundreds of charters. Municipal incorporation became a right rather than a privilege. Any locality of a few hundred residents could establish itself as a municipal corporation.

This pattern of permissive incorporation reflected the emerging American deference to grassroots rule. Generally, if local voters wanted to incorporate or annex their property to an existing city, it was their business and the state governments would allow them to determine their own destinies. State legislatures did interfere with city governments, and sometimes ignored the petitions of mayors and city councils. But interference usually arose at the behest of some local faction seeking a favor; it did not necessarily express the state's interest in systematic central control. If a locality requested a charter amendment or authorization to exercise a new power and there was no organized opposition at the local level to the request, then the legislature would generally approve the petition. In the United States, local governments were means by which the people of a locality could express and fulfill their needs. In contrast, throughout most of the world, local government was a means by which the central authorities could administer localities, and ensure the order and safety of the populace. This contrast between top-down authority and grassroots rule continued to differentiate American city government from local government elsewhere in the world.

During the nineteenth century, the largest cities in America and Europe expanded their functions and developed bureaucracies. In 1829, the creation of a uniformed, full-time police force in London, supplanting the earlier constable-night watch system, set an example for cities throughout the English-speaking world. Between 1840 and 1870, the major American cities established professional, full-time police and fire-fighting forces, the largest of which included thousands of men by the close of the nineteenth century. The growing cities also constructed great water supply systems and expansive sewerage networks. Moreover, by the end of the nineteenth century, New York City, Philadelphia, and Chicago could each boast of thousands of acres of public parkland rivaling the great urban preserves of London and Paris.^[2]

Accompanying these developments in the United States was a growing chorus of complaints about corruption and incompetence in city government. The native-born elites resented the rise of immigrant voters and politicians in the nineteenth-century city, and they were ever ready to expose any malfeasance of an Irish politician or police officer. Although city government provided unprecedented services, the newspapers and native-born commentators could point to some notorious examples of corruption in their indictments of city government. Thus, by the close of the nineteenth century, American city government was widely criticized and deemed a failure by many influential observers.

REFORM

The prevailing late nineteenth-century American obsession with corruption in city government was to have a profound influence on twentieth-century municipal development. Demands for administrative reform dominated the municipal dialogue of the early twentieth century, and there was persistent tinkering with the structure of city rule to achieve the honesty and efficiency, supposedly lacking in the government of America's great cities. The goals of honesty and efficiency were to determine the reform agenda of the twentieth century.^[3]

Beginning in the 1880s, civil service reform was a favorite of American good-government advocates, and gradually an increasing proportion of city employees enjoyed the protection of a civil service system. Consequently, the patronage power of city politicians declined, and demands for professionalism in government eroded the clout of party leaders. This was especially evident in police departments. Whereas nineteenth-century police forces were often tools of the politicians, and appointment and promotion were political favors, in the twentieth century, police reformers sought to cleanse departments of political influence and ensure a nonpartisan professionalism.^[4]

Meanwhile, reformers sought to limit the power of neighborhood-based ethnic politicians by curbing the authority of ward-elected city councils and enhancing the authority of the mayors. This contrasted markedly with developments in Canada and Great Britain, where city council committees remained the dominant policy-making element of city government and the mayor was largely a ceremonial figurehead. Suspicion of plebeian, neighborhood representatives led to a strong mayor structure in the United States with the executive, elected by the city's voters at large, exercising veto authority and broad appointive powers.^[5]

Another reform intended to ensure honesty and efficiency in American municipal government was the adoption of at-large election of city council members.

To eliminate ward representation and destroy the power of the supposedly venal ward politicians, municipal reformers advocated that all elected officials be chosen by the city electorate at large. Consequently, an increasing number of cities abandoned ward representation, or at least provided for mixed councils with some members chosen by neighborhoods and others by the citywide electorate.

To further improve city administration, Americans experimented with two new forms of rule: commission government and city manager rule. Originating in Galveston, Texas, in 1901, the commission plan placed all executive and legislative authority in the hands of a small board of commissioners elected at large.^[6] During the 1910s, the city manager plan superseded the commission scheme on the reform agenda. Under manager government, the city council remained in charge of basic policy making, but implementation of that policy and the administration of city services was the responsibility of the city manager. This manager was ideally a nonpartisan professional appointed by the council to ensure expert administration. The early city managers were often engineers with a devotion to efficiency and an expert knowledge of public works.^[7] Whereas commission government lost favor, the city manager plan was adopted in municipalities throughout the United States and in Canada as well. By the close of the twentieth century, more American municipalities operated under the manager plan than under the mayor-council scheme.^[8]

During the late nineteenth and early twentieth centuries, the tradition of deference to the locality was reinforced through the adoption of home rule measures. Viewing the state legislature as a channel through which corrupt local interests could obtain their selfish ends, good-government reformers in the United States sought to close this avenue to nefarious gain. Home rule provisions in state constitutions forbid state lawmakers to legislate on questions solely of local concern, and authorized cities to draft their own charters and charter amendments that would take effect subsequent to approval by the city electorate.^[9]

LATE TWENTIETH-CENTURY DEVELOPMENTS

During the second-half of the twentieth century, the reform mentality remained influential in the United States, although some questioned the merits of the structural and administrative changes of earlier years. The number of American cities adopting manager rule continued to increase. The new breed of managers often were graduates of university city management programs, not simply engineers with some background in the planning and administration of public works. Moreover, an increasing number of mayor-council

cities were hiring chief administrative officers to assume charge of municipal administration without compromising the strong mayor's policy-making or appointive powers. In other words, these officers provided administrative expertise, but unlike city manager municipalities, the mayor was not reduced to a ceremonial figure, presiding over council meetings and cutting ribbons.

Not all elements of the reform agenda survived unscathed. At-large election of city councils came under attack, especially because it seemed to deprive ethnic minorities—who dominated a single neighborhood but were outnumbered citywide—of their fair share of council representation. Thus, there was some shift back to district representation, but at-large elections survived in many municipalities.

Meanwhile, the municipal bureaucracy became more entrenched, and police forces grew increasingly hostile to any civilian interference that threatened their professional independence. Amid the racial disorder of the 1960s, many observers called for civilian review boards to judge charges of police brutality. But such reforms made little headway, and elected officials often seemed impotent when faced with resistance from the city bureaucracy. Militant action by municipal employee unions in the 1960s and 1970s made this quite evident. Despite tight municipal budgets and citizen complaints about city employees, mayors and councils often had to concede cash and favors to the powerful municipal employee unions.

City leaders in the United States did, however, enjoy some new sources of revenue that helped to alleviate money worries. Before World War II, city governments depended overwhelmingly on property taxes for their revenues. Since the 1940s, state legislatures have granted cities the authority to levy a range of new taxes, weaning them from their overwhelming dependence on the property tax. In 1939, Philadelphia became the first American city to levy a municipal income tax, and some other large cities were to follow Philadelphia's example in the following decades. Similarly, municipal sales taxes proved a boon to many cities. In addition, during the 1960s and 1970s, federal aid to cities rose as Washington financed a larger share of the local government burden. Beginning in the late 1970s, however, the federal share began to decline, and Uncle Sam's short-lived interest in city government waned.^[10] The federal government's program of general revenue sharing with municipalities lasted only 14 years, from 1972 to 1986. Federal budget woes and declining political interest in a perceived urban crisis doomed prospects for federal munificence.

The aging central cities of the Midwest and the Northeast especially needed additional funding. Ringed by independent suburban municipalities, these cities could not annex additional territory. As well-to-do

residents and lucrative businesses moved beyond their boundaries, these cities watched their tax bases erode. Metropolitan reformers claimed that the answer to this problem was the creation of some form of metropolitan government that would distribute the burdens of governance more equitably throughout the urban region. Other nations had experimented with metropolitan governments. As early as 1888, the British parliament created the London County Council to assume certain areawide functions in the London region without destroying the many existing borough and city governments that continued to exercise local responsibilities. In 1953, the Ontario parliament established a metropolitan structure for Toronto, comprising the central-city municipality and surrounding suburban municipalities. The municipalities performed specified local functions, while the overarching metropolitan government was allocated regional duties. Political scientists praised the Toronto scheme as a model for American metropolitan reform, but in the United States overarching metropolitan bodies with authority over a range of functions were not a realistic option. In Great Britain, Canada, and much of the rest of the world, the central government could impose reform schemes on localities without winning prior approval from the voters in those localities. No popular referendum preceded the adoption of metropolitan government in Toronto. In the United States, however, the long-standing deference to grassroots opinion generally prevented the states from forcing their will on the localities and supplanting existing patterns of fragmented rule with metropolitan federations abridging the autonomy of area municipalities. Special-purpose metropolitan authorities did develop in the United States. Existing city governments were willing to abdicate responsibility for a single function such as water supply or sewerage, but Americans did not embrace metropolitan behemoths designed to largely supersede city rule.^[11]

CONCLUSION

At the close of the twentieth century, then, the American tradition of deference to local autonomy survived in marked contrast to the pattern of central control over local government found elsewhere in the world. In the United States, Canada, and Europe, the forces of professional city administration remained unabated. This was evident in the continuing popularity of the city manager plan in the United States and the increased reliance on chief administrative officers in American mayor-council cities and in European municipalities as well. Given rising concern about such regional problems as urban sprawl and

wasteful land-use patterns, it remains unclear to what extent American city governments will retain their autonomy, or whether city governments will yield power to regional authorities or confront greater interference from state lawmakers.

REFERENCES

1. Teaford, J.C. *The Municipal Revolution in America: Origins of Modern Urban Government 1650–1825*; University of Chicago Press: Chicago, 1975.
2. Teaford, J.C. *The Unheralded Triumph: City Government in America, 1870–1900*; Johns Hopkins University Press: Baltimore, 1984.
3. Schiesl, M.J. *The Politics of Efficiency: Municipal Administration and Reform in America: 1880–1920*; University of California Press: Berkeley, CA, 1977.
4. Fogelson, R.M. *Big-City Police*; Harvard University Press: Cambridge, MA, 1977.
5. Griffith, E.S. *The Modern Development of City Government in the United Kingdom and the United States*; Oxford University Press: London, 1927.
6. Rice, B.R. *Progressive Cities: The Commission Government Movement in America, 1901–1920*; University of Texas Press: Austin, 1977.
7. Stillman, R.J. *The Rise of the City Manager: A Public Professional in Local Government*; University of New Mexico Press: Albuquerque, 1974.
8. Renner, T.; Desantis, V.S. Municipal form of government: issues and trends. *Munic. Year Book* **1998**, *65*, 30–41.
9. McBain, H.L. *The Law and Practice of Municipal Home Rule*; Columbia University Press: New York, 1916.
10. Teaford, J.C. *The Rough Road to Renaissance: Urban Revitalization in America, 1940–1985*; Johns Hopkins University Press: Baltimore, 1990.
11. Hamilton, D.K. *Governing Metropolitan Areas: Response to Growth and Change*; Garland Publishing: New York, 1999.

Civic Duty

Vera Vogelsang-Coombs

*Maxine Goodman Levin College of Urban Affairs, Cleveland State University,
Cleveland, Ohio, U.S.A.*

Larry Bakken

School of Law, Hamline University, St. Paul, Minnesota, U.S.A.

INTRODUCTION

This article discusses the civic duty of public administrators, based on legal, ethical, and practical interpretations of democratic citizenship. Civic duty refers to the rights, obligations, feelings, and habits of a citizen. A citizen has membership in a democratic state. A state is an independent country with defined territory. A state compels obedience from all who reside in its geographic domain because its sovereign commands absolute power internally. The sovereign's institutions promote a country's permanent interests—law and order, social stability, long-term prosperity, and national security. The relationship between a state and citizens is two-way. The state provides protection to citizens. In exchange, citizens accept the duty to obey the sovereign, to fulfill the requirements of citizenship, and to defend the state from external aggression. Adult residents who are not citizens do not enjoy the full protection or the rights therein of a state.

A democratic state is one in which the “demos”—the people collectively—are the sovereign; its defining characteristic is self-governance. In America, the founders vested the new country's sovereignty in the citizenry. The founders believed that citizens, as differentiated from a king's subjects, could govern themselves, because they could regulate themselves (Ref. ^[1], pp. xvii–xviii). Citizens regulate themselves when they are faithful to the duties associated with the status of citizenship. Citizenship takes two forms: legal citizenship and ethical citizenship; ethical citizenship has two subclasses—national citizenship and democratic citizenship. The meaning of civic duty is dynamic because the American state embeds the norms of citizenship in changing political institutions, legal rules, and democratic culture. American democracy involves citizens in the roles of private individual, political participant, and officeholders. Officeholders can have three identities: 1) political representatives, i.e., elected officials or lawmakers; 2) public administrators, i.e., bureaucrats or the civil service; and 3) and independent judges. In this analysis, we will limit our discussion to

citizenship and the civic duties of the public administrator. The following discussion explains each row of [Table 1](#) in turn.

LEGAL CITIZENSHIP AND LIMITED GOVERNMENT

Democratic government sets the foundation for legal citizenship. American government is a form of liberal democracy. Liberal democracy rests on the principle of limited government. This principle assumes that the citizens voluntarily surrender part of their absolute power as sovereign to government so that government can make decisions for the state. In exchange for the people's grant of some sovereign power, liberal democracy functions according to a constitution or charter. In America, the Constitution is written and sets the reciprocal relationship between individual citizens and democratic government. Besides enumerating the powers that the people have surrendered to government, the Constitution also lists the restraints on government. Its significance is that it is the supreme law of the land. All who reside in America are subject to it, including the popularly elected president, the head of state. The U.S. Constitution uses three other legal devices to protect citizens from government. First, it establishes a federal republic with dual sovereigns. Federalism means that two distinct governments exist that derive their powers directly and independently from the sovereign people: the national government and states. The Constitution assures a republican form of government in every state and at the national level. Citizens of one state enjoy the privileges of citizenship in every other state. It also guarantees to protect states from external invasions and domestic violence. Second, the Constitution secures citizens' civil rights and liberties. The U.S. Bill of Rights, constitutional amendments, and federal laws express these protections. Liberties, such as free speech, are inalienable and are immune from the arbitrary exercise of governmental power. Neither government nor a citizen can repudiate them or transfer them to another person.

Table 1 Citizenship, liberal democracy, and civic duty

Types of citizenship	Characteristics of liberal democracy	Norms for citizenship	The civic duty of public administrators
Legal citizenship	Limited government	Personal responsibility	The higher standard of conduct
National citizenship (ethical citizenship)	Good government	Caring for public interests	Office as public service
Democratic citizenship (ethical citizenship)	Potential responsiveness	The civic culture	Citizenship as office

Third, the Constitution authorizes the principle of reserved powers. Amendment X gives to the states or the people the powers that the Constitution has not explicitly granted to the national government or withheld from the states.

Pluralism and Civil Society

Liberal democracy presupposes the existence of a pluralist civil society. A pluralistic society divides a democratic state's power into open and competing groups instead of concentrating it in a ruling elite. Pluralism creates a civil society. Civil society is broader than government, giving citizens a private (social) life and a political life. Citizens get a unique identity by pursuing personal interests, including family life, in private (non-governmental) associations. The premise is that only citizens can define their interests. Self-interested citizens advance their interests by joining groups of similarly minded individuals. Pluralism assumes that a check on group power occurs as self-interested individuals shift their allegiances from group to group in response to changing perceptions of their own interests.^[2] Citizens' groups align with others in the political arena to advocate for policy preferences and to achieve favorable outcomes. Politics is the mechanism for aggregating and converting individual preferences into collective decisions. In theory, this conversion takes place in the competitive processes of responsible parties and within a set of legal rights and political rules.^[3] Backed by political authority, these collective choices compel obedience.

Civic Duty and Personal Responsibility

Citizens in liberal democracy can assume the identity of a private citizen, a political participant, or an officeholder. All have a duty to be responsible. Personal responsibility for private citizens means that individuals voluntarily restrain their behavior. They choose to restrict themselves not to interfere with others' private lives, liberties, and interests. In exchange, responsible citizens have the right to expect others to do the same. Furthermore, citizens' private lives have

legal protections. Democratic government has a duty to avoid arbitrary intrusions into the citizenry's private affairs and can only do so reluctantly after following due process. Personal responsibility also means that citizens have a civic duty to participate in democratic political processes. Political participation is necessary, because the legitimacy of democratic government depends on it. It allows the citizenry to show consent. Showing consent legitimates democratic government's use of the state's power. Citizens show consent indirectly by voting for political representatives who, in turn, conduct the "people's business." Voting depends on three principles. First, citizens stand as political equals. Second, liberal democracy holds open and free elections at regularly scheduled intervals. Third, democratic government has an obligation to count each citizen's vote. Personal responsibility for elected officials means that they encourage pluralistic participation, grant access to citizen groups, broker exchanges among competing interests, and enforce contracts. Political representatives grant the right to exercise limited discretion to public administrators so that democratic government can listen and respond to legitimate group interests. Elected officials control administrative discretion by organizing government agencies into hierarchies and through legislative oversight. The electorate has the power to replace democratic government if the voters perceive officeholders as unresponsive or even to change the Constitution.

The Higher Standard of Conduct

In democratic theory, the position of a free and autonomous citizen is the most esteemed identity. Citizens who serve in government must surrender more rights and democratic freedoms to government. The democratic sovereign restricts the conduct of officeholders because of its distrust of concentrated political authority. Therefore, some actions of private citizens are illegal if done by political representatives or public administrators. The restrictions of office translate into a higher standard of conduct. Legislative codes, conflict of interest statutes, and criminal sanctions

operationalize this higher standard. In exchange for discretionary authority, public administrators must follow the rules of their office. Legal requirements exist for public administrators to carry out their office with political impartiality and professionalism. They have a duty to respect election outcomes, to support democratic processes, and to act as rule-abiding trustees (Ref.^[4], pp. 154–155). Administrators must submit to sanctions if a legislative review finds that they have failed to discharge these duties properly. Impartiality and professionalism counteract the partisanship and ideologies of popularly elected officials. One school of thought frees public administrators from legal restraints, obliging them to disobey evil-minded elected officials in times of crises. If nefarious elected officials capture the state's absolute power and present their unconstitutional activities as those of the democratic sovereign, then dictatorship happens, as in Nazi Germany. According to March and Olsen, public administrators in similar situations must actively thwart the wishes of iniquitous political officials (Ref.^[4], pp. 86–89).

NATIONAL CITIZENSHIP AND GOOD GOVERNMENT

The legal foundations of citizenship are incomplete. Citizens who possess legal and civil rights in a pluralist society do not automatically feel political allegiance to their country.^[5] To stimulate national citizenship, the American founders vested the sovereignty of the new country in the political community, not the electorate. This political community encompasses the nationwide body of citizens as they interact informally and formally in social, religious, economic, cultural, and community associations in their localities.^[6] By vesting sovereignty in the national political community, the founders broadened the idea of citizenship to include an ethical dimension. Ethical citizenship refers to the moral beliefs, values, fraternal feelings, habits, and reciprocal commitments of individuals that guide their lives. Whereas legal citizenship is expressed in laws and rules, ethical citizenship is in standards that most people acknowledge but are not codified. Ethical citizenship is always a matter of voluntary behavior and is subject to different interpretations by different people.

The founders rooted the political community in a culture of “good government.” Political culture refers to the citizenry's attitudes about the style of democratic governmental operations.^[7] Good government, as opposed to perfect government, is based on legitimacy that extends beyond electoral popularity. This means that democratic government must defend its legitimacy on higher (normative) values. It is not

enough for government to appear good; its actions must be good. To make government's actions “good,” officeholders are bound to justify publicly and receive approval from the democratic sovereign (not just the electorate).^[8] In obtaining prior approval, officeholders determine whether their proposed actions are ethically appropriate, i.e., have wide community support. For March and Olsen, the ethic of appropriateness means that government acts consistently with cultural and political norms.^[4] This ethic reassures the democratic political community that officeholders will not act arbitrarily nor for purely private interests. The actions of good government strengthen the democratic sovereign's confidence in those to whom it has entrusted the state's power.

Civic Duty and Caring for Public Interests

Membership in the political community gives citizens a public life besides a private life. In public life, citizens actively maintain public purposes, based on a perception of national solidarity and a sense of a common destiny. For Norton Long, a precondition of legal citizenship is that citizens have a sense of “*moi commun*.”^[9] Communal feelings stir in citizens, he says, if “a public thing” quite literally exists for which they feel responsible. For John Dewey, this public thing emerges when citizens rediscover the “public and its interests.”^[10] A political community of citizens, he says, cares for public interests. To care for public interests, citizens must engage each other to discern common needs. This task requires citizens to reflect and to distinguish between the special case (their own interests) and the general case (the interests of others). In other words, citizens use moral reasoning to perceive the difference between public interests and private interests. According to Dewey, people create private interests when they join voluntarily, and only those people involved in a transaction experience the direct or indirect consequences of those interactions. Public interests are not fixed. They emerge when individuals as a group are seriously and indirectly harmed by the consequences of transactions in which they have not engaged. Although citizens may disagree, their open discussions about public interests allow the political community to search for remedies.

Caring for the public gives diverse citizens a common ethical identity. Public spirit awakens this identity in citizens. The civic virtue of private citizens is the source of public spirit. Public-spirited citizens participate in public life when they perceive that the personal advantage of each member of the community is consistent with working for the good of all. Alexis de Tocqueville called this disposition of American citizens “the principle of self interest rightly understood”

(Ref.^[11], pp. 129–132). Citizens live this principle daily in small acts of self-restraint, not through great sacrifices. Such daily actions, said de Tocqueville, gradually draw individuals through their habits in the direction of the national political community. Virtuous citizens, acting on this principle, voluntarily help others. In place of doing it themselves, they willingly give up additional personal resources to sponsor community programs aimed at serving the poor and at improving society. De Tocqueville said that this public role for private citizens is based on custom, not law. This custom binds together the nation, allowing diverse citizens in the political community to lead a good life. For Dewey, public life is a good life. A good or high-quality common life emerges only if a democratic state is a good state. A good state comes into existence when trustees of the public (lawmakers and public administrators) systematically care for public interests. In the good state, citizens have greater liberty and security in their private affairs because they are relieved from the waste of negative struggle and needless conflict (Ref.^[10], pp. 71–72). Its community-based processes, by promoting reflection, moral reasoning, and open discourse, civilize humanity, allowing citizens to show empathy, tolerance, and generosity (Ref.^[4], p. 60).

Office as Public Service

For Dewey, the public is vital to the creation of government. Government is the public organized to carry out the functions of the state in a way that strengthens bonds among citizens. The sovereign political community connects the power of the democratic state to care for public interests through government officials. It gives them the use of the democratic state's power to regulate the conjoint actions of individuals and groups that cause a public interest to emerge. Different issues create different public interests. As trustees of the public and its interests, public administrators hold the office of a public servant. The duty of public servants is to recognize the dynamics and the consequences of individuals "acting as a public." According to March and Olsen, the political community's discussions about public interests should guide and restrain public servants' actions. By listening to these discussions, public servants will get an idea about what governmental actions are acceptable to the larger political community on moral grounds. These discussions reveal the issues that government should avoid or remove from the policy agenda because of a lack of wide acceptance. By withdrawing these issues temporarily, public servants help to prevent the polarization of the community. Moreover, the influence of community discussions on public servants allows the democratic sovereign to civilize government.

Ethical Maturity

This task of public service requires ethically mature officeholders. Ethically mature public servants, said Stephen K. Bailey, appreciate the reality of self-interest and its impact on motivating behavior, both public and private:

... Public service is the capacity to harness private and personal interests to public interest causes. Those who will not traffic in personal and private interests (if such interests are within the law) to the point of engaging their support on behalf of causes in which both public and private interests are served are, in terms of moral temperament, unfit for public responsibility (Ref.^[12], p. 47).

Ethically mature public servants, said Bailey, possess three moral qualities: optimism, courage, and fairness tempered by charity. Optimism allows public servants to see possibilities in shifting political priorities so that they can perform the functions of government without becoming cynical. Optimistic public servants are motivated to better society and to improve the quality of human life. Cynical officials who perceive humankind as unworthy of such efforts are the ultimate source of corruption in a free society. Courage overcomes the forces of private relations, especially friendship, and provides a degree of neutrality to override requests for special favors. Ethical courage means that public servants are willing to decide even with inadequate information. Ethical courage is meaningless unless it enables public servants to act justly. Fairness tempered by the virtue of charity promotes justice. Charitable public servants discipline themselves by controlling persistent inner claims for personal recognition, power, and status. Charity is also a source of visions for the "good society" without which government, said Bailey, becomes "a sullen defense of existing patterns of privilege" (Ref.^[12], p. 52). Fairness compensates for the subjective nature of all decision making. Fairness is necessary because of administrative discretion. Fair-minded public servants control their discretion by balancing flexibility and consistency in their decisions. As Bailey pointed out, public administrators who never deviate are rigid and lack compassion; those who deviate too much are subversive. The ethical compromises of mature administrators allow them to act with enough consistency and flexibility to survive and succeed in public service.

CHALLENGES TO NATIONAL CITIZENSHIP

At the end of the twentieth century, scholars have identified at least five challenges to national citizenship

in America. First, Robert Putnam argued that American life lacks communal solidarity. Instead of a “nation of joiners” with a vibrant public life, Americans are “bowling alone,” and their civic life has all but disappeared (Refs.^[11], pp. 114–118,^[13]). Second, scholars have identified a societal malaise: widespread alienation of citizens from government. For example, Evan Berman’s national survey identified three sources of citizen discontent: 1) government officials whom citizens perceive as indifferent or hostile to their needs; 2) poor service delivery; and 3) governmental processes that exclude citizens.^[14] Third, the emergence of the welfare state has significantly undermined the citizen–government relationship. Specifically, Don Eberly said that Americans have assumed problematic identities as clients, customers, consumers, and claimants of government:

The provision of public assistance to tens of millions of Americans is often accompanied by the subtle message that recipients are hopelessly trapped in conditions that require the permanent help of advocates, interest groups and government workers. It produces a mindset that dwarfs citizenship for the poor and the non-poor alike. It leaves the poor feeling justified in doing little to reclaim control of their lives and the non-poor feeling no obligation to intervene with neighborly aid (Ref.^[1], p. xxix).

Fourth, in a comprehensive study of citizenship, Rogers Smith found that liberal democracy is not the core meaning of the American civic identity.^[15] Clashing traditions and popular politics have excluded most Americans from enjoying full citizenship. Fifth, Chester Newland argued that a pervasive system of spoils at the national level excludes most of the public administration community from important matters of governance (Ref.^[16], pp. 45–56). This exclusion undermines the capacity of American democracy to function adequately or to achieve social change. Some fixes to these problems of national citizenship, such as Benjamin Barber’s “strong (grassroots) democracy,” aim at weakening liberal democratic institutions and basic protections.^[17] A more constructive approach is not to deny democratic liberal traditions. It draws the norms of ethical citizenship from the operations of successful liberal democracies. If democracy is to endure, then pragmatic realism must temper the idealistic notions of citizenship.^[18]

DEMOCRATIC CITIZENSHIP AND POTENTIAL RESPONSIVENESS

E.E. Schattschneider provided a realist’s view of democratic citizenship. The most damage to the democratic

cause comes from idealists who attribute a “mystical, magical omnipotence” to the people (Ref.^[19], pp. 135–136). This attribution takes no heed of what the democratic system can do. Involving 280 million people directly in all aspects of governance is physically impossible. Instead, he assumed that democratic citizens were moral equals: Each person is a precious and unique human. Morally equal citizens have potential involvement. Potentially involved citizens become active in public affairs because of a democratic dynamic: conflict is contagious. Different public issues activate different citizens. If only 0.10% of the citizenry mobilizes around an issue, then the latent force of the remaining citizens is 999 greater than that of the activists. Schattschneider’s point is that any estimate of citizen participation based on the number of original activists is fatuous. Potentially involved citizens are not neutral spectators. Public issues succeed or fail because the original activists mobilize or exclude the spectators. The balance of citizen involvement in any public issue, he said, is not fixed until *everyone* is involved (Ref.^[19], pp. 4–5).

From this analysis, Schattschneider recasted the meaning of citizenship. He made its characteristics consistent with what most citizens can and are willing to do in public life. He said that the role of democratic citizens is the same as in private relationships: People take advantage of what other people know and judge results. In public affairs, citizens rely on government leaders and organizations, or their responsible opposition, to give them choices about policy alternatives. These choices must allow citizens the potential to participate in meaningful ways. In this democratic system, the sovereign people “act through the government.” This system satisfies citizens only if it meets the condition of “potential responsiveness.” Potential responsiveness in a democracy, said Hannah Pitkin, requires citizens to have: “. . . access to power rather than its actual exercise . . . There need not be a constant activity of responding, but there must be a constant condition of responsiveness, of potential readiness [of government] to respond.”^[20] If citizens perceive potential responsiveness in the democratic system, then they feel that they can initiate action or trust leaders to do the right thing. Citizens accept governmental authority despite their lack of awareness of what government is doing most of the time. These beliefs are a part of the democratic civic culture.

The Civic Culture

According to Gabriel Almond and Sidney Verba, the civic culture shapes the norms for citizens and their relationships to each other and to their democratic government.^[21] Their comparative analysis of five

democracies, though drawn in 1963, remains the classical formulation. It showed that stable democracies have democratic civic cultures. They described the American civic culture as mixed. Citizens' activity in public life combines involvement and activity with passivity and non-involvement. They take on and balance the duties of three identities: the influential activist, the passive governmental subject, and the parochial (family-oriented, private) person. Almond and Verba found that the norms of trust and confidence from citizens' social interactions transfer to their political relationships and temper them. They also found that the interactions of citizens and government showed a cyclical pattern. In normal times, most citizens are quiescent. They are uninterested in what decision makers do, but officeholders can act only within legal and ethical bounds. If an issue becomes salient, then citizens will increase their involvement with government, making demands on elected officials and public administrators. If government officials make appropriate adjustments, then the importance of politics will lose saliency for mobilized citizens, and they will withdraw from active participation. This cycle allows a democratic system to change slowly and humanely. If issues become and remain intense, then they destabilize a democracy.

Almond and Verba's research showed that the cycle of political participation strains neither people nor liberal democracy. The reason is that the civic culture controls decision makers without checking them so tightly as to render democratic government ineffective. The check and balance occur for four reasons. First, democratic government officials subscribe to the belief in the influence potential of citizens. Second, government decision makers anticipate citizens' reactions even if citizens are not active or making demands on them. Third, government officials and citizens are part of the same political community. They share the same civic culture that exposes them to the same democratic values. Fourth, public and private life are not opposites in the civic culture. The democratic culture mixes the personal, social, and political identities of citizens. It socializes government officials to uphold a continuum of public-private relationships that flourish legitimately in the political community.

Citizenship as Office

Terry Cooper formulated an ethic of citizenship for democratic public administration. He assumed that citizenship underpins all normative relationships in the political community. Citizenship, in his conceptualization, is the public office of virtuous individuals in the political community. Public administrators are citizens first. They are virtuous citizens "employed

as one of us to work *for* us."^[22] In office, they are fiduciary citizens. Acting as trustees for the political community, virtuous administrators make citizens part of the administrative role. Efficient service delivery and law enforcement are their penultimate obligations. Their ultimate duty is to build and contribute to community life by sharing the sovereign's power, upholding civic values, and pursuing long-term interests. The authority relationships of virtuous administrators are both vertical and horizontal. They look at popularly elected politicians and to the citizenry for direction. For Cooper, the ultimate loyalty of virtuous administrators belongs to the citizenry because the citizenry's authority is that of the sovereign. The sovereign's authority transcends loyalty to specific elected officials or particular governmental agencies.

Virtuous administrators use transformational leadership practices to facilitate (or awaken) the involvement of citizens in public affairs. They combine their professional expertise and civic values in a way that gives citizens new and meaningful opportunities to participate in governance and for self-reflection (Refs.^[16], p. 53 and 55,^[23]). Their practices stimulate positive changes in individuals, institutions, and the community (Ref.^[4], p. 55). "Coproduction," "indirect administration," and "transparent decision making" are three examples of such practices. Coproduction, for Charles Levine, meant that virtuous administrators involve active citizens as partners in allocation decisions regarding service delivery.^[24] For Cooper, virtuous administrators engage in indirect administration by subordinating their technical expertise to active community leaders. They engage in policy deliberations by working through and respecting existing communities and their associations and authorities. Cooper said that administrators are justified to shed their neutrality and to act as policy advocates if their judgments are consistent with their roles as fiduciary citizens. Transparent decision making, said Carol Lewis, establishes the ethical credibility of public administrators.^[25] It is necessary because most administrative decisions take place without the physical presence of citizens. Credible administrators, she said, "go on record." By going on record, they publicly justify their decisions. This practice socializes them to account for their decisions even if citizens will not judge the results for years.^[26] Besides enhancing governance and enriching community life, these practices develop the character of democratic citizens.

Political Socialization

The development of democratic citizenship is a task of political socialization. Political socialization is concerned with the "personal and social origins of political

outlooks.’’^[27] It is the process through which citizens orient themselves to the political community, democratic government, and the civic culture. Political socialization begins early and is lifelong. It takes place in the family, peer groups, public schools, workplaces, and in political processes, although the quality of these experiences will vary. A liberal education is a vehicle for building citizenship in young adults. This broad education allows young people to reflect carefully about the human condition. It encourages them to use this learning to shape their lives as moral citizens and civilized human beings.^[28] Socialization also takes place in university-based Master of Public Administration (MPA) programs and through the associations of professional administrators, such as the American Society for Public Administration (ASPA) and the International City/County Management Association (ICMA). Finally, virtuous administrators socialize citizens. Their transformational practices connect citizens and democratic government. Virtuous administrators teach citizenship, using the experience of the political community.

CONCLUSION

Underpinning this analysis of citizenship and civic duty is the relationship between elected officials and public administrators. It touches on a fundamental political question: What is the proper balance between representative institutions and the civil service in a liberal democracy? Answers to this question are political as much as matters of law, ethics, and best practice. The terms associated with civic duty—the “common good,” “the public interest,” “ethical maturity,” and “the virtuous administrator”—are political terms. Political terms have ambiguous meanings, because no consensus exists on their definitions. This definitional ambiguity accommodates multiple interpretations of civic duty. For example, citizen-centered norms of civic duty prevail in the literature at the end of the twentieth century. Embedded in these norms is the following proposition: The citizen is supreme in governance, despite what legitimate elected officials want public administrators to do. This is an explosive proposition. It implies that the ultimate duty of the public administrator is to enable citizens to realize their view of the public interest. In effect, those professional administrators, through their citizenship identity, are preeminent over elected officials in democratic governance. As a political proposition, one can place this view at an extreme point on a definitional scale of civic duty. Given this conceptual placement, one can see why this interpretation of civic duty is controversial and is likely to elicit reactions. Moreover, political shifts in a changing environment are inevitable, and they will lead people

to challenge the prevailing conventions. For example, before the terrorists attacked America on September 11, 2001, national citizenship was a minor identity for Americans (Refs.^[1], p. xx,^[4] p. 72, also see the section on [Challenges to National Citizenship](#) in this article). Putnam’s “bowling alone” metaphor captures the view that civic duty was irrelevant in American society. After the attacks, national citizenship has become a primary identity for many Americans. In the catastrophe’s wake, the definition of civic duty has tipped to the other extreme. New interpretations have placed the preferences of elected officials as preeminent in American governance. These views contain another controversial political proposition: The ultimate duty of public administrators is to enable elected officials to realize their view of the public (i.e., national) interest. It implies that the public interest is whatever elected officials say it is. Our point is that the question of civic duty, as a political question, is one that a liberal democracy can never resolve once and for all. Inherent in the literature on civic duty are different and successive political interpretations of “who should govern.”

INTERNET RESOURCES

The National Civic League is the oldest civic organization in the country; see www.ncl.org. The National Association of Schools of Public Affairs and Administration is the accrediting body of MPA programs; see www.naspa.org. The American Political Science Association has formed a Civic Education Network; see www.apsanet.org/CENnet. This network has a list of more than 60 civic organizations; see www.apsanet.org/CENnet/organizations. Its list of teaching and research resources is found at www.apsanet.org/CENnet/resources.

REFERENCES

1. Eberly, D.E. The Quest for a Civil Society. In *Building a Community of Citizens: Civil Society in the 21st Century*; Eberly, D.E., Ed.; University Press of America: Lanham, MD, 1994.
2. Shafritz, J.M. Pluralism. In *The HarperCollins Dictionary of American Government and Politics*, 1st Ed.; Harper Collins Publishers: New York, 1993; 354–355.
3. Lowi, T.J. *The End of Liberalism*; W.W. Norton: New York, 1979.
4. March, J.G.; Olsen, J.P. *Democratic Governance*; The Free Press: New York, 1995; 154–155.
5. Heywood, A. *Political Ideas and Concepts, An Introduction*; St. Martin’s Press: New York, 1994; 156.
6. Hult, K.M.; Walcott, C. *Governing Public Organization: Politics, Structures, and Institutional Design*; Brooks/Cole Publishing: Pacific Grove, CA, 1990.

7. Shafritz, J.M. Political Culture. In *The HarperCollins Dictionary of American Government and Politics*, 1st Ed.; Harper Collins Publishers: New York, 1993; 362.
8. Madison, J. Federalist Paper #49. In *The Federalist Papers*; New American Library: Mentor Book, New York, 1788/1961; 313–317.
9. Long, N.E. An Institutional Framework for the Development of Responsible Citizenship. In *The Polity*; Press, C., Ed.; Rand-McNally: Chicago, 1962; 184.
10. Dewey, J. *The Public and Its Problems*; Swallow Press: Chicago, 1927.
11. de Tocqueville, A. *Democracy in America*; Bradley, P., Ed.; Vintage Books–Random House: New York, 1838/1945; vol II; 129–132.
12. Bailey, S.K. Ethics and the Public Service. In *Combating Corruption/Encouraging Ethics*; Richter, W.L., Burke, F., Doig, J.W., Eds.; American Society for Public Administration: Washington, D.C., 1964/1990; 47.
13. Putnam, R. Bowling alone: America's declining social capital. *J. Democr.* **1995**, 6 (1), 65–78.
14. Berman, E. Dealing with cynical citizens. *Public Adm. Rev.* **1997**, 57 (2), 105–112.
15. Smith, R.M. *Civic Ideals: Conflicting Visions of Citizenship in U.S. History*; Yale University Press: New Haven, 1997.
16. Newland, C.A. Public executives: Imperium, sacerdotium, collegium? Bicentennial leadership challenges. *Public Adm. Rev.* **1987**, 47 (1), 45–56.
17. Barber, B. *Strong Democracy: Participatory Politics for a New Age*; University of California Press: Berkeley, 1984.
18. Baldwin, H.W. Warfare and Civilization. In *The Peloponnesian War*; Bantam Books: New York, 1960; 8Thucydides; 4th Printing, Introductory Essay.
19. Schattschneider, E.E. *The Semi-Sovereign People: A Realist's View of Democracy in America*; Reissued Ed.; The Dryden Press: Hinsdale, IL, 1975; 135–136.
20. Pitkin, H.F. *The Concept of Representation*; First Paperback Ed.; UCLA Press: Berkeley, CA, 1972; 233.
21. Almond, G.A.; Verba, S. *The Civic Culture: Political Attitudes and Democracy in Five Nations*; Princeton University Press: Princeton, NJ, 1963; 5.
22. Cooper, T.L. *An Ethic of Citizenship for Public Administration*; Prentice-Hall: Englewood Cliffs, NJ, 1991; 139.
23. Burns, J.M.G. *Leadership*; Harper and Row: New York, 1978; 434.
24. Levine, C.H. Citizenship and Service Delivery: The Promise of Coproduction. In *Democracy, Bureaucracy, and the Study of Administration*; Stivers, C., Ed.; Westview Press: Boulder, CO, 2001; 280–300.
25. Lewis, C.W. *The Ethics Challenge in Public Service: A Problem-Solving Guide*; Jossey-Bass Publishers: San Francisco, CA, 1991; 58.
26. Thompson, D.F. *Political Ethics and Public Office*; Harvard University Press: Cambridge, MA, 1987; 113.
27. Dawson, R.E.; Prewitt, K.; Dawson, K.S. *Political Socialization*, 2nd Ed.; Little, Brown and Co.: Boston, 1977; 1.
28. Gray, H. The Aims of Education. In *The Aims of Education*; Boyer, J.W., Ed.; The University of Chicago Press: Chicago, 1987/1997; 63–79.

Civil Disobedience

Matthew Tedesco

John Harris

Department of Philosophy, University of Colorado at Boulder,
Boulder, Colorado, U.S.A.

INTRODUCTION

An act of civil disobedience involves a violation of law with the intent of effecting a change in current policy, regarded as unjust by the citizens taking action. Important issues involved in civil disobedience include fidelity to the state, publicity of the disobedient act, the permissibility of violence in civil disobedience, and the acceptance of punishment on the part of the protesters. Civil disobedience was first explicitly discussed in detail by Henry David Thoreau, and later refined by activists such as Martin Luther King, Jr. and Mohandas Gandhi, and theorists such as John Rawls and Ronald Dworkin; however, its first important treatment is offered by Socrates, the ancient Greek philosopher, as presented in Plato's *Crito* and *Apology*.

KEY FIGURES AND IDEAS

As a matter of policy, it may appear strange for political theorists to attempt to justify a citizen's refusal to obey the law. One might think that if the government is, for the most part, just, then the citizens under the authority of the government always have a duty to obey the laws. Socrates (469–399 B.C.E.) appears to advocate just this position in the *Crito* when he stated, "...one must not give way or retreat or leave one's post, but both in war and in courts and everywhere else, one must obey the commands of one's city and country, or persuade it as to the nature of justice."^[1] It would appear that Socrates thought that one ought to obey the government regardless of what it commands. This is, in part, because Socrates thought that the laws of the state provided so much for him that he owed a debt of gratitude, and thus, fidelity to the law compelled Socrates to go so far as to give his life if necessary. However, in the *Apology*, Socrates appears to have said something quite different. Here, Socrates proclaimed before the court of Athens that, "...I will obey the gods rather than you, and as long as I draw breath and am able, I shall not cease to practice philosophy..."^[2] This example, taken from antiquity, shows that even Socrates, a man who eventually killed himself because he was commanded

to do so by the laws, thought that there are some laws that one should not obey. Even if we accept, as Socrates appeared to, that there are some laws that should be disobeyed, the question then becomes what differentiates an act of civil disobedience from a simple violation of the law, and when, if ever, is civil disobedience justified.

Modern interest in civil disobedience was rekindled by Henry David Thoreau. His essay "Resistance to Civil Government" (later known as "Civil Disobedience") was first published in 1849,^[3] and it served as an important influence on prominent social and political activists in decades to follow. In it, Thoreau addressed his reasons for refusing to pay a poll tax over the course of several years, an act of defiance that landed him briefly in jail until, to his consternation, an associate paid it for him and procured his release. Thoreau was primarily concerned with the evil of slavery, and specifically, the way in which his tax dollars causally implicated him in it—supporting the government that supported slavery. Thoreau took no issue with the immediate direction of taxes, such as his highway tax, in that paying it made him simply a "good neighbor." Furthermore, while Thoreau strongly disagreed with slavery, he did not think that any person's duty was to be devoted to combating serious wrongs. Crucially, Thoreau considered it every person's duty to make sure that a role, even indirect, in supporting serious wrongs was not played. It is this crucial causal link that led Thoreau to take action—in his case, refusing to pay the tax that indirectly attached him to the wrong.

Thoreau was explicit in his endorsement of breaking the law in actions of civil disobedience; even if these actions come at some substantial cost to the agent, these costs are simply outweighed by the importance of removing oneself from blameworthiness in the wrong being protested. He specifically acknowledged that this cost might include imprisonment or loss of property, along with other harassments. This is not to say, however, that he endorsed acceptance of punishment and, by extension, the legitimacy of government; he simply was at the mercy of something stronger than him.

Thoreau, then, introduced several important components to contemporary understandings of civil

disobedience, including the moral requirement of action, likely illegal, by private citizens, and the acknowledgment of the inevitability of punishment. In general, a key feature of Thoreau's presentation of civil disobedience is that it is in many ways indistinguishable from outright revolution. Thoreau blurs any distinction here by speaking explicitly and approvingly of revolution and "war with the State" within the essay, including the permissibility of bloodshed. This conflation can be contrasted with prominent advocates of civil disobedience in the twentieth century, beginning with Mohandas K. Gandhi.^[4]

Gandhi is perhaps most famous for endorsing and engaging in non-violent civil disobedience (an approach that Thoreau clearly did not advocate). Gandhi was primarily concerned with the British rule of India and sought through civil disobedience to achieve independence for his nation. For Gandhi, civil disobedience was merely a part of the broader movement of what he called *Satyagraha*, often described in English as something like "truth-force."

Gandhi offers two primary, related reasons for preferring non-violence to violence in acts of civil disobedience. The first concerns the nature of the action. That is, violent action, even taken in the service of some supposed greater good, is still by definition violent, and so evil. This evil cannot be eradicated, given the way it is contained within the act, and thus, the outcome of the action, while good, can only be temporary. Because violent action is by definition evil, while non-violent action is untainted by anger and hatred, Gandhi offers a second reason for preferring non-violent action: it is more effective. That is, because the pressure exerted on the state by non-violent civil disobedience is grounded in "goodwill and gentleness," it will more effectively and lastingly accomplish its ends than will violent action.

Another key element of Gandhi's discussion of civil disobedience was the notion of "self-purification" that the protester is required to first undergo. As Gandhi presents it, this self-purification is a deep and profound preparation, whereby the protester acknowledges the nature of the protest and what actions must be taken because of it. Like Thoreau, Gandhi acknowledges the hardships the protester is likely to endure; unlike Thoreau, Gandhi believed that these hardships must be voluntarily accepted. This voluntary acceptance and emphasis on non-violence are two of Gandhi's most important contributions toward a contemporary understanding of civil disobedience.

Martin Luther King, Jr. took up these elements in addressing civil disobedience in relation to the civil rights movement in the United States. In King's "Letter From Birmingham City Jail,"^[5] he echoes many of the key elements of civil disobedience presented by Gandhi before him. In advocating

non-violence as the appropriate means of civil disobedience, King explained that it is right because of its nature as a moral act, and because this moral foundation makes the act more effective in making the issue a focal point of public awareness. Like Gandhi, he advocated acceptance of the punishment, also calling the process "self-purification" and advocating this acceptance via workshops concerning appropriate behavior in civil disobedience. It was this acceptance of the state's reaction to civil disobedience that placed him willingly in jail, from where he wrote this famous letter.

King went further than Gandhi in making explicit the reasoning behind accepting the punishment of the state following civil disobedience. While Thoreau advocated violence and outright revolution in objecting to the state, King acknowledged the narrower focus that distinguished civil disobedience from revolution. That is, for King, it was not the state that needed to be toppled in the face of injustice, but rather the specific elements of law. In just this way, King preserved the notion of fidelity to the state that revolution denied. King mirrored the natural law tradition in drawing this distinction, explicitly citing Augustine's famous line that "an unjust law is no law at all." King took great care to point out the distinction between just and unjust laws, noting that he was solely targeting the latter.

Through Thoreau, Gandhi, and King, then, the notion of civil disobedience was importantly refined. Thoreau's primary concern was addressing the moral duty of the private citizen to be critical of the state through civil disobedience, while acknowledging the reaction that this protest would certainly elicit. Gandhi stressed the acceptance of this reaction, and crucially emphasized non-violent activity as the proper means of civil disobedience. King echoed Gandhi in these important refinements, adding a more detailed discussion of the manner in which civil disobedience targets specific laws of the state while preserving fidelity to the state, recognizing its legitimacy in spite of its endorsement of some unjust laws.

Contemporary philosophers and theorists, notably John Rawls and Ronald Dworkin, contributed to our understanding of civil disobedience by continuing to refine how it is defined. Rawls, in his book *A Theory of Justice*,^[6] argued that an adequate theory of civil disobedience would define and differentiate it from other forms of disobedience and specify the conditions under which civil disobedience is justified. An adequate theory also will explain the role that civil disobedience will play within the political system.

Civil disobedience, according to Rawls, involves a number of important features that serve to distinguish it from other forms of political action or protest. First, for Rawls, civil disobedience is an essentially political act, and thus is done openly and publicly. By engaging

in civil disobedience, one hopes to bring a claim of injustice to the attention of the populace. Because civil disobedience is political, one must appeal to those political principles that are widely accepted by the members of that society. Civil disobedience is an appeal to the majority and its sense of justice, and by engaging in civil disobedience, one hopes to show the majority that the policy or law in dispute is inconsistent with the society's shared political values.

This public, political appeal must be non-violent, because it is the final appeal to the society to have one's claim of injustice addressed. Like Gandhi and King, Rawls thought that civil disobedience should be non-violent, because it will be more effective in accomplishing one's goals. Violence or violation of another's civil liberties, Rawls argued, could obscure the justness of the claim that one hopes to bring before the majority.

Further, Rawls shared King's conviction that non-violence expresses one's fidelity to the law, and insofar as civil disobedience is a means of altering rather than replacing the current political system, non-violence is an important way of expressing such fidelity. Rawls explained that, while civil disobedience is done with the hopes of changing the laws, it necessarily involves the violation of laws. One need not violate the law that is being protested, but nonetheless, the civilly disobedient act is an act that violates the law. Thus, according to Rawls, civil disobedience is defined as "...a public, non-violent, conscientious yet political act contrary to law usually done with the aim of bringing about a change in the law or policies of the government."^[6]

Rawls contrasted civil disobedience with conscientious refusal. Conscientious refusal is the violation of laws or administrative orders because the laws or orders violate one's personal convictions. One may, for instance, refuse to obey a certain law because to do so would violate a strongly held religious belief. This is conscientious refusal, not civil disobedience as Rawls understood it. Civil disobedience and conscientious refusal are distinguished by a variety of features. For instance, conscientious refusal is not a public act. That is, it is not intended to address the majority's sense of justice. In fact, when engaged in conscientious refusal, one may wish that the disobedience remain unknown to the general population and the government. Further, conscientious refusal is not necessarily based on the shared political principles of the society. Conscientious refusal is not necessarily a political act and thus may be grounded in one's personal moral, religious, or philosophical convictions.

Rawls argued that civil disobedience is justified when three conditions are satisfied. First, civil disobedience is appropriate only when there are substantial and clear injustices. Rawls thought that the

most serious injustices will involve violations of liberty as opposed to economic injustices. Second, civil disobedience should generally be a last resort for bringing about political change. Before resorting to civil disobedience, one should exhaust all other options that may redress the injustice. That is, all other appeals to the majority as well as legal appeals have failed to resolve the problem, and thus, the only remaining option to rectify the injustice is civil disobedience. Finally, Rawls argued that one must not disobey if this will threaten the existence of the state. Rawls explained that for civil disobedience to be justified, it must not threaten the stability of the government. Because civil disobedience is intended to bring about improvements within the government, individuals must be certain they are not threatening its existence by raising the objection. If the government or society is particularly unstable at the time, or a number of other people or groups are engaging in civil disobedience, then this is good reason to wait until the society is stable enough to address another claim of injustice.

According to Rawls, civil disobedience will play the role of a stabilizing device within the state. Rawls claimed that, much like free and regular elections or a judiciary empowered to interpret the constitution, civil disobedience will play an important role in preserving the integrity of the legal or political system. Here, again, one can see the distinctively political role that Rawls reserved for civil disobedience. Civil disobedience is just another part of the system of checks and balances within Rawls's theory of justice. The legal philosopher, Ronald Dworkin, disagrees with Rawls and King on this account, and sees no reason to limit civil disobedience only to cases of political protest.

While Dworkin and Rawls's accounts of civil disobedience were similar, the two philosophers disagreed in a number of important respects. In his book *A Matter of Principle*,^[7] Dworkin argues that when one's morals, conscience, or integrity forbid obedience to a law or policy, then that person need not attempt to alter the law by other methods before refusing to obey. Such instances of what Dworkin called "integrity-based civil disobedience" justify one's refusal to obey, because when people violate their morals, conscience, or integrity in order to obey a law, they suffer immediate, irreparable harm by doing so. Thus, Dworkin broke from Rawls insofar as Rawls insisted that civil disobedience must be one's last resort. For Dworkin, one is justified in disobeying if the law requires a significant compromise of that person's moral, religious, or philosophic beliefs.

This is contrasted with justice-based civil disobedience which, according to Dworkin, is disobedience intended to challenge an unjust political policy or program. In those instances in which people

object to the law because they have a justice-based complaint, Dworkin argued that they must first attempt to exhaust all other political avenues for bringing about the desired change before resorting to civil disobedience.

Additionally, Dworkin, unlike Rawls, did not consider publicity to be an essential feature of civil disobedience. Some instances of civil disobedience, for instance, dodging the draft or ignoring slave laws, are best accomplished covertly. People may wish to be arrested and charged with a crime to bring their cause to the attention of the majority. But, in order to be an act of civil disobedience, an individual need not have to be charged with a crime or be publicly violating the laws that the individual deems objectionable.

CIVIL DISOBEDIENCE AND PUBLIC POLICY IN THE UNITED STATES

The civil rights movement of the mid-twentieth century in the United States offers no shortage of illustrations of civil disobedience in practice, and through them, we can evaluate their relation to public policy. Perhaps one of the more famous cases begins with Rosa Parks, who, on December 1, 1955, violated segregation laws in Montgomery, Alabama, by refusing to yield her seat to White passengers boarding a bus after her. Her subsequent arrest touched off a year of monumentally important social upheaval, culminating in the November 13, 1956, decision by the U.S. Supreme Court declaring bus segregation laws unconstitutional. This set the stage for further pivotal features of the civil rights movement, including the national prominence of Dr. Martin Luther King, Jr., the protests in Birmingham in 1963, and the Civil Rights Act of 1964.

Immediately following the arrest of Rosa Parks, the African-American citizens of Montgomery, led in part by Dr. King and his associates, launched a bus boycott that would last for months. In lieu of buses, participants in the boycott took taxis, engaged in car pools, or even walked, maintaining a solidarity that would inform and inspire the civil rights movement. All of these actions, however, stood in violation of existing anti-boycott laws, and led to the arrests of 89 African-American leaders, including King. When this tactic failed, local officials turned to less scrupulous means in an attempt to undermine the boycott. Fabricated violations led to indiscriminate arrests of African-Americans; those awaiting carpools were charged with crimes ranging from illegal hitchhiking to loitering. Laws were passed with the clear purpose of undermining the boycott; African-American taxi drivers, who had taken to charging African-Americans a lesser fare in order to support the boycott, were prohibited from doing so. The city then attacked the car

pools as a public nuisance interfering with private enterprise, a charge affirmed by local courts on November 13, 1956. This last step would have been devastating for the movement, had it not been for the Supreme Court's decision that day agreeing with a lower federal court's opinion that the segregation laws were unconstitutional—a case originating from Rosa Parks' refusal to pay the fine demanded of her for refusing to surrender her seat nearly a year earlier.

Montgomery, then, saw acts of civil disobedience against bus segregation laws and antiboycott laws, first by Rosa Parks and then by the African-American community participating in the bus boycott. These acts all challenged existing laws that were clearly unjust and did so while maintaining the non-violence described and practiced by King, among others. While official local response was frequently disgraceful, the federal judicial response saw through the moral bankruptcy of the laws and local officials in question and supported the Montgomery African-American community.

With regard to public policy, it is worth noting one of the two ways that acts of civil disobedience seek to effect change: through judicial review. That is, while acts of civil disobedience most often seek to generally raise public awareness of a wrong, change is most frequently brought about by two avenues: through judicial review or through legislative action. For law enforcement officials, such as the police responding to cases of civil disobedience, response options are likely quite limited; insofar as civil disobedience involves violation of law, law enforcement officials cannot themselves change policy. They must simply enforce the laws impartially. Acts of civil disobedience, however, are aimed beyond law enforcement at the laws being enforced. These acts seek to set a course of judicial review underway or to raise awareness such that lawmakers will respond through legislation. When Rosa Parks' lawyers advised her to decline to pay her fine, their purpose was to pursue a legal challenge that would impact public policy via judicial decision. Importantly, the civil rights movement set in motion in part by Parks' action led to the Civil Rights Act of 1964, illustrating the legislative avenue of change.

This example of civil disobedience allows a clear evaluation of official response, precisely because of the way in which the protesters were plainly supported by justice and morality. In retrospect, we see clearly not only that the laws were unjust, but also that local law enforcement was in the wrong in administering punishment partially and arbitrarily. Looking beyond the civil rights movement, other cases where we are less clear about the moral standing of protesters and those they are protesting are also available, such as those involving protest against nuclear power and weaponry.

Perhaps one of the most famous movements against the development of nuclear power occurred during the

late 1970s in Seabrook, New Hampshire. In 1977, a group calling themselves the Clamshell Alliance occupied the construction site of the Seabrook nuclear power plant. During the protest, the New Hampshire police arrested over 1400 members of the Clamshell Alliance. Refusing to post bond of \$1500, the protesters were in prison for 2 weeks until the charges were dropped, and they were released.

The Alliance not only brought attention to the anti-nuclear power movement, but also helped further develop techniques of civil disobedience. The Clamshell Alliance had no centralized leadership, but instead, it was composed of smaller support groups or cells. Without centralized leadership, decisions had to be made democratically, and this was at times a slow process; however, the advantage was that there was no figurehead, and thus those who opposed the movement had no clear target to attack. Decentralized group organization was adopted by a number of other groups whose members sought to resist nuclear power in their communities.

There have also been a number of civilly disobedient actions taken against nuclear weapons and their production. On September 9, 1980, eight people entered the General Electric Nuclear Missile Re-entry division in King of Prussia, Pennsylvania. While in the factory, the eight poured bottles of their own blood on documents and hammered on, and thus destroyed, two nose cone casings for nuclear warheads. This group, whose members called themselves the plowshares eight (alluding to a biblical passage that references beating swords into plowshares), was convicted of burglary, conspiracy, and criminal mischief. In February of 1984, the Pennsylvania Superior Court overturned the conviction of the plowshares eight. The appellate process continued, but the plowshares eight were never sent back to prison. This act started what has since become known as the plowshares movement.

The plowshares movement has continued to draw attention to the nuclear weapons debate by engaging in similar acts of civil disobedience. Similar to those committed by the plowshares eight. For instance, the group that came to be known as the minuteman III plowshares was clearly influenced by the actions of the plowshares eight. On August 6, 1998, the 53rd anniversary of the bombing of Hiroshima, Japan, activists Daniel Sicken and Sachio Ko-Yin entered a nuclear silo in Weld County, Colorado. Sicken and Ko-Yin poured blood and stenciled a broken rifle, the international symbol of disarmament, onto the concrete cap of the silo. They then hammered on the tracks used to slide the concrete cap back so that the missile may be launched. After bending the tracks and breaking the anchoring bolts that hold the tracks in place, the two men sat down on the concrete cap and waited to be arrested. Sicken and Ko-Yin were

later convicted of sabotage, conspiracy, and destruction of government property, and were sentenced to 41 months and 30 months in prison, respectively.

The prosecution and subsequent conviction of Sicken and Ko-Yin highlight a further matter of public policy related to civil disobedience. There is a great deal of disagreement about what is the appropriate governmental response to acts of civil disobedience. While most of the theorists and activists discussed here hold that those who violate the law must accept their resulting arrest willingly, it does not follow from this that the government should prosecute all those arrested. In fact, most will agree that the intentions and purposes of the individuals who engage in civil disobedience clearly distinguish them from common criminals. What is unclear, however, is whether the appropriate legal response to the common criminal is, for these reasons, inappropriate for the individual who engages in civil disobedience.

In his book, *The Morality of Civil Disobedience*,^[8] Robert T. Hall presented a number of possible legal responses to civil disobedience. The options discussed by Hall can be divided into three separate categories: one may refuse to prosecute any case of civil disobedience, one may prosecute selectively, or one may prosecute all cases.

The first possibility, advocated by legal philosopher Hugo A. Bedau,^[9] is to attach a rider to all laws stating that anytime a person violates a law in an act of civil disobedience or conscientious refusal, that person should not be prosecuted for the legal violation. The value of such a proviso is that the government would protect an important form of political or social expression.

Ronald Dworkin^[10] argued that civil disobedience does not necessitate legal action or the drastic legal protection suggested by Bedau. Instead, Dworkin argues, whether an act of civil disobedience should or should not be prosecuted is a matter of prosecutorial discretion. For instance, if the violation of law associated with a particular act of civil disobedience did not harm anyone, then the prosecutor may have good reason to refuse to press the case against the individual who broke the law. In cases that involve harm, be it to persons or institutions, or the destruction of property, the prosecutor may then have a compelling reason to press the case against the individual. It is important to note that Dworkin believes no special actions or arrangements need to be made for cases of civil disobedience, prosecutorial discretion is already a feature of the legal landscape.

Another possibility is to allow a jury to address the problem of civil disobedience. That is, if the jury feels that the defendant was justified in violating the law, then they may acquit, despite the fact that they believe the defendant is genuinely guilty of a violation of

the law. Allowing the jury to decide the issue may include the complication of forcing the courts to allow testimony regarding the defendant's motivations for violating the law as part of the defense, but this should not be an insurmountable problem. Indeed, motives already play an integral role in legal actions. However, within the United States, the legal system has typically sought to limit the extent of jury nullification rather than extend it.

CONCLUSION

While the appropriate legal response to civil disobedience remains a significant problem for scholars and administrators, there can be no doubt that progress has been made in refining its definition and justification. Thoreau provides one of the first sustained discussions of civil disobedience. His argument for one's obligation to violate laws, when those laws contribute to injustices, helped lay the theoretical foundation for a more precise understanding of civil disobedience. Gandhi and King built on this foundation and helped permanently establish non-violence as a necessary component of civil disobedience. Further, Gandhi and King agree with Thoreau that one must accept the consequences for violating the law. However, they differed from Thoreau in that they believe that by accepting the punishment, one reveals a general attitude of fidelity to the law, thus bolstering the moral strength of one's position. The works of Rawls and Dworkin further clarified the definition, but perhaps more importantly, both substantially

clarified possible justifications for, and the role of, civil disobedience within a political system. Yet, it is not the writings but the actions of Thoreau, Gandhi, King, the Clamshell Alliance, the plowshares movement, and others like them that makes clear the important role that civil disobedience can play in addressing injustices within a society.

REFERENCES

1. Plato. *Crito*; (Trans. G.M.A. Grube). In *Plato: complete works*; Cooper, J.M., Hutchinson, D.S., Eds.; Hackett Publishing Company: Indianapolis, 1997; 45 pp.
2. Plato. *Apology*; (Trans. G.M.A. Grube). In *Plato: Complete Works*; Cooper, J.M., Hutchinson, D.S., Eds.; Hackett Publishing Company: Indianapolis, 1997; 27 pp.
3. Thoreau, H.D. Resistance to civil government. In *Political Writings*; Cambridge University Press: Cambridge, 1996; 1–21.
4. Gandhi, M.K. *Selections From Gandhi*; Bose, N.K., Ed.; Navajivan Publishing House, 1957; 149–172.
5. King, M.L., Jr. Letter from a birmingham jail. In *Civil Disobedience*; Bedau, H.A., Ed.; Pegasus: New York, 1969; 72–89.
6. Rawls, J. *A Theory of Justice*, 2nd Ed.; The Belknap Press, 1999.
7. Dworkin, R. *A Matter of Principle*; Harvard University Press, 1985; 104–116.
8. Hall, R.T. *The Morality of Civil Disobedience*; Harper Row, Publisher, Inc.: New York, 1971; 131–151.
9. Bedau, H.A. On civil disobedience. *J. Philos.* **1961**, *58* (21), 653–665.
10. Dworkin, R. *Taking Rights Seriously*; Harvard University Press, 1977; 206–222.

Civil Service

J. Edward Kellough

*School of Public and International Affairs, University of Georgia,
Athens, Georgia, U.S.A.*

INTRODUCTION

In a general sense, the civil service is comprised of all individuals employed in government, with the exception of those in the military and those who work for the legislature and judicial branches of government. As such, civil service systems exist at every level of government and in every country. Because those who work in the civil service are charged with the responsibility for implementing and managing government operations and programs, the relationship between the civil service and political authorities at the heads of governments has always been of critical importance. Competence, responsiveness, and accountability in the civil service are issues often defined in terms of the context of that relationship. It is expected that public employees will possess the knowledge, skills, and abilities requisite for their positions; that they will be responsive to the public that they serve; and that they will be held accountable to the public and government leaders for their official actions.

THE CIVIL SERVICE SYSTEM

Early in the history of the United States, positions in the civil service were controlled largely by political authorities and were obtained primarily on the basis of political connections. At the federal level, the earliest presidents sought to guarantee that appointments went to people who were capable, but they also preferred people who held political views largely consistent with those advocated within their administrations. Eventually, as political parties became more active and more important nationally, party affiliation became a central criterion for appointment. This system, which became known as political patronage, was defended on the grounds that most jobs in government required little technical skill and that appointment on the basis of political connections would help to ensure responsiveness and accountability to the chief executive. Eventually, however, questions of competence were made subordinate to political considerations, and corruption and inefficiency were common. Patronage practices had given rise to a spoils system in which the offices of government were considered the rightful

plunder of those who were victorious in the electoral process. As the parties alternated in control of the executive branch in the mid-1800s, instability, graft, and the abuse of public authority became common problems.

Significant reform of the spoils system at the federal level began with passage of the Pendleton Act of 1883. The law was passed following the assassination of President James A. Garfield by an individual who had lobbied the Garfield administration unsuccessfully for an appointment in the civil service. Garfield's assassination, along with generally declining fortunes of the Republican party, prompted the Republican majority in Congress to support the Pendleton bill. The legislation, which originally covered only approximately 10% of the federal workforce, specified that employees would obtain their jobs on the basis of their performance on open and competitive examinations designed to measure their abilities to perform the work at issue. Employees covered by the system were also protected from political abuse or removal from office solely for political reasons. As time passed, this approach to public employment, known as the merit system, was extended to cover more than 90% of the federal workforce, and today, most state and local governments have civil service systems similarly staffed on the basis of the merit principle.

Indeed, the prevalence of merit systems for public employment is such that the term civil service is often equated only with those positions filled on the basis of demonstrated ability rather than political connections. Defined in this manner, the civil service is governed by elaborate systems of rules and procedures intended to insulate public employees from undue political influence. For example, in addition to hiring, there are rules governing promotions, transfers, performance appraisal, and compensation. Politically neutral competence on the part of employees and equity or fairness in the manner in which employees are treated within the system are the core values served by the structures of rules and procedures. Obviously, such systems have the effect of constraining managerial discretion or the freedom with which managers may assign and utilize public employees. That is, in fact, the intention behind the array of civil service rules. Some observers, however, argue that greater flexibility

for management is necessary to more effectively and efficiently achieve the purposes of government.

Currently, popular civil service reform proposals are designed to bring about those greater levels of flexibility. Consistent with concepts known as the “new public management,” recent reform initiatives stress reinvention, reengineering, decentralization, and deregulation of the civil service. Advocates for these approaches argue that authority for such key personnel actions within the civil service as selection, promotion, pay, and the classification of jobs should be moved from central personnel offices and placed in the various line agencies, where agency managers will be able to use that authority to better pursue organizational goals and objectives. In this way, human resources functions within the civil service could be better integrated with the planning and management control functions within public organizations. There is also a sense that by placing greater authority for the civil service in the hands of agency heads or directors, who often are political appointees, accountability from the civil service to political authority, especially executive authority, will be enhanced. While that may be the case, there may also be fewer safeguards designed to protect employees and promote equity in the public service. In effect, the great difficulty is to find a way to properly balance the legitimate need for flexibility and political control with appropriate levels of employee protection consistent with the concept of politically neutral competence. The U.S. federal civil service system has been reformed moderately in recent years to allow for greater flexibility in areas such as employee examination and selection procedures, but some state governments have pushed those reforms much further. The state of Georgia, for example, implemented a wide-ranging set of reforms in which authority over personnel policy within the state civil service has been highly decentralized. There is little empirical evidence, however, to demonstrate that the new personnel system is a significant improvement over the more traditional arrangement that had been in place earlier.

CONCLUSION

It is obvious that the civil service performs numerous essential tasks in any society. It is the means by which government services are delivered. One can generally find an extremely wide range of jobs within the civil service, ranging from skilled crafts and trades to engineers, scientists, technicians, and a host of professions from law to medicine. There are also numerous administrators skilled in organizing and managing organizational activities. Certainly, the quality of

public services and even the perceived legitimacy of government are determined in large part by the quality of the civil service and the way in which it operates. Given the range of occupations and complexity, the administration and management of a civil service personnel system is not an easy task. Competent and capable employees must be recruited and selected. Jobs must be classified according to the duties associated with them, and systems of compensation that are fair to employees and competitive with employers in the private sector must be devised. Employee training and development programs must also be implemented. Performance appraisal systems must be established, and policies for employee discipline must also be placed into operation. All of these personnel tasks must be accomplished in an effective manner if the civil service is to function properly.

BIBLIOGRAPHY

1. Ban, C.; Riccucci, N.M. *Public Personnel Management: Current Concerns, Future Challenges*, 3rd Ed.; Longman: New York, 2002.
2. Berman, E.M.; Bowman, J.S.; West, J.P.; Wart, M.V. *Human Resource Management in Public Service: Paradoxes, Processes, and Problems*; Sage Publications: Thousand Oaks, CA, 2001.
3. Daley, D.M. *Strategic Human Resources Management: People and Performance Management in the Public Sector*; Prentice Hall: Upper Saddle River, NJ, 2002.
4. Gatewood, R.D.; Field, H.S. *Human Resource Selection*, 4th Ed.; The Dryden Press: Fort Worth, TX, 1998.
5. Hays, S.W.; Kearney, R.C. *Public Personnel Administration: Problems and Prospects*, 4th Ed.; Prentice Hall: Englewood Cliffs, NJ, 2002.
6. Ingraham, P.W. *The Foundation of Merit: Public Service in the American Democracy*; The Johns Hopkins Press: Baltimore, MD, 1995.
7. Ingraham, P.W.; Ban, C. *Legislating Bureaucratic Change: The Civil Service Reform Act of 1978*; SUNY Press: Albany, NY, 1984.
8. Kearney, R.C.; Hays, S.W. Reinventing government, the new public management and civil service systems. *Rev. Public Pers. Adm.* **Fall 1998**, 18, 38–54.
9. Kellough, J.E.; Selden, S.C. Pay-for-performance systems in state government: perceptions of state agency personnel managers. *Rev. Public Pers. Adm.* **Winter 1997**, 17, 5–21.
10. Nigro, L.G.; Kellough, J.E. Civil service reform in Georgia: going to the edge? *Rev. Public Pers. Adm.* **Fall 2000**, 41–54.
11. Shafritz, J.M.; Rosenbloom, D.H.; Riccucci, N.M.; Naff, K.C.; Hyde, A.C. *Personnel Management in Government: Politics and Process*, 5th Ed.; Marcel Dekker: New York, 2001.
12. Van Riper, P.P. *History of the United States Civil Service*; Row, Peterson and Company: Evanston, IL, 1958.

Civil Service Reform Act of 1978

Jerrell D. Coggburn

*Department of Public Administration, University of Texas at San Antonio,
San Antonio, Texas, U.S.A.*

INTRODUCTION

The Civil Service Reform Act of 1978 (CSRA) is considered a watershed event in the evolution of federal personnel administration. This stems from the fact that it was the first comprehensive civil service reform passed at the federal level since the Pendleton Act of 1883. The CSRA made a number of major structural changes to the federal civil service system and embraced a number of important administrative values. In so doing, the CSRA established the general framework under which federal human resource management is conducted today.

BACKGROUND

The passage of the CSRA in 1978 represented the culmination of decades of growing dissatisfaction with the federal government's personnel system. In the period between the passage of the Pendleton Act—which created the Civil Service Commission (CSC) and a merit-based personnel system for the federal government—and the CSRA, fundamental changes occurred in the scope and size of government. Throughout the twentieth century, the federal government grew in size and complexity as the problems it addressed multiplied and became increasingly intractable (e.g., poverty, civil rights). Predictably, the public began to question government's growth and cost and to associate policy failures with bureaucratic waste, laziness, and ineptitude. Such sentiment was understandable given the public's historic skepticism and distrust of the government, especially in the post-Watergate era.

Perceptions from inside the federal government were equally critical. The Pendleton Act had created a civil service system that promoted merit and politically neutral competence, but such an emphasis was not without consequence. What emerged was a professionalized federal workforce, protected from political coercion, that became resistant to political control. This created tension for presidents and their political appointees who were responsible for government performance but felt unable to control it. Similarly, federal managers felt hampered by restrictive and cumbersome personnel procedures and regulations.

Employees, for their part, felt that they did not receive adequate recognition for their contributions. While a number of reforms were enacted over the years to address problems related to the civil service, they were piecemeal: as particular problems arose, the common approach was to pass legislation or to enact rules in an ad hoc fashion. Over time, however, these reforms served to create a civil service system widely perceived to be ineffective, inefficient, and unresponsive to presidential direction.

So, by the eve of the CSRA's passage, observers could agree upon a number of problems afflicting the federal personnel system, including the CSC's dual and conflicting mission (discussed below); the insularity of civil servants from political direction; excessive administrative procedures; a socially unrepresentative federal bureaucracy; and an inability to reward excellent employee performance.^[1] In order to remedy the civil service's myriad problems, wholesale change was needed.^[2] That wholesale change came in the form of the CSRA, which is generally considered one of President Jimmy Carter's most significant domestic policy achievements.

STRUCTURAL CHANGES

The CSRA included a host of changes to the organizational arrangements for carrying out the federal government's human resource function. Most of these structural changes were achieved through Reorganization Plan Number 2. Though Reorganization Plan Number 2 is not technically a component of the actual CSRA, it established the framework for carrying out the CSRA's provisions, so they are generally discussed as one. The most obvious structural change was the abolition of the CSC. The CSC, created by the Pendleton Act, was a three-member, bipartisan panel that represented an attempt to remove overt political control (i.e., the "spoils") over the federal personnel system. While largely successful in removing political abuses, experience showed that the CSC was pressured by its dual mission of administering the personnel function, including protecting merit principles, and serving as the president's personnel advisor. In its place, the CSRA created several administrative agencies.

Office of Personnel Management

Under the CSRA, the Office of Personnel Management (OPM) became the main personnel management agency of the federal government. Among its mandates, OPM is responsible for making federal personnel policy, advising federal agencies on personnel-related issues, and monitoring compliance with personnel laws and regulations. In contrast to the former CSC's format, OPM is headed by a single director, appointed by and responsible to the president.

Merit Systems Protection Board

The CSRA assigned CSC's role of protecting merit principles to the newly created Merit Systems Protection Board (MSPB). The MSPB is a bipartisan, quasi-judicial entity that carries out its merit protection role by hearing employee appeals and reviewing the government's personnel regulations. Additionally, the MSPB conducts research studies to ensure federal agencies' actions (including those of OPM) comply with merit principles. The MSPB's three members are appointed by the president and serve fixed, 7-year terms.

Office of Special Counsel

Within the MSPB, the CSRA created the Office of Special Counsel (OSC). The CSRA gives the OSC investigative and prosecutorial powers. In other words, the OSC has authority to investigate cases of alleged violation of prohibited personnel practices and to prosecute cases where civil service laws and regulations have been violated. OSC places special emphasis on providing protection to whistleblowers, that is, those federal employees who report violations of law, government mismanagement, or waste. The counsel is appointed by the President, with the Senate's advice and consent, and serves a 5-year term.

Federal Labor Relations Authority

The act also created the Federal Labor Relations Authority (FLRA). The "authority," as it is commonly referred to, is charged with overseeing union-management relationships, including determining bargaining units; certifying bargaining agents; and resolving labor-management impasses through its Federal Service Impasses Panel. The provision creating the authority (Title XII) was not originally contained in the CSRA as presented to Congress, but it was added later during the legislative process. Title XII is significant, because it gives statutory recognition to collective bargaining at the federal level. Prior to the

CSRA, collective bargaining rights for federal employees had only been recognized in presidential executive orders. The three members of the authority are appointed by the president (with senate approval) and serve overlapping 5-year terms.

Senior Executive Service

One of the most significant features of the CSRA was the creation of the Senior Executive Service (SES). The act envisioned the SES to be a cadre of top-tier executives, drawn from the highest levels of the federal service, serving as managerial generalists. In exchange for forgoing some protections they enjoyed under the civil service system, members of the SES are given the opportunity to compete for performance-based pay increases and bonuses (which can be as high as \$20,000), to assume policy-making roles, and to develop their managerial skills through rotating assignments. Rotating assignments are also important from a political standpoint, because presidents and their political appointees (e.g., departmental or agency heads) are given flexibility to move senior career managers to the programs they deem appropriate. The basic idea for the SES is similar to Great Britain's higher civil service model that establishes "rank in the person"—members of the SES retain their rank (pay, status), even if their specific positions change.

VALUES AND CONCEPTS EMBODIED IN THE CSRA

A Foundation of Merit

In addition to the CSRA's major structural changes, the act is important for the administrative values and managerial concepts it embodies. Obviously, the act sought to improve the efficiency and effectiveness of the federal personnel system. These represent classic public administration values and are typically used to sell all sorts of administrative reform, not just civil service reform. But, in addition to these instrumental performance goals, the act reaffirmed the federal government's commitment to a number of important principles, not the least of which is the merit principle.

According to Alan K. Campbell, a chief architect of the CSRA and the first director of OPM, the merit principle was the foundation upon which the entire federal personnel function was to be based.^[2] The drafters of the CSRA considered merit so important that the act codifies a statement of the merit principle at its outset (see [Table 1](#)). Accompanying this delineation was a statement of prohibited personnel practices ([Table 2](#)) which sought to clarify any ambiguity about

Table 1 Merit system principles

- Recruiters should seek to achieve a workforce made up of qualified people from all segments of society, and selection and promotion should be based solely on merit, after fair and open competition.
- All employees and applicants should be treated fairly, without discrimination, and with proper regard for their privacy and constitutional rights.
- Equal pay should be provided for work of equal value, with appropriate consideration for the rates paid by private sector employers. Appropriate incentives and recognition should be provided for excellent performance.
- Employees should maintain high standards of integrity, conduct, and concern for the public interest.
- The federal workforce should be used efficiently and effectively.
- Retention should be based on performance, inadequate performance should be corrected, and employees who cannot or will not improve their performance to meet required standards should be separated.
- Employees should receive effective education and training in order to achieve better organizational and individual performance.
- Employees should be protected against partisan politics and may not use their offices for partisan political purposes.
- When employees legally disclose information evidencing wrongdoing (known as “whistle-blowing”), they should be protected from reprisal.

Source: Merit System Protection Board, [http://www.mspb.gov/foia/forms-pubs/intro.html#The Merit System](http://www.mspb.gov/foia/forms-pubs/intro.html#The%20Merit%20System) (accessed August 2001).

who (anyone with personnel authority) could be held responsible for what (violating prohibited personnel practices). Together, these provisions underscore the importance of merit as a guiding value in federal personnel under the CSRA.

Executive Control

While the act clearly reaffirmed the federal government’s commitment to merit, it also represented the desire to make the federal bureaucracy more directly accountable to presidential control. The act strengthens such political accountability in a number of ways. For example, making OPM’s director a presidential appointee helps to ensure that federal personnel policy will be responsive to presidential (political) priorities. Also, the structure of the SES, with career members working on a contractual basis in agencies led by presidential appointees, means that the reward and advancement of SES members depend in part on the performance assessments of these political appointees.

Table 2 Summary of prohibited personnel practices

- Discriminating on the basis of race, color, religion, sex, national origin, age, disability, marital status, or political affiliation.
- Soliciting or considering statements concerning a person who is being considered for a personnel action unless the statement is based on personal knowledge and concerns the person’s qualifications and character.
- Coercing the political activity of any person, or taking any action as a reprisal for a person’s refusal to engage in political activity.
- Deceiving or willfully obstructing anyone from competing for employment.
- Influencing anyone to withdraw from competition for any position to help or hurt anyone else’s employment prospects.
- Giving unauthorized preferential treatment to any employee or applicant.
- Nepotism.
- Taking or failing to take, or threatening to take or fail to take, a personnel action because of an individual’s legal disclosure of information evidencing wrongdoing (“whistle-blowing”).
- Taking or failing to take, or threatening to take or fail to take, a personnel action because of an individual’s exercising any appeal, complaint, or grievance right; testifying or lawfully assisting any individual in the exercise of any appeal, complaint, or grievance right; cooperating with or disclosing information to an agency Inspector General or the Special Counsel, or refusing to obey an order that would require the individual to violate a law.
- Discriminating on the basis of personal conduct that does not adversely affect the performance of an employee or applicant or the performance of others, except that an employee or applicant’s conviction of a crime may be taken into account in determining suitability or fitness.
- Taking or failing to take any other personnel action if the act or omission would violate any law, rule, or regulation implementing or directly concerning the merit system principles.
- Violating a veterans’ preference provision in connection with a personnel action (for purposes of disciplinary action only).

Source: Merit System Protection Board, [http://www.mspb.gov/foia/forms-pubs/intro.html#Prohibited Personnel](http://www.mspb.gov/foia/forms-pubs/intro.html#Prohibited%20Personnel) (accessed August 2001).

Thus, an additional increment of political control is provided by the CSRA.

Social Representation

The CSRA reaffirmed the federal government’s commitment to a socially representative workforce.

This reaffirmation came in various ways. For example, the act explicitly states the federal government's policy of having a workforce reflective of the nation's diversity. Toward that end, the act established a minority recruitment program—the Federal Equal Employment Opportunity Recruitment Program—within OPM. In addition, the CSRA shifted responsibility for enforcing non-discrimination in the federal government to the Equal Employment Opportunity Commission. Finally, the act, as originally submitted to Congress, attempted to significantly curtail veteran's preference by limiting veteran's preference to a one-time, initial hiring preference. This was important, because granting hiring preferences to veterans, the overwhelming majority of whom are male, hampers diversification efforts. This provision, however, was dropped during congressional consideration of the act in favor of less far-reaching limitations.^[3]

Decentralization

The CSRA gives OPM authority to delegate significant areas of responsibility to line agencies within the federal government. In so doing, the CSRA makes a statement about the desirability of personnel decentralization. The delegation and decentralization of examining, for example, is intended to give agencies direct control over determining employee qualifications and hiring, while freeing OPM to devote more of its energies to providing technical assistance to agencies.^[4] In such an arrangement, agency managers have greater flexibility over personnel decisions and can make these decisions more quickly. It also emphasizes a service orientation for OPM instead of a centralized decision-making role.

Whistle-Blower Protection

The CSRA's inclusion of whistle-blower protections embodies at least two administrative values: employee rights and efficiency. Employee rights are bolstered by the OSC's charge to investigate cases of alleged reprisals suffered by whistle-blowers. The intent is to encourage employees to come forward with allegations of waste, mismanagement, and wrongdoing. This, in turn, will improve the efficiency of government, as these costly problems are rooted out.

Private Sector Management

Certain provisions of the CSRA represent the federal government's acceptance of private sector management models. The clearest examples of this are the act's merit pay provisions for mid-level managers

(GS 13–15) and the model of executive compensation created for the SES. Both forms, which rely on individual performance evaluation, seek to increase performance by offering employees financial incentives. The logic used by the CSRA's drafters was simple: if the federal government was to become more efficient, it needed to adopt private sector techniques like performance evaluation and pay for performance.^[5]

Experimentation

Finally, the CSRA provides for personnel experimentation through demonstration projects. Implicitly, this recognizes the inadequacy of synoptic, "one size fits all" approaches to personnel management. By allowing experimentation with novel personnel practices, the CSRA seeks to improve personnel management while reducing the risk of applying reforms or new management practices government-wide prior to discovering their full implications.

APPRAISING THE CSRA

In general, academic treatments of the CSRA's implementation and impact have been unfavorable. At worst, the act has been described as a failure; at best, it has been said to have fallen short of its highest objectives. The act's shortcomings have been tied to its internal contradictions that reflect the tensions its designers faced, including: 1) the desire for decentralization to promote better management versus the desire for centralization to protect merit; 2) the maintenance of control through rules and regulations versus delegating authority and stressing individual accountability; 3) the desire for better management versus political responsiveness; and 4) reorganization of personnel structures to effect positive change versus revitalization of the civil service.^[6] Political and economic factors have also been blamed for the act's failings. For example, implementation of significant pieces of the act occurred during the Reagan Administration. In the throes of economic stagflation, Reagan, who has never been described as a friend of the civil service, focused more on government cutbacks than on achieving the implementation objectives of the CSRA. Thus, the CSRA suffered from design and implementation problems. Specific examples of the act's problems are easy to identify.

First, consider the agencies created by the CSRA. In a survey of federal personnelists conducted by the MSPB, OPM received poor marks in key areas like technical assistance, monitoring and oversight, and policy guidance.^[7] Academic researchers have been equally critical of OPM, noting that it has not

modernized federal personnel management as envisioned by the CSRA and that many of the problems it was intended to remedy persist.^[8] The MSPB and OSC have been criticized for not being aggressive in their efforts to protect merit and, in the case of MSPB, for favoring management in its decisions.^[9] Also, the CSRA's language left unclear the relationship between the OSC and the MSPB. The tensions this created ultimately led to OSC being established as a separate, independent agency by the Ethics Act of 1989. Finally, researchers have given the FLRA credit for the timeliness and impartiality of its decisions but have also pointed out the ineffectual nature of its decisions: FLRA decisions are often overturned upon judicial review.^[10] So, while the CSRA's structural changes have been effected, each of the new agencies has faced its share of operational difficulty.

Second, problems have plagued the SES. Attracted by the lure of advancement and monetary incentives, almost all of those eligible for the SES joined in the first year. Many felt betrayed when, within the year, Congress cut the number of SES members eligible for performance bonuses from 50 to 25%. Rotation in office has also proven to be troublesome. Intended to be a managerial development tool, many in the SES view reassignments as punitive measures. Early on, problems such as these resulted in low morale among SES members, rapid turnover within the SES (more than half of the original members were gone in 3 years), and a majority of SES members saying they would not recommend a career in federal government.^[11] More recently, however, the SES has seen some positive developments. Resentment over pay issues was ameliorated when the Reagan administration lifted salary caps for the SES, and a 1992 MSPB survey^[12] found SES members to have the highest levels of job satisfaction among all government employees. Despite this, most observers agree that the SES is one of the more disappointing aspects of the CSRA.

Finally, several other objectives of the act have achieved disappointing results. Merit pay, for example, has proven to be virtually unworkable in the federal government. The problems, which have been examined extensively by academics (for example, see Ref.^[13]), center on employee perceptions about the fairness of merit pay determinations, the link between performance and pay, and the size of merit pay increases. Recognition of the problems in the merit pay for middle manager system came quickly, as Congress replaced the original system after only 3 years with the Performance Management System. This replacement system faced its own problems and was eliminated in 1993.

The act sought to afford additional protections to whistle-blowers, thus encouraging them to come forth with allegations. This objective has met with mixed success. A report by the MSPB^[14] shows a decrease in

the number of employees who reported witnessing wasteful activities. This would suggest that less waste has occurred over time. The same report shows that the number of employees blowing the whistle when waste was witnessed has increased over time. This again would suggest some success: fewer incidents, but a higher rate of reporting on the incidents that are observed. The problem, however, is that unacceptably high percentages (37%) of whistle-blowers report suffering reprisals for reporting the waste they observed.

The story is similar in the area of adherence to merit principles and prohibited personnel practices. As stated above, the CSRA reaffirmed the merit principle as the foundation for federal personnel management. However, alarmingly high numbers of federal employees report problems with their respective agencies' adherence to merit. For example, one-third of those responding to a 1996 MSPB survey indicated a belief that their agencies regularly fail to uphold merit in hiring and promotions, over one-third indicated their agencies do a poor job of protecting employees from personal favoritism, and one-fourth indicated their agencies do a poor job of ensuring equal pay for equal work.^[15] In the area of equal employment opportunity, the same survey found a significant minority of employees who reported being denied a job or job benefit: 14% reported suffering racial discrimination, 13% reported sexual discrimination, and 11% reported age discrimination.^[16] Findings such as these illustrate the difficulty of ensuring adherence to the most fundamental civil service tenets in a government of approximately 3 million employees.

Finally, experimentation through personnel demonstration projects has not lived up to expectations. This is due in part to limits placed on the number of demonstration projects that can be authorized (no more than 10 at one time), the time-consuming nature of receiving approvals to conduct the projects, and OPM passivity in encouraging agency demonstration projects.^[17] Still, the demonstration projects are not without value. The China Lake Naval demonstration project, for example, experimented with the use of broad banding pay. This approach collapses pay grades into wide pay bands, thereby increasing managerial flexibility in pay administration. The positive results achieved at China Lake have helped to make broad banding a mainstay of contemporary personnel reform proposals. Even here, however, government-wide implementation remains elusive.

CONCLUSION

The CSRA held out the promise of creating a more efficient, effective, and responsive personnel system. Given all that it promised, coupled with the novelty

of many of its reform components—at least as applied to the *public* sector—it is not surprising that the act has not been able to deliver on all fronts. Indeed, the continued interest in reforming the federal civil service (e.g., the Clinton Administration's National Performance Review) attests to the fact that the CSRA was not the success story proponents had hoped for. Despite its shortcomings, the CSRA retains prominence in the public administration field, because it still constitutes the broad framework for federal human resource management.

REFERENCES

1. Ingraham, P.W. The civil service reform act of 1978: its design and legislative history. In *Legislating Bureaucratic Change: The Civil Service Reform Act of 1978*; Ingraham, P., Ban, C., Eds.; State University of New York Press: Albany, NY, 1984; 13–28.
2. Campbell, A.K. Testimony on civil service reform and reorganization (March 14, 1978). In *Classics of Public Personnel Policy*; Thompson, F., Ed.; Moore: Oak Park, IL, 1979; 77–102.
3. Rosenbloom, D.H.; Berry, C.R. The civil service reform act and EEO: The record to date. In *Legislating Bureaucratic Change: The Civil Service Reform Act of 1978*; Ingraham, P., Ban, C., Eds.; State University of New York Press: Albany, NY, 1984; 182–199.
4. Ingraham, P.W.; Reed, P.N. The civil service reform act of 1978: the promise and the dilemma. In *Public Personnel Administration: Problems and Prospects*, 2nd Ed.; Hays, S., Kearney, R., Eds.; Prentice Hall: Englewood Cliffs, NJ, 1990; 277–291.
5. Ingraham, P.W.; White, J. The design of civil service reform: lessons in politics and rationality. *Policy Stud. J.* **1989**, *17* (2), 315–330.
6. Ingraham, P.W. The design of civil service reform: good politics or good management? In *The Promise and Paradox of Civil Service Reform*; Ingraham, P., Rosenbloom, D., Eds.; Pittsburgh University Press: Pittsburgh, PA, 1992; 19–36.
7. Merit Systems Protection Board. *Federal Personnel Management Since Civil Service Reform: A Survey of Federal Personnel Officials, A Report to the President and the Congress of the United States*; Merit Systems Protection Board: Washington, DC, 1989; 1–23.
8. Lane, L.M. The administration and politics of reform: the office of personnel management. *Policy Stud. J.* **1989**, *17* (2), 331–351.
9. Vaughn, R.G. The U.S. merit systems protection board and the office of special counsel. In *The Promise and Paradox of Civil Service Reform*; Ingraham, P., Rosenbloom, D., Eds.; Pittsburgh University Press: Pittsburgh, PA, 1992; 121–140.
10. Rosenbloom, D.H. The federal labor relations authority. In *The Promise and Paradox of Civil Service Reform*; Ingraham, P., Rosenbloom, D., Eds.; Pittsburgh University Press: Pittsburgh, PA, 1992; 141–156.
11. Huddleston, M.W. Is the SES a higher civil service? *Policy Stud. J.* **1989**, *17* (2), 406–419.
12. Merit Systems Protection Board. *Working for America: An Update, A Report to the President and the Congress of the United States*; Merit Systems Protection Board: Washington, DC, July 1994; 1–42.
13. Perry, J.L. The merit pay reforms. In *The Promise and Paradox of Civil Service Reform*; Ingraham, P., Rosenbloom, D., Eds.; Pittsburgh University Press: Pittsburgh, PA, 1992; 199–215, see also Gaertner, K.N.; Gaertner, G.H. Performance evaluation and merit pay: results in the environmental protection agency and the mine safety and health administration. In *Legislating Bureaucratic Change: The Civil Service Reform Act of 1978*; Ingraham, P., Ban, C., Eds.; State University of New York Press: Albany, NY, 1984; 87–111.
14. Merit Systems Protection Board. In *Working for America: An Update, A Report to the President and the Congress of the United States*; Merit Systems Protection Board: Washington, DC, July 1994; 1–42.
15. Merit Systems Protection Board. *Adherence to the Merit Principles in the Workplace: Federal Employees' Views, A Report to the President and the Congress of the United States*; Merit Systems Protection Board: Washington, DC, September 1997; 1–12.
16. Merit systems protection board. *The Changing Federal Workplace: Employee Perspectives, A Report to the President and the Congress of the United States*; Merit Systems Protection Board: Washington, DC, March 1998; 1–41.
17. Ban, C. Research and demonstration under CSRA: Is innovation possible. In *The Promise and Paradox of Civil Service Reform*; Ingraham, P., Rosenbloom, D., Eds.; Pittsburgh University Press: Pittsburgh, PA, 1992; 217–235.

Civil Service Reform and Reinvention

Jerrell D. Coggburn

*Department of Public Administration, University of Texas at San Antonio,
San Antonio, Texas, U.S.A.*

INTRODUCTION

Civil service reform and reinvention refers to the purposeful modification of governmental human resource management (HRM) systems with the goal of maximizing important administrative values. The particular modifications adopted and the values they manifest vary over time and circumstance.

THE IMPORTANCE OF ADMINISTRATIVE VALUES TO CIVIL SERVICE REFORM

In the most basic terms, civil service systems exist to meet three purposes:^[1]

1. To recruit qualified workers for public sector jobs.
2. To reward and develop the public workforce.
3. To provide rules for the public workforce to meet public objectives.

The simplicity of this depiction, however, masks important tensions inherent to civil service systems. These tensions are often described by academics as resulting from a number of competing administrative values seeking expression (Several authors have discussed the importance of administrative values. For example, see Ref.^[2]). At different points in time, some of these values receive greater emphasis relative to others, and these shifts in emphasis are reflected in the reforms adopted. Because an emphasis on one value can come at the expense of others, there remains continual pressure for reform. Thus, civil service reform has become an enduring feature of the American political system.

David H. Rosenbloom (This discussion of administrative values draws heavily on Ref.^[3]) provided one framework for understanding the values contained in civil service systems. Specifically, Rosenbloom contended that public administration can be understood from three perspectives: managerial, political, and legal. Each perspective is associated with certain administrative values. The following discussion highlights these administrative values within the context

of civil service reform (For complete treatments the history of American civil service reform, see Ref.^[4]).

Political Values

First, the political perspective emphasizes two administrative values: representativeness and responsiveness. Representativeness means that the composition of the civil service should mirror American society. This, in turn, should allow the actions undertaken by administrative agencies to reflect (or “represent”) the will of the majority in the populace. This value has taken on several forms during the history of the civil service. Under George Washington, representativeness took on a geographic connotation because having all regions of the country represented was viewed as a way to insure legitimacy and support for the federal government. Representativeness has also entailed partisan considerations. This was evident in Thomas Jefferson’s efforts to balance the proportions of Republicans and Federalists within the federal bureaucracy so that they reflected the nation’s partisan composition. In more recent times, the value has taken on a demographic tone, which can be seen in governments’ Equal Employment Opportunity and Affirmative Action policies. Despite these variations, the general objective is the same: to have a civil service reflective of the nation.

Responsiveness is the second value associated with the political perspective. In the context of the civil service, this value means having a government that does not thwart the political will of citizens as reflected in their choice of elected officials. Essentially, this means that elected executives should have control over government’s policies and programs and, by necessity, the civil servants who implement them. This value was prominent in the patronage or “spoils” systems that emerged full-blown at the federal level following the election of Andrew Jackson in 1829. Under spoils, newly elected executives fill administrative positions with political supporters, whether they are technically qualified or not. Being supporters, such individuals are likely to be sympathetic (or “responsive”) to the policies of the newly elected. While the most egregious abuses of patronage have virtually vanished due to

various civil service laws and court decisions,^a the desire for responsiveness endures. For example, President Franklin Roosevelt established many “New Deal” agencies outside of the federal civil service system, President Eisenhower created a new category of federal policy-making positions (known as “Schedule C”), with appointees who were exempt from the civil service system’s provisions, while the Carter Administration’s Civil Service Reform Act (CSRA) of 1978 gave presidents power to make political appointments to high-ranking posts within the Senior Executive Service and to appoint the director of the Office of Personnel Management (OPM). In all of these cases, the underlying intent was the same: make the federal bureaucracy (i.e., the civil service) more responsive to presidential direction.

Managerial Values

The second perspective identified by Rosenbloom is the managerial perspective. This perspective emphasizes the values of efficiency, effectiveness, and economy. These values are to be achieved through the apolitical, business-like administration of the human resource (HR) function. As is the case with the values of the political perspective, the importance of managerial values has fluctuated over time. For example, in the late 1800s, the Progressive movement sought to remove rampant partisan abuses from the civil service. Under the spoils approach, it was not uncommon for unqualified workers to be appointed to office and then to use their positions for personal gain. Because new elections brought new rounds of patronage appointments, the activities of governmental agencies also lacked consistency and continuity. The passage of the Pendleton Act in 1883 sought to remedy this by creating a merit-based civil service system. This meant that civil servants were to be selected on the basis of competence—that is, their ability to perform the job as determined by open, competitive exams—and political neutrality. In so doing, qualified persons would be hired (efficiency and effectiveness) in a systematic fashion (efficiency and economy), and much of the abuse would be eliminated (economy). In other words, the Pendleton Act was an attempt to combat

the consequences of the spoils system’s emphasis on political values by establishing a civil service system emphasizing managerial values.

The Pendleton Act did not mark the end of managerial values’ importance to the civil service. Indeed, the desire to take political considerations out of the public HR equation and to make the civil service more “business-like” was a central theme in the evolution of civil service systems during the twentieth century, and it remains a salient feature today. The Classification Acts of 1923 and 1949, for example, established position classification as a key personnel function. Position classification systematically organizes government jobs according to the nature of work done and the knowledge, skills, and abilities required. Ideally, this represents an efficient way to specify what work is to be done, ensure that people doing similar work are treated similarly, establish the basis for reward, and establish career paths for civil servants. The passage of the Hatch Act at the federal level and “Little Hatch Acts” at the subnational level are also indicative of attempts to remove politics from the “administrative” questions of the civil service. In general, the desire to remove politics from civil service decisions and instead to base them on technically efficient, effective processes has been a defining characteristic of the development of recruitment, hiring, compensation, and promotion systems.^[1]

Legal Values

Finally, the legal perspective emphasizes values like employee rights and equity. These values have steadily gained importance in the United States, especially since the demise of the “doctrine of privilege.” This doctrine held that government employment was a privilege—not a right—and that governments could set strict conditions on public employment. Public employees, in turn, “voluntarily” agreed to these restrictions when they accepted a government position. In practice, this meant that civil servants had no recourse when they suffered an adverse personnel action (e.g., being denied a job or promotion, receiving a reprimand or demotion, being separated) for exercising rights they normally enjoyed as citizens (e.g., free speech, freedom of expression, etc.). With the fall of this doctrine by the 1970s, many of the rights and liberties public employees enjoyed as private citizens were extended into the public workplace. So, for example, an employee that blows the whistle on agency wrongdoing can no longer be dismissed for disloyalty; rather, such expression is a protected right.

Due process is another important employee right. In *Board of Regents v. Roth*, the Supreme Court held

^aThe Pendleton Act of 1883 eroded patronage by creating a limited merit system for the federal government. The merit system’s coverage was gradually expanded over time. The Hatch Act of 1939 (as amended) places additional limits on the partisan activities of both public employers and employees. The courts have also played an important role in curbing patronage. Notable cases include *Elrod v. Burns* (1976), *Branti v. Finkel* (1980), and *Rutan v. The Republican Party of Illinois* (1989).

that public employees can establish property interests in their government positions. Since the Constitution, through the fifth and 14th amendments,^b establishes that no person can be denied life, liberty, or *property* without due process of the law, civil service systems at all levels of government had to be modified to afford employees due process. The practical result has been the creation, in many instances, of elaborate employee grievance and appeals procedures designed to ensure the protection of these rights.

Finally, the legal perspective values the equitable treatment of all applicants and employees. In other words, similarly situated applicants or employees should be treated in a similar fashion regardless of their gender, race, age, ethnicity, religious beliefs, or physical abilities. So, for example, when the federal government's Professional and Administrative Career Examination (PACE) was found to have an adverse impact on minority applicants, the federal government was forced by the courts to take remedial action. While PACE represented an attempt to make the selection of entry-level professional and career candidates more efficient (thus, maximizing managerial values), its demise represented the importance of legal values (equity) to the civil service. As a matter of practice, civil service regulations and procedures emphasizing legal values often come at the frustration of managerial values.

Rethinking the Purposes of Civil Service

Returning to the basic purposes mentioned at the outset, the values emphasized in civil service systems have profound effects on how individuals are recruited and selected for government positions (through patronage or competence?), how employees are rewarded and developed (through political loyalty or on-the-job performance?), and how employees are directed toward meeting public objectives (through overt political direction or through elaborate position descriptions?). Thus, the basic purposes of civil service are inextricably linked to managerial, political, and legal purposes and the administrative values they encompass. Since there has yet to be an agreement on what an appropriate balance is among the various administrative values,^[5] there is little question that civil service reform will continue to be a feature of the American political landscape.

^bDue process for federal government employees is established constitutionally by the 5th amendment and statutorily by the Lloyd-LaFollette Act of 1912. Due process for state and local government employees is established constitutionally through the 14th amendment.

THE RECENT FACE OF REFORM: CIVIL SERVICE REINVENTION

Civil service reform has proceeded in an ad hoc fashion in the United States. As particular problems have arisen, the tendency has been to enact laws, rules, and regulations designed to prevent those problems from occurring again. Typically, scant attention has been given to the cumulative effect of adding rules and regulations on top of existing rules and regulations. In many jurisdictions, this has resulted in the creation of rigid, complex, and ineffective civil service systems (For example, see Ref.^[6]). Critics cite a litany of problems to illustrate this, including lengthy and bureaucratic recruitment processes that repel quality applicants; labyrinthine grievance and appeal processes that stifle managers' ability to deal with poor performers; rigid and complex classification and compensation systems that make it difficult to assign work and reward performance; and central HR offices that emphasize regulatory compliance over HR effectiveness.

It is from this general depiction of public sector HR management that the most recent civil service reform efforts have emerged, largely under the broad "reinventing government" banner. The reinvention moniker is a reference to the work of David Osborne and Ted Gaebler, whose 1992 book, *Reinventing Government*, touched off a wave of interest in administrative reform, not just of civil service, at all levels of government. In fact, many aspects of reinvention mirror administrative reform efforts in other parts of the world (for example, in Canada, New Zealand, and Great Britain). Generally, reinvention-type reforms undertaken attempt to dramatically streamline regulatory burdens while simultaneously decentralizing decision-making authority to managers in line agencies. "Let managers manage" is the common rallying cry, reflecting reformers' desire to improve the quality of services government delivers to citizens (or "customers," in reinventing lingo).

Because an effective HR function is intimately related to government's ability to perform, civil service reform has become a prominent feature of reinvention. Among the key civil service reinvention themes are deregulation, decentralization, a shift to a "service" orientation for central HR offices, and privatization.

Deregulation

In attempting to combat the abuses and corruption associated with patronage, civil service systems developed rules and procedures that emphasized merit and neutral competence. While largely successful in battling problems associated with patronage,

proponents of reinvention argue that this has come at a significant cost to managerial flexibility and, ultimately, government performance. Thus, deregulation entails removing administrative rules and procedural requirements that are thought to impede performance. Deregulation can affect almost every aspect of HRM. For example, in the area of recruitment and selection, deregulation might include eliminating requirements for all job applicants in a given jurisdiction to use the same standardized application or to take a written exam. Instead, applicants might be allowed to submit résumés containing information more directly related to the jobs they are applying for. Similarly, deregulation can come into play in employee discipline and termination. As mentioned above, managers often feel constrained by the inability to effectively deal with poor performers or non-performers. Streamlining (i.e., “deregulating”) grievance and appeals procedures by consolidating reviews and relaxing management’s evidentiary burdens attempts to balance employees’ due process rights with government’s desire to effect employment actions in a timely, efficient fashion. Finally, in compensation, deregulation might provide agency managers the freedom to set or adjust employee pay within broad pay bands based upon prevailing market wages and an employee’s skill and on-the-job performance (this, as opposed to rigid classification-based pay schedules and step increases). As these examples show, deregulation attempts to remove “red tape” that is thought to undermine HR management effectiveness and efficiency.

Decentralization

Decentralization is another aspect of reinvention. It entails parceling out authority and discretion over many important HR functions from central HR offices to managers in line agencies. The basic idea is to empower managers responsible for producing programmatic results by giving them more control over HR management. For example, decentralizing hiring authority to agencies allows managers to formulate job criteria, control the evaluation of applicants, and hire qualified individuals. The rationale for this is straightforward: managers in line agencies have greater technical knowledge about their vacant positions than personnelists working in central HR offices; thus, they are in a better position to ensure the recruitment and selection process results in hires that meet their respective agencies’ needs. Moreover, placing HR management authority in the hands of the agencies actually doing the hiring, promoting, training, etc., expedites these processes. In general, a decentralization strategy

seeks to increase managerial authority over HR management, thereby enabling managers to utilize it as an effective performance improvement tool.

Service Orientation

This aspect of reinvention is directly related to the preceding two. Traditionally, central HR offices assumed responsibility for creating HR policy, administering various HR functions (e.g., classification, competitive exams, training, and development), and for ensuring regulatory compliance through agency oversight. Proponents of reinvention argue that these central “counterbureaucracies,” as they are sometimes called, have the tendency to act as “guardians” of the merit principle and the merit system. From the reinvention standpoint, this represents a form of “goal displacement,” in that the true goals of the civil service system (e.g., attracting and retaining the most qualified workforce) are obscured by an overemphasis on ensuring adherence to rules. In contrast, a service orientation would have central HR offices focus on providing assistance and advice to line agencies, as these agencies assumed increased responsibilities for the day-to-day management of HR functions within a decentralized, deregulated environment.

Privatization

A final aspect of reinventing the civil service is privatization. While there are many varieties of privatization, a common approach is for government to contractually “outsource” services to non-profit or private (for-profit) organizations. In other words, government makes a provision for some “public” service or good, but the actual production is left to a contracted entity. Free of the civil service’s regulatory burdens and armed with the ability to design service delivery, private contractors are thought to be able to provide better services at lower costs. Importantly, privatizing also sends the politically symbolic message that big government—as exemplified by the number of civil servants—is being cut. Privatization of public services like prisons and solid waste collection has been around for decades. Importantly, however, privatization is not isolated to programmatic areas. Indeed, traditional HR functions like payroll and benefits administration, workforce training and development, employee assistance programs, and employment testing have been targeted for privatization. Proponents argue that outsourcing these “housekeeping” activities frees HR offices to focus on their strategic and advisory roles.

CONTEMPORARY EXAMPLES OF REFORM AND REINVENTION

Recent examples of civil service reinvention are plentiful. At the federal level, reinvention activities were initiated under the Clinton Administration's National Performance Review (NPR). In *Reinventing Human Resource Management*, a report that accompanied the original NPR, the administration specified a number of goals for federal HR management (Table 1).

Table 1 Reinventing Human Resource Management Under the National Performance Review

Enable managers to create and maintain a quality, diverse workforce.

- Create a flexible and responsive hiring system.
- Reform the general schedule classification and basic pay system.

Enable managers to empower, develop, train, reward, and discipline employees.

- Authorize agencies to develop programs for improvement of individual and organizational performance.
- Authorize agencies to develop incentive reward and bonus systems to improve individual and organizational performance.
- Strengthen systems to support management in dealing with poor Performers.
- Clearly define the objective of training as the improvement of individual and organizational performance; make training more market driven.

Enable employees to manage work and family responsibilities.

- Enhance programs to provide family-friendly workplaces.

Hold managers accountable for adherence to principles of merit and equal opportunity.

- Improve process and procedures established to provide workplace due process for employees.
- Improve accountability for Equal Employment Opportunity goals and accomplishments.
- Improve interagency collaboration and cross-training of human resource professionals.

Create a system that is self-renewing and continually improving.

- Strengthen the Senior Executive Service so that it becomes a key element in the government-wide culture change effort.
- Eliminate excessive red tape and automate functions and information.
- Form labor-management partnerships for success.
- Provide incentives to encourage voluntary separations

Source: National Performance Review. *Reinventing Human Resource Management*; Accompanying Report of the National Performance Review. National Performance Review: Washington, DC, 1993.

The NPR recommended a variety of specific initiatives designed to accomplish these broad reinvention objectives. In many instances, these recommendations were followed. For example, the NPR resulted in several instances of deregulation. The most visible was the elimination of the *Federal Personnel Manual*: the administration went so far as to use a wheelbarrow to ceremoniously discard it in a dumpster.^[7] In another example of deregulation, the OPM also eliminated the requirement for all federal agencies to use a standard application form (the form is known as SF-171). The NPR also accomplished decentralization—as of 1998, virtually all of the federal government's hiring authority has been decentralized to line agencies.^[8]

An important question is whether the reinvention reforms initiated by the Clinton Administration will last under the administration of George W. Bush. There is early evidence that many broad reinvention themes will, in fact, be continued. In the *President's Management Agenda*, for example, President George W. Bush calls for the flattening of agencies to make them “citizen centered,” for the contracting out of “commercial positions,” and for agencies to more fully utilize the existing HR flexibilities already available. Additionally, the administration will be searching for other ways “to enhance management flexibility, permit more performance-oriented compensation, correct skills imbalances, and provide other tools to recruit, retain, and reward a high quality workforce.”^[9]

Not all reinvention activity—or perhaps even the most impressive—has occurred at the federal level. The most far-reaching, perhaps radical,^[10] reinvention has occurred at the state level. Georgia is the prime example. In 1996, under the direction of then Governor Zell Miller, the state effectively eliminated its traditional merit system. The justification for this change was classic reinvention. According to Miller, the state's merit system was outdated, overregulated, and provided cover for bad workers instead of rewarding good workers (As quoted in, see Ref.^[11]). The reform adopted, Act 816, contained deregulation and decentralization components aimed at speeding hiring and disciplinary processes, providing agency managers greater HR flexibility, and reorienting the state HR office to an advisory and service role. The initial impacts are only beginning to surface, and they appear mixed. For example, there is some evidence that deregulation and decentralization have allowed cronyism and favoritism to creep back into the state's HR functions, and while most state workers in Georgia agree that the reforms have expedited certain HR processes, most do not perceive an increase in overall HR effectiveness (see Ref.^[12]).

Wisconsin is another state that has been widely recognized for recent civil service reforms. Responding to many of the same frustrations as Georgia, Wisconsin

undertook a number of civil service reforms during the 1990s. Under the direction of Robert Lavigna, the state's merit system administrator at the time, Wisconsin instituted a number of reforms intended to improve the state's HR function. Examples include decentralizing HR authority, establishing walk-in testing procedures for some administrative positions, and creating an on-line jobs bulletin service. The state also created two specialized recruitment programs: the Critical Recruitment Program and Entry Professional Program. The Critical Recruitment Program gives agencies the flexibility to quickly hire workers in competitive areas like engineering and health care. Similarly, the Entry Professional Program eschews traditional written exams in favor of more flexible recruitment techniques so that qualified college graduates can be recruited and hired more quickly. These reforms, which incorporate reinvention themes like decentralization and deregulation, appear to have improved the civil service system in the state. Specifically, the state reports that hiring is more timely, applicant pools are better qualified, workforce diversity has improved, and HR costs have decreased (These reforms and their outcomes are described in Ref.^[13]).

Most recently, the state of Florida announced its intention to privatize major portions of its HR function. In September 2001, the state issued an "intent to award" for a \$40 million per year contract with a private firm, Convergys, Inc. The scope of work to be performed by the contractor, as indicated in the invitation for bids issued by the state, is broad. Included are many transactional processes like advertising vacancies, assisting in employee training and development, collecting time sheets, processing new appointments, administering payroll and benefit systems, and maintaining employee personnel records.^[14] Obviously, shifting these functions to a private sector provider represents a large-scale HR experiment. The rationale for conducting it is economical and political. Economically, the state figures that it can save the \$80 million needed to update its antiquated personnel computer system, plus an estimated \$25 million per year in personnel salaries over the 5- to 7-year life of the contract. Politically, privatizing HR represents the priorities of Governor Jeb Bush, who ran on the pledge to make Florida's government "leaner." It is worth noting that the state plans to maintain control over critical areas like employee selection, discipline, grievances, and collective bargaining. Still, the scope of the proposed outsourcing is enormous.

CHALLENGES OF REINVENTED CIVIL SERVICE

The direction that civil service reform and reinvention has taken in recent years, as exemplified above, leads

to a number of important challenges. Among these are ensuring that the requisite HR management knowledge, skills, and abilities are present in central HR offices and line agencies; addressing equity issues associated with the removal of standard, centralized HR processes; and preventing the potential political abuses from materializing. How these challenges are addressed will be telling to the next, perhaps inevitable, wave of civil service reform.

HR Management Expertise

First, a decentralized personnel system means that expertise formerly found in central HR offices must now be disseminated throughout the line agencies of government. Managers and HR generalists in these agencies must possess the knowledge, skills, and abilities needed to effectively and legally assume a range of HR responsibilities, potentially including recruitment, selection, performance appraisal, classification, compensation, and training. Some solace is offered by reinvention's call to have central HR offices move to consultative roles in which they offer advice and assistance to line agencies. This, however, entails new knowledge, skills, and abilities on the part of those in central HR agencies, for example, they need to become skilled consultants with a breadth of HR knowledge as opposed to specialized personnel experts. If either of these necessities fails to materialize, then the effectiveness, efficiency, and perhaps legality of public HR management practices may be called into question.

Equity

A second major concern related to civil service reinvention is equity. Traditionally, a primary responsibility of central HR offices has been to ensure equity. Examples include ensuring open competitive exams, classification and compensation systems that award equal pay for equal work, and consistent procedures for employee grievances. When control is decentralized and the processes are deregulated, there remains a real possibility that similar cases will be treated in a dissimilar fashion. Such a possibility has been raised by many in the public administration community who warn that public employees have come to expect equitable treatment (See for example, Ref.^[15]) if civil servants experience disparate treatment as a result of where (i.e., which agency) or for whom (i.e., which manager) they work, then there will likely be a backlash that favors the reassertion of standardized, centralized controls. In other words, managerial flexibility may run into renewed demands for equity.

Politics

Finally, reinvention has to contend with the “politics problem” (For example, see Ref.^[16]). By removing HR rules and regulations and by decentralizing authority, a reinvented civil service is one that is more open to political and partisan abuses. As mentioned above, the historical development of the civil service placed a premium on combating such problems. When rules are relaxed or eliminated and when HR authority is decentralized, there is a greater likelihood that political officials will use their positions of control and influence to reward friends and political supporters. If and when such political favoritism or cronyism occurs, the result will likely be new calls for reform—reforms that echo traditional antipatronage themes. It is important to recall, however, that not all civil service protections are embodied in civil service laws and regulations. Indeed, the courts have placed severe limits on the ability of government officials to “clean house” on purely political grounds. Similarly, widespread fixation on government performance means that political figures have an incentive to use their HR authority to hire competent individuals who can help government “do more for less.” Still, the possibilities of political intrusions into HR decision making raise the chances that reinvention will encounter difficulties.

CONCLUSION: MORE REFORM AND REINVENTION

Any description of civil service reform and reinvention must be viewed as prologue. Indeed, civil service reform has become “virtually a national pastime.”^[17] At this point, it is difficult to say what the full effects of civil service reinvention will be, as it will take many more years, perhaps decades, for them to materialize. However, a dynamic environment, influenced by shifting political, economic, and social forces, virtually ensures that the nature of civil service’s problems will fluctuate, as will the administrative values civil service systems emphasize.

REFERENCES

1. Ingraham, P.W. *The Foundation of Merit: Public Service in American Democracy*; The Johns Hopkins University Press: Baltimore, MD, 1995.
2. Kaufman, H.F. Emerging conflicts in the doctrines of public administration. *Am. Polit. Sci. Rev.* **1956**, *50* (4), 1057–1073; and Thompson, F.J. Managing within civil service systems. In *Handbook of Public Administration*; Perry, J.L., Ed.; Jossey-Bass Publishers: San Francisco, 1989; 359–372.
3. Rosenbloom, D.H. Public Personnel Administration and Collective Bargaining. In *Public Administration: Understanding Management, Politics, and Law in the Public Sector*, 4th Ed.; McGraw-Hill: New York, 1998; 207–273.
4. Van Riper, P.P. *History of the United States Civil Service*; Row, Peterson and Co.: Evanston, IL, 1958; Cayer, N.J. Evolution of the personnel system. In *Public Personnel Administration in the United States*, 3rd Ed.; St. Martin’s Press: New York, 1996; 16–41; and Ingraham, P.W. *The Foundation of Merit: Public Service in American Democracy*; The Johns Hopkins University Press: Baltimore, MD, 1995.
5. Thompson, F.J. Managing Within Civil Service Systems. In *Handbook of Public Administration*; Perry, J.L., Ed.; Jossey-Bass Publishers: San Francisco, 1989; 359–372.
6. Ban, C. *How do Public Managers Manage?* Jossey-Bass Publishers: San Francisco, CA, 1995; Jorgensen, L.; Fairless, K.; Patton, D. Underground merit systems and the balance between service and compliance. *Rev. Public Pers. Adm.* **1996**, *16* (2), 5–20; Cohen, S.; Eimicke, W. The overregulated civil service: The case of New York City’s personnel system. *Rev. Public Pers. Adm.* **1994**, *14* (2), 10–27; Kettl, D.F.; Ingraham, P.W.; Sanders, R.P.; Horner, C. *Civil Service Reform: Building a Government that Works*; The Brookings Institute: Washington, D.C., 1996. Several reform commissions have also singled-out civil service systems as a major source of government’s performance problems, including The Volcker Commission (1989), the Winter Commission (1993), and National Performance Review (1993).
7. Gore, A. *The Best Kept Secrets in Government*; The Fourth Report of the National Performance Review; Government Printing Office: Washington, DC, 1996.
8. Ban, C. The national performance review as implicit evaluation of CSRA: building on or overturning the legacy? In *The Future of Merit*; Pfiffner, J.P., Brook, D.A., Eds.; John Hopkins Press: Baltimore, MD, 2000; 57–80.
9. Executive Office of the President, U.S. Office of Management and Budget. In *The President’s Management Agenda: Fiscal Year 2002*; Office of Management and Budget: Washington, DC, 2002; 13.
10. Condrey, S.E. Georgia’s civil service reform: a four year assessment. In *Radical Reform of the Civil Service*; Condrey, S., Maranto, R., Eds.; Lexington Books: Lanham, MD, 2001; 179–194.
11. Nigro, L.G.; Kellough, E.J. Civil service reform in Georgia: Going to the edge? *Rev. Public Pers. Adm.* **2000**, *20* (4), 41–54.
12. Condrey, S.E. Georgia’s Civil Service Reform: A Four Year Assessment. In *Radical Reform of the Civil Service*; Condrey, S., Maranto, R., Eds.; Lexington Books: Lanham, MD, 2001; 179–194 and Nigro, L.G.; Kellough, E.J. Civil service reform in Georgia: going to the edge? *Rev. Public Pers. Adm.* **2000**, *20* (4), 41–54.

13. Lavigna, R. Creating a responsive personnel system. *PA Times* **1996**, 19 (11), 10–14; and Lichster, J.E. Reinventing personnel systems. *State Government News* **1997**, 40 (8), 25, 31.
14. This information is drawn from the state of Florida's website: http://www.myflorida.com/dms/hrm/hrout/HR_Outsourcing_Project.html (accessed on September 2001).
15. Peters, B.G.; Savoie, R.J. Managing incoherence: The coordination and empowerment conundrum. *Public Adm. Rev.* **1996**, 56 (3), 281–290; Kellough, E.J. The reinventing government movement: A review and critique. *Public Adm. Q.* **1998**, 22 (1), 6–20.
16. Kellough, E.J. The reinventing government movement: A review and critique. *Public Adm. Q.* **1998**, 22 (1), 6–20; Ban, C. The national performance review as implicit evaluation of CSRA: building on or overturning the legacy? In *The Future of Merit*; Pfiffner, J.P., Brook, D.A, Eds.; John Hopkins Press: Baltimore, MD, 2000; 57–80; and Thompson, F.J. The winter commission report: is deregulation the answer for public personnel management? *Rev. Public Pers. Adm.* **1994**, 14 (2), 5–9.
17. Cayer, N.J. Merit System Reform in the States. In *Public Personnel Administration: Problems and Prospects*, 2nd Ed.; Hays, S., Kearney, R., Eds.; Prentice Hall: Englewood Cliffs, NJ, 1990; 263.

Collaborative Governance

David E. Booher

*Center for Collaborative Policy, California State University Sacramento,
Davis, California, U.S.A.*

INTRODUCTION

Collaborative governance refers to practices that engage stakeholders, citizens, and public agencies in joint deliberations on policy decisions. Collaborative governance has special features that distinguish it from traditional government. At least five examples can be identified from many substantive fields of public policy such as environment, health, and city planning. These include consensus building, community visioning, negotiated rule making, network structures, and public involvement. Implementation of collaborative governance presents problems because of the differences between collaborative governance and traditional governance. Hence, collaborative governance often emerges as an adjunct to policy making by government agencies. Scholars believe that collaborative governance practices have emerged in response to contemporary conditions that affect policy making. The future for collaborative governance depends upon the development of theory and the pursuit of more research.

FEATURES OF COLLABORATIVE GOVERNANCE

Collaborative governance has at least five special features as summarized in [Table 1](#).

Collaborative governance ideally includes the full array of legitimate stakeholders concerned about the policy issue who recognize their interdependence in finding solutions. The process is self-organizing, with the participants setting their own ground rules, agenda, and decision-making rules. Participant deliberations are unconstrained by the convener and the status quo, and all assumptions can be challenged. The dialogue is carefully managed to ensure that all are heard and respected. The participants engage in joint fact finding. Finally, the task is meaningful to the participants and there is clarity about how decisions will be implemented.

Differences between traditional governance and collaborative governance can be highlighted for 11 key dimensions, as summarized in [Table 2](#). These differences create the issues for the implementation of collaborative governance that are discussed below.

EXAMPLES OF COLLABORATIVE GOVERNANCE PRACTICE

There are at least five examples of collaborative governance practice in wide use today: consensus building, community visioning, negotiated rule making, network structures, and public involvement.^[1]

Each of these examples typically occurs in different stages of the policy process as illustrated in [Fig. 1](#). Collaborative public involvement is used to inform policy analysis. Consensus building and visioning often are used to formulate policy for implementation by public agencies. Negotiated rule making and network structures are often used to implement policy. However, in some cases, deliberations in network structures may lead to the formulation of policy, as organizations find that they must seek changes to policy for effective implementation.

Consensus building brings all relevant stakeholders together to seek agreements regarding specific public policy issues in situations where traditional governments have been unable to craft solutions, resulting in a stalemate. These processes are often convened by affected public agencies.^[2] Once agreement has been reached, the consensus building process may end, and implementation may be turned over to traditional institutions. However, stakeholders often monitor implementation and respond to changed conditions. For example, they may transform into another example of collaborative governance, such as a network structure.

Community visioning is a process whereby members of a community build consensus on a description of the community's desired future and on actions to help create goals for the future reality.^[3] This process is usually followed by the organization of working groups to carry out specific steps and to monitor implementation of results. Similar to consensus building, a community visioning process may be transformed into a network structure.

Negotiated rule making is a process in the United States whereby a regulatory agency brings stakeholders together to engage in negotiation with each other and with regulatory agencies regarding the promulgation of a new regulation.^[4] If an agreement is achieved, the regulatory agency then carries out

Table 1 Features of collaborative governance

Features	Summary description
Participants	Full diversity of legitimate stakeholder representatives related to the issue and dependent on other stakeholders for resolution
Process	Self organizing with participants setting ground rules, agenda, and decision rules
Deliberations	Unconstrained by convener. Status quo and all assumptions can be challenged. Dialogue focuses on interests and the problem, not positions and avoids ad hominine attacks. Dialogue is managed to assure all are heard and discourse is sincere, accurate, and comprehensible
Information	Joint fact finding with all information accessible and fully shared among participants
Task	Meaningful to participants with clarity about how decisions will be implemented

the approval of the new regulation through the traditional process. Negotiated rule making is especially common for environmental regulation. The U.S. Environmental Protection Agency (EPA) has been using negotiated rule making since the 1980s, and in 1990, two new federal laws, the Negotiated Rulemaking Act and the Administrative Dispute Resolution Act, were enacted to enable these processes.

Network structure refers to a configuration that brings together affected parties, often including both government agencies and NGOs, to work together to address an ongoing problem or issue of mutual concern that they cannot individually address.^[5] A network structure is not focused on reaching agreement on a specific policy dispute, but rather seeks to build capacity to accomplish shared goals. In this way, a network structure goes beyond mere communication and coordination. The participants take part in a new entity, taking on activities that are more than the simultaneous tasks of independently operating organizations. They have a common mission, collaboration

norms, and a unique structural arrangement that is often provided for in a memorandum of agreement.

Public involvement refers to the processes for engaging the public in policy decision making that go beyond traditional processes of notice, comment, and public hearings typical in the United States.^[6] It facilitates dialogue among the public and with decision makers about options for policy, and it ensures follow up to keep the public informed about the final decision and how the decision has been influenced by the involvement.

Collaborative governance typically proceeds through a series of four steps, as summarized in Table 3. Although these steps occur sequentially, it is not uncommon for the steps to be recursive. For example, upon entering into deliberation, participants may discover that additional stakeholders must be involved, causing them to revisit the organization step. Or, during deliberation, they may discover that more information is needed, and thus return to the analysis step. Likewise, during implementation, new conditions

Table 2 Comparison of traditional governance and collaborative governance

Governance dimension	Traditional governance	Collaborative governance
Structure	Hierarchical	Networks
Source of direction	Central control	Coordination
Goals	Clear with defined problems	Various and changing goals and problems
Origin of system behavior	Determined by roles of participants	Determined by relational interactions of participants
Role of manager	Organization controller	Mediator, process manager
Managerial tasks	Planning & guiding organization process	Guiding interactions, providing opportunities
Managerial activities	Planning, designing, and leading	Engaging participants and resources, influencing conditions
Leadership style	Directive	Generative
Criteria for success	Attainment of the goals of formal policy	Realization of collective action
Organizational context	Single authority	Multiple authorities
Source of democratic legitimacy	Representative democracy	Deliberative democracy

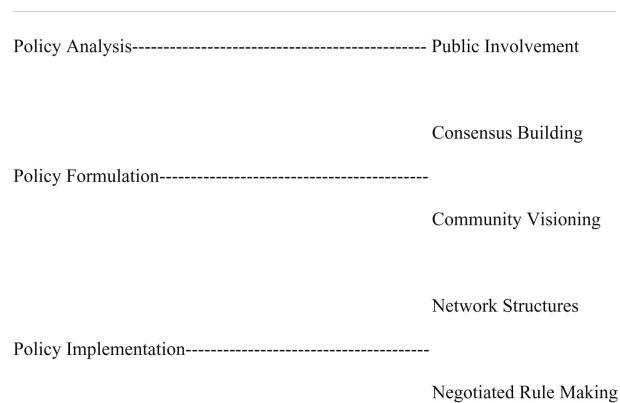


Fig. 1 Location of collaborative governance examples in policy process stages.

may emerge that require returning to the deliberation step. In addition, each context for collaborative governance is unique, and hence, specific activities may vary to reflect the situation.

IMPLEMENTATION OF COLLABORATIVE GOVERNANCE

As shown in Table 2, collaborative governance requires different structures, ideas about goals and the determination of system behavior, managerial functions, leadership styles, and organizational context than traditional governance. Collaborative governance may also imply a different ideal for the source of democratic legitimacy than traditional governance. These differences may create tensions between the practices of traditional governance and collaborative governance. Therefore, the most prevalent obstacle to implementation of collaborative governance is the way the existing institutional context is structured around adversarial methods for collective decision making. Collaborative governance often conflicts with

the political and bureaucratic styles that define much public policy practice.^[7] Moreover, collaborative governance requires a willingness to transcend traditional debates about agency discretion and to experiment with nontraditional forms of accountability.^[8]

As a result, collaborative governance practices are often created as an adjunct to public agencies. For example, consensus building and visioning are sponsored by public agencies seeking to formulate policy that will be widely agreed to by stakeholders and citizens. Participants may find that, to be successful, they must work with both the different features of collaborative governance and the traditional features of public agencies.

Another important obstacle to implementation of collaborative governance is presented by activists who argue that significant injustices and social harms exist in society that must be confronted by adversarial activism. These activists fear that collaboration may prevent this confrontation if processes are not inclusive, or they set unacceptable constraints on the terms of collaboration. They also argue that participants are influenced by a common societal discourse based on a product of structural inequality.^[9] As a result, some activists may be reluctant to participate in collaborative governance.

Collaborative governance has been shown to have outcomes such as creating innovative policies, eliminating stalemates, and producing political, social, and intellectual capital that enable future problem solving.^[10] Participants often report satisfaction from the processes and the expectation that they will use such processes in the future.

Finally, collaborative governance is not appropriate for all policy making. For example, for issues in which there is little complexity, interdependence, or uncertainty or for issues in which the goals and the authority to act are clear, traditional governance may be more effective and efficient. In addition, the features of collaborative governance are critical to achieving the outcomes promised. Failure to involve all stakeholders in a self-organizing process of deliberation and joint fact finding may produce few or even undesirable

Table 3 Steps for a collaborative governance process

Step	Typical activities
Organization	Assess whether collaboration is appropriate. Engage stakeholders or public. Design process. Provide resources
Analysis	Develop mutual understanding of concerns and issues. Develop common information base. Agree on issues to be addressed in deliberations
Deliberations	Agree on problem(s). Generate alternatives. Develop evaluation criteria and evaluate alternatives with criteria. Decide on alternative(s) to recommend or pursue
Implementation	Identify tasks and responsibilities to implement decisions. Determine methods to monitor results and respond to changed conditions

outcomes and increase public cynicism, and may be seen as cooptation.^[11]

WHY IS COLLABORATIVE GOVERNANCE EMERGING?

Scholars have argued that five societal conditions for contemporary policy are important to understanding the emergence of collaborative governance.^[12] First, new spaces are being created for governance. In this spatial metaphor, traditional government institutions, with their specific jurisdictions for policy issues and geography, are seen as spaces in which policy making historically occurs. Today, these spaces are being augmented by configurations of multilevel governance among these institutions and between the institutions and non-governmental organizations (NGOs). Collaborative governance is a practice that occurs in these new spaces in interstices among traditional agencies.

Second, the complexity and rapid change of contemporary society have created an increasing awareness among the public of the limits to centralized hierarchical control by government agencies guided by expertise.^[13] This awareness leads to more uncertainty about policy and a focus on the importance of managing that uncertainty. Collaborative governance may be one strategy to manage uncertainty because it emphasizes joint responsibility, continuous learning, adaptation, and shared capacity for future problem solving in the face of changed circumstances.

Third, society has become more culturally diverse. Solutions to public policy problems require decision makers to deal with an array of publics with different languages, values, perspectives, cognitive styles, and world views. Collaborative governance responds to this challenge by including all relevant publics and by creating processes in which dialogue and deliberation can lead to mutual understanding.

Fourth, officials and the public are increasingly aware of interdependence in policy making as agencies find that they cannot be successful by working unilaterally. While diversity poses challenges to communication and understanding, interdependence creates the need to overcome these challenges. Collaborative governance focuses on interdependence by creating the capacity to interact, share power, and find shared problem definitions and paths to solutions.

Finally, arising out of all of these is the issue of the changing dynamics of trust. Trust in traditional government originates from the legitimacy of constitutional institutions. In the context of new spaces, uncertainty, diversity, and interdependence, this trust can no longer be assumed. Leaders and the public find that, to work across boundaries and across different values and cultures in the face of continuous change,

they must find additional means to engender trust and solve shared problems. Collaborative governance supports proceeding independent of trust and has been shown to engender trust among its participants.^[14]

THE FUTURE OF COLLABORATIVE GOVERNANCE

A central issue for the future of collaborative governance is the need for development of theory and related research. The practice of collaborative governance has evolved without a robust theory. Scholars of many disciplines from around the world are now giving it careful attention. For example, 50 scholars recently published an open call to their colleagues to extend and enhance teaching and research related to collaborative governance because of its importance to the future of democracy.^[15]

A robust theory would provide: (1) the basis for anticipating the probability that collaborative governance might emerge or be appropriate in specific policy contexts; (2) the basis for the design of new institutional forms that incorporate collaborative governance; (3) a framework to evaluate specific collaborative governance methods for specific situations; and (4) a framework to analyze existing research and to design future research for empirical testing.

A theory of collaborative governance needs to address: (1) the conditions when collaborative governance is appropriate in contrast to the conditions when traditional governance is appropriate; (2) the justification for the democratic legitimacy of collaborative governance; (3) the appropriate methods for effective use; and (4) the interaction of collaborative governance with traditional governance.

Some of the areas for future research include representation and equality in collaborative governance processes, impact on policy outcomes, accountability, and paths to institutionalization.

REFERENCES

1. Booher, D. Collaborative governance practices and democracy. *Natl. Civic. Rev.* **2004**, 93 (4), 32–46.
2. Susskind, L., McKernan, S., Thomas-Larmer, J., Eds.; *The Consensus Building Handbook*; Sage: Thousand Oaks, CA, 1999. For an example of one process see the Sacramento Water Forum, <http://www.csus.edu/ccp/projects/recent.htm#waterforum> (accessed March 2006).
3. Moore, C.; Longo, G.; Palmer, P. Visioning. In *Consensus Building Handbook*; Susskind, L., McKernan, S., Thomas-Larmer, J., Eds.; Sage: Thousand Oaks, CA, 1999.

4. Weber, E. *Pluralism by the Rules: Conflict and Cooperation in Environmental Regulation*; Georgetown University Press: Washington, DC, 2001.
5. Keast, R.; Mandell, M.; Brown, K.; Woolcock, G. Network structures: working differently and changing expectations. *Public Adm. Rev.* **2004**, *64* (3), 363–371.
6. Innes, J.; Booher, D. Reframing public participation: strategies for the 21st century. *Plan. Theory Pract.* **2004**, *5* (4), 419–436.
7. Innes, J.; Gruber, J. Planning styles in conflict: the metropolitan transportation commission. *J. Am. Plan. Assoc.* **2005**, *71* (2), 177–188.
8. Freeman, J. Collaborative governance in the administrative state. *UCLA Law Rev.* **1997**, *45* (1), 1–97.
9. Young, I.M. Activist challenges to deliberative democracy. In *Debating Deliberative Democracy*; Fishkin, J.S., Laslett, P., Eds.; Blackwell: Oxford, UK, 2003.
10. Connick, S.; Innes, J. Outcomes of collaborative water policy making: applying complexity thinking to evaluation. *J. Environ. Plan. Manag.* **2003**, *46* (2), 177–197.
11. Booher, D.; Innes, J. Network power in collaborative planning. *J. Plan. Edu. Res.* **2002**, *21* (3), 221–236.
12. Hajer, M.; Wagenaar, H., Eds.; *Deliberative Policy Analysis: Understanding Governance in Network Society*; Cambridge University Press: Cambridge, UK, 2003.
13. Scott, C.J. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*; Yale University Press: New Haven, CT, 1998.
14. On proceeding independent of trust see: Fisher, R.; Ury, W. *Getting to Yes: Negotiating Agreement Without Giving*, 2nd Ed.; Penguin Books: New York, 1991.

On engendering trust see: Leach, W.; Sabatier, P. To trust an adversary: integrating rational and psychological models of collaborative policymaking. *Am. Political Sci. Rev.* **2005**, *994*, 491–503.

15. http://www.csus.edu/ccp/CDN/Call_to_Scholars_-_May_2005.pdf (accessed June 2006).

BIBLIOGRAPHY

1. Bardach, E. *Getting Agencies to Work Together: The Practice and Theory of Managerial Craftsmanship*; Brookings Institution Press: Washington, DC, 1998.
2. Chrislip, D.; Larson, C. *Collaborative Leadership: How Citizens and Civic Leaders can Make a Difference*; Jossey-Bass: San Francisco, CA, 1994.
3. Crosby, B.; Bryson, J. *Leadership for the Common Good: Tackling Public Problems in a Shared-Power World*, 2nd Ed.; Jossey-Bass: San Francisco, CA, 2005.
4. Fung, A.; Wright, E.O. *Deepening Democracy: Institutional Innovations in Empowered Participatory Democracy*; Verso: New York, 2003.
5. Gastil, J.; Levine, P. *The Deliberative Democracy Handbook*; Wiley: New Jersey, NJ, 2005.
6. Roberts, N., Ed.; *The Transformative Power of Dialogue*; Elsevier Science: Amsterdam, 2002.
7. Straus, D. *How to Make Collaboration Work*; Berrett-Koehler: San Francisco, CA, 2002.
8. Yankelovich, D. *Coming to Public Judgment: Making Democracy Work in a Complex World*; Syracuse University Press: Syracuse, NY, 1991.

Collective Bargaining

Robert Hebdon

Faculty of Management, McGill University, Montreal, Québec, Canada

INTRODUCTION

By permitting the formation and legal representation of unions or associations, collective bargaining processes reconcile the competing interests of managers and employees in the workplace. The potential for conflict exists in both private and public sectors—not only over economic issues but also the control of work processes. A third element of collective bargaining in a North American context is the institutionalization of disputes over disciplinary practices and the interpretation of collective bargaining agreements signed by management and unions. Thus, the goals of collective bargaining are to provide orderly and peaceful mechanisms for the resolution of disputes between management and labor over the distribution of economic rewards, the control over work processes, and the interpretation and application of collective bargaining agreements.

SCOPE

By the year 2000, over 37% of the U.S. public sector workforce were members of unions or employee associations. Collective bargaining had become entrenched as a mechanism for providing employee voice in public sector establishments at the local, state, and federal levels. Not all states, however, have collective bargaining laws. Twenty-four states have comprehensive laws that impose a “duty to bargain” on public employers for all public employees—state, teacher, police, fire, and other municipal employees (Table 1). At the other extreme, there are 11 states that have no duty to bargain in the law and five states that have no law at all (Arkansas, Colorado, Louisiana, Missouri, and West Virginia).

Differences Between Public and Private Sector Collective Bargaining

It is widely believed that the private sector model of collective bargaining is inappropriate for the public sector. Consequently, numerous important differences have emerged in the laws that govern two bargaining systems.

Scope of bargaining

Generally, public employee unions are more restricted in the scope of negotiable issues than their private sector counterparts. For example, employees of the federal government cannot bargain over wages or benefits. In the public sector, management’s rights are more broadly defined to include matters of public policy.

Economic context

To some extent, public sector bargaining is insulated from the same market forces that exist in the private sector. By definition, such public goods as public health and safety are offered in a less competitive environment than in the private sector. Although the monopoly of services was seen as a bargaining advantage by public employee unions, particularly in the early years, fiscal conservatism and the privatization threat have eliminated any such advantage.

The right to strike

Unlike private sector collective bargaining, most public employee unions do not have a legal right to strike. As a matter of public policy, the services that public employees provide are viewed as too essential to permit disruptions. Strikes may threaten the health and safety of the public. Even in states where public sector strikes are legal (e.g., Pennsylvania), strikes may be stopped by injunction if they pose a threat to the health and safety of the public.

Four Decades of Bargaining

Since the 1960s, collective bargaining has gone through four distinct phases roughly corresponding to the four decades since then. The first generation represented the growth phase of employment and unions of the 1960s, the second was characterized by the taxpayer revolts of the 1970s, and the third generation in the 1980s involved greater emphasis on the performance and productivity of public services.^[1] In what may be described as a period of heightened aggression against public sector collective bargaining, the 1990s represented the fourth generation of public sector collective bargaining.^[2] Since 1990, public employees have been

Table 1 Legislated bargaining duty

Occupational coverage	Number of states and DC
All	24
All less state	3
Police, fire, and education	2
Education and municipal	2
Education only	5
Police and fire	4
None	11

Source: From Ref.^[6].

increasingly under attack on the related fronts of job security and compensation through privatization and challenges to collective bargaining rights.

Dispute Resolution Procedures

Given that a primary goal of collective bargaining is the peaceful resolution of disputes between management and labor, we examine the evolution of dispute resolution procedures. Most public sector dispute resolution mechanisms were designed over 20 years ago to avoid strikes. The belief was that essential public employees simply could not be allowed to walk off their jobs because of the irreparable harm that might be done to the public and because “excessive” union bargaining power would result in huge wage gains in negotiations. This provided the rationale for extensive intrusions into the collective bargaining process in the public sector, in contrast to the voluntarism of private sector dispute resolution. The result was the current North American legislative patchwork of arbitration,^a fact finding,^b and compulsory mediation^c (Table 2).

In recent years, public sector strikes have declined and new issues of concern have emerged, associated with such factors as wage freezes, layoffs, restructuring, contingent workers, privatization, and employee benefits. In this new environment, it is not surprising that policy makers are reassessing the necessity of these intrusive procedures that focus on avoiding the strike

^aInterest or “contract” arbitration involves the establishment of agreement provisions by a third party; grievance or “rights” arbitration deals with matters of contract interpretation during the life of the agreement.

^bAs the name implies, fact finding is a process that normally involves a formal hearing with briefs and evidence in support of the parties’ positions, followed by a written public report in which the fact finder makes a non-binding recommendation on each issue in dispute. Interest arbitration differs mainly in the private and binding nature of the arbitrator’s (or board’s) award.

^cCompulsory mediation is usually an intermediate stage where a neutral third party acts as an informal facilitator. Unlike its private sector counterpart, it is often a compulsory step in the public sector.

Table 2 Law by terminal resolution procedure

Procedure	Police and fire	State	Education	Municipal
Mediation	5	9	9	9
Fact finding	6	13	19	15
Arbitration	22	5	8	6

Source: From Ref.^[6].

outcomes, not the process for dealing with these new issues. It is entirely possible that both fact finding and arbitration will come under even more scrutiny as their costs and benefits are reevaluated. Public sector dispute resolution procedures are at the center of the assault on public sector bargaining. An important issue, for example, is the extent to which interest arbitration (a dispute resolution procedure adopted when the focus was on prohibiting strikes) is appropriate to the current environment, when the focus is on retrenchment and restructuring.

Recent Trends

Dispute resolution in the public sector adapted to meet the more extreme economic, political, and financial environment of the 1990s. Intensifying fiscal pressures and more conservative political forces combined to create the apparent need for major cost reductions and/or layoffs in many jurisdictions. Paradoxically, this crisis atmosphere created a unique opportunity for labor and management to experiment with cooperative approaches to dispute resolution. Experiments in “interest-based bargaining”^d abound at federal, state, and local levels of government covering a wide range of occupations from blue-collar municipal employees in Wisconsin to clerical and administrative employees in the Federal Department of Labor. In Wisconsin, for example, the Wisconsin Employment Relations Commission (WERC) now offers training to the parties in “consensus bargaining,” a problem-solving approach to negotiations and dispute resolution. The WERC^[3] reported that it trained 25 sets of municipal negotiators (over 525 persons) in this new approach. Interest-based and problem-solving bargaining has also been employed in bargaining at the federal level. For example, the Labor Department and the American Federation of Government Employees negotiated two “major agreements” using a win-win approach (Ref.^[4], p. 506; Ref.^[5], p. 5). Probably the most pervasive

^dInterest-based bargaining, which has many variants (win-win, principled, consensus, mutual gains, cooperative, etc.), attempts to replace the adversarial negotiations that often accompanies collective bargaining with a more cooperative or integrative approach that emphasizes the common concerns of management and labor.

changes in collective bargaining are occurring as a result of reform of the U.S. educational system. As teachers become increasingly involved in educational policy making, more cooperative approaches to bargaining are beginning to appear, although research indicates that this has not spelled the end of adversarial bargaining.

Although numerous strikes has remained low after the steep decline of the early 1980s, other forms of conflict such as grievance arbitrations and unfair labor practices have steadily increased in some jurisdictions. To meet this new challenge, public sector agencies have institutionalized new forms of mediation, particularly grievance mediation.^e

The courts continue to play an inconsistent but important role in shaping dispute resolution procedures Lund and Maranto.^[6] The right to strike has been extended by decisions of state supreme courts to all public employees in Louisiana (1990) and Colorado (1992), but denied in common law to public employees in West Virginia (1990). In addition, interest arbitration awards have been reviewed in the supreme courts of Iowa (1992) and New Jersey (1993), and a binding interest arbitration law was found to be unconstitutional by Nebraska's supreme court (1991).

The antigovernment mood has placed public sector collective bargaining under severe scrutiny. What appeared just a few years ago as acceptable dispute resolution methods are increasingly under assault from taxpayers, employers, and politicians. For example, the utility of interest arbitration is being questioned in Wisconsin, Iowa, New Jersey, and New York. For both fact-finding and interest arbitration, the debate focuses on the appropriate weight to be attached to the criteria of taxpayer concerns and ability to pay.

This reassessment of dispute resolution procedures comes after two or three decades of experience in many states.

CONCLUSION

Despite the challenges of public sector collective bargaining since the 1960s, it remains a popular mechanism for providing employees voice in the workplace. After the growth period in the 1960s, close to 40% of public service employees at the local, state, and federal levels of government are covered by collective bargaining.

REFERENCES

1. Lewin, D.; Feuille, P.; Kochan, T.A.; Delaney, J.T. *Public Sector Labor Relations: Analysis and Readings*; Lexington Books: Lexington, MA, 1988.
2. Hebdon, R. Public sector dispute resolution in transition. In *Public Sector Employment in a Time of Transition*; Belman, D., Gunderson, M., Hyatt, D., Eds.; Industrial Relations Research Association, 1996; 85–125; Ch. 3.
3. Biennial Reports. Wisconsin Employment Relations Commission: Madison, WI, 1992.
4. McKee, J. Beyond litigation: new approaches to federal labor relations—let's give change a chance. In *Proceedings of the IRRRA Meeting, Seattle, WA*; 1993; 506 pp. (Spring).
5. Annual Reports. Federal Labor Relations Authority: Washington, DC, 1992 and 1993.
6. Lund, J.; Maranto, C. Public sector labor law: an update. In *Public Sector Employment in a Time of Transition*; Belman, D., Gunderson, M., Hyatt, D., Eds.; Industrial Relations Research Association, 1996; 21–39.

^eUnfair labor practices refer to complaints to a labor relations administrative agency (e.g., the National Labor Relations Board), either by a union or employer over such matters as bargaining in bad faith or interference and/or coercion in the organizing process. Grievance mediation represents the application of voluntary mediation techniques to grievance disputes as an intermediate step before formal binding arbitration.

Collective Bargaining, The Politics of

Robert Hebdon

Faculty of Management, McGill University, Montreal, Québec, Canada

INTRODUCTION

A concern in the early period of public sector bargaining (1960 to early 1970s) was that public employee unions may have too much power, particularly political power vis-à-vis weaker interest groups.^[1] Subsequent evidence, however, has not supported this fear. Most research has shown that union-non-union wage differentials, for example, are greater in the private sector than the public sector (see Ref.^[2]). Nonetheless, public employee unions have exercised their right to engage in various political activities to significantly increase wages, employment levels, and departmental budgets (summarized in Ref.^[3]; see also Ref.^[4]).

Three broad questions have been the focus of early exploratory research into union political activities in the public sector. The few studies that are available examined the types of political activities at the municipal level and the considerable occupational and union variation. Second, studies examined the factors that explain the variation in type and level of political activity. Finally, some research analyzed the relationship of political action to collective bargaining. For example, do public sector unions compensate for weak bargaining laws by engaging in political action?

After reviewing some theoretical considerations, this paper will examine the nature and scope of union political action, some factors that affect levels and intensity, and the evidence on the relationship between political action and collective bargaining. The paper begins with a brief case study of union political action in New York state.^a

A CASE STUDY OF UNION POLITICAL ACTION

The New York State Public Employees Fair Employment Act (Taylor Act) was amended in 1977 to eliminate fact-finding for police and firefighter bargaining units and to replace it with binding arbitration of disputes. The amendment provided for a 2-year sunset provision. When the 2 years was up, New York state unions and municipalities lobbied the state legislators for their respective positions. The unions, who wanted

to keep arbitration, argued that arbitration contributed to labor peace and produced comparable settlements to those under fact-finding. The employers—the New York Conference of Mayors—wanted to scrap arbitration, because it increased the rate of impasse; it constituted an invasion of the authority of government; it cost a great deal; and it took away the right of the parties to say no to a settlement. For his part, the Governor wanted to keep arbitration but make it binding only on the union. The Public Employment Relations Board (PERB), the neutral agency, recommended keeping arbitration without modification. PERB argued that 2 years was not a long enough period to judge the arbitration process. The police and firefighter unions extensively lobbied the state legislators. They held face-to-face meetings, submitted briefs, appeared before select committees, etc. A compromise was finally worked out when government finally realized that the unions had the votes to renew the arbitration process. The final bill that was passed and signed by the Governor provided for a 2-year extension with minor changes to the arbitration process.

In the final analysis, the unions, employers and their organizations, neutral agencies, and the public, through their elected representatives, engaged in an intelligent debate about the merits of arbitration versus fact-finding as a dispute resolution process. It is true that this case demonstrates the exercise of considerable union political power, but it also shows that conducted in a pluralistic setting, a wide range of legitimate views were exchanged, and in the end, a reasonable outcome was achieved.

THEORETICAL CONSIDERATIONS

Freeman^[6] offers a plausible explanation for the failure of public sector unions to wield excessive political influence and power. While it is true that public sector unions can affect employer behavior through the political process, because they help elect executive and legislative branches, the demand for public sector labor is not necessarily more inelastic than in the private sector. Freeman suggests that the discipline of the budget, the high labor intensity of public services, and the ability of citizens to act as a constraint on public sector

^aBased on a paper by Kochan.^[5]

Table 1 Political actions and dominant collective bargaining law by state 1978–1988

State	Number of cities	Political actions		Law ^a
		Total	Per city	
AK	2	5	2.5	2
AL	4	0	0.00	0
AR	8	8	1.00	0
AZ	7	3	0.43	0
CA	139	259	1.86	0
CO	8	9	1.13	0
CT	40	32	0.80	1
DC	1	6	6.00	0
DE	1	1	1.00	0
FL	52	131	2.52	0
GA	3	5	1.67	0
HI	2	10	5.00	2
IA	16	30	1.88	1
ID	5	15	3	2
IL	60	37	0.62	0
IN	8	18	2.25	0
KS	9	8	0.89	0
KY	4	6	1.50	0
LA	4	7	1.75	0
MA	37	40	1.08	1
MD	8	6	0.75	0
ME	9	9	1.00	1
MI	45	85	1.89	0
MN	40	45	1.13	2
MO	16	22	1.38	0
MS	3	5	1.67	0
MT	4	11	2.75	2
NC	8	8	1.00	0
ND	1	0	0.00	0
NE	7	10	1.43	0
NH	9	5	0.56	0
NJ	64	53	0.83	1
NM	1	1	1.00	0
NV	6	21	3.50	0
NY	29	44	1.52	0
OH	45	53	1.18	0
OK	19	39	2.05	0
OR	14	23	1.64	2
PA	63	23	0.37	2
RI	7	4	0.57	1
SC	2	0	0.00	0
SD	2	2	1.00	0

(Continued)

Table 1 Political actions and dominant collective bargaining law by state 1978–1988 (Continued)

State	Number of cities	Political actions		Law ^a
		Total	Per city	
TN	4	7	1.75	0
TX	44	83	1.89	0
UT	5	12	2.40	0
VA	6	9	1.50	0
WA	8	20	2.50	0
WI	25	51	2.04	1
WV	3	11	3.67	0
WY	3	10	3.33	0
Totals	910	1302		
Mean			1.66	

^aLegend: Arbitration = 1; Right to Strike = 2; No Finality = 0. Note that in this table, the dispute resolution law is based on the dominant state public sector collective bargaining law, but in reality, it varies by occupational group in many states.

wages through such activities as referenda, limit the power of public sector unions. Anderson and Delaney^[7] develop a conceptual framework for the study of political activities. They make a distinction between activities directed at collective bargaining and those devoted to long-run political benefits for union members generally.^b

TYPES AND INCIDENCE OF POLITICAL ACTIVITY

In the summer of 1988, the International Cities and Counties Management Associations surveyed city managers in all U.S. cities over 10,000 population, where some of their employees were in a union or association. The 910 survey respondents provided a rare glimpse into the scope and character of union political activities in America. City managers were asked to list the number and type of union political activity over the previous 10-year period (1978–1988). Table 1 shows the breadth of coverage of political activities throughout the country. In only a few states were there no reports of union political activities (Alabama, North Dakota, and South Carolina), and in these states, there were only one or two cities reporting. In terms of the raw numbers, California, Florida, Michigan, Pennsylvania, and Texas have the highest state totals of political activities. When expressed as per city rates, the District of Columbia, Hawaii,

^bFor a thorough review of the literature on this topic, see Anderson and Delaney.^[7]

Table 2 Union municipal political activities by occupational groups

	No. of cities	Per city rate of electoral politics	Per city rate of lobbying	Per city rate of publicity campaigns	Per city rate of total political activity
Police	804	1.84	0.84	0.59	3.27
Fire	619	2.29	1.06	0.71	4.06
Sanitation	316	2.39	1.02	0.79	4.2
Public works	679	1.85	0.84	0.62	3.31
Other	665	2.01	0.85	0.63	3.49

Nevada, West Virginia, and Wyoming had the most significant levels of political action.

In Table 2, the data are broken down by occupational group (police, fire, sanitation, public works, and other) and by electoral politics (candidate endorsements, financial contributions, and manpower or in-kind support), publicity campaigns (mismanagement disclosure or publicity), state lobbying (including referendum campaigns), and total political actions.

Electoral politics is by far the most popular form of political activity across all occupational groups reported, with sanitation and firefighter unions showing the highest rates per city and police and public works the lowest. In order of declining importance, lobbying and publicity campaign rates were distributed in the same manner across occupational units as electoral politics.

Overall, there were on average between 3.3 and 4.2 incidents of political activity reported by city over the previous 10-year period. It was not surprising that sanitation unions were the most politically active over this period, because they were confronting the issue of the contracting out of their services, and the collective bargaining process was ill-equipped to handle it.^[8] Additionally, high levels of political activity were expected for firefighter unions local, given the rich history of the International Association of Fire Fighters as an American Federation of Labor craft union.

DETERMINANTS OF UNION POLITICAL ACTIVITIES

Anderson and Delaney^[7] show that union political action is inversely related to success in bargaining and weak legislation. They also found that factors that determine union involvement in electoral politics differ from those that determine union involvement in legislative politics. Chandler and Gely^[3] find that electoral politics is a complement to collective bargaining.

Hebdon and Stern^[9] postulate that where the collective bargaining law fails to provide finality, there will be higher than average levels of political action.

In this case, because unions cannot strike or force an agreement through arbitration, greater use of political pressure is expected. Here, publicity campaigns are more likely to be used as pressure tactics as part of the bargaining process. Bargaining laws were shown to have a negligible impact on electoral politics and state lobbying.^[9] Only under final offer arbitration^c is there a higher propensity to engage in electoral politics when compared to right to strike laws, but this is only significant at the 10% level. The unambiguous result is the stronger propensity to conduct publicity campaigns when under a strike ban law compared to right to strike laws. The Hebdon and Stern interpretation of this finding is that unions use mismanagement threats and publicity campaigns as substitute pressure tactics for threatening or organizing a strike. In other words, laws that permit legal strikes channel conflict away from publicity campaigns against public employers to more traditional strike pressures.

Also tested was Anderson's^[10] hypothesis that union political action, necessary for the introduction of the collective bargaining law, diminishes over time as the union-management relationship matures. Hebdon and Stern found no support for Anderson's^[10] hypothesis that political action diminishes as the union-management relationship matures. In fact, political actions increased significantly with the age of the relationship over this 1978–1988 period.

POLITICAL ACTION AND COLLECTIVE BARGAINING

Unlike their counterparts in the private sector, public sector unions may play a direct role in electing their bosses.^[11] Collective bargaining performs a political function in the sense that a system of governance is established. Unlike the private sector bargaining model, in the public sector, it takes on a multiparticipant character that includes taxpayers, unions, politicians,

^cUnder final offer arbitration, the arbitrator must choose the last position of management or of the union.

school boards, employer organizations, and other interest groups.

Thus, to achieve their goals, collective bargaining and political action may present strategic choices for unions in the public sector. There is some evidence from California, for example, that unions in the public sector seem to prefer political action (forming coalitions and lobbying) over industrial action and strikes.^[12] This substitution of political actions for striking also appears in the work of O'Brien,^[4] who showed that the political activities of police and firefighters' unions could be more important than collective bargaining in pay determination. Finally, Hebdon and Stern^[9] showed that a hidden cost of strike ban laws is the redirection of conflict away from strikes to such other expressions as publicity campaigns.

REFERENCES

1. Wellington, H.H.; Winter, R.K., Jr. The limits of collective bargaining in public employment. In *The Unions and the Cities*; The Brookings Institution: Washington, DC, 1971; 12–32.
2. Belman, D.; Heywood, J.S. The structure of compensation in the public sector. In *Chapter in Public Sector Employment in a Time of Transition*; Belman, D., Gunderson, M., Hyatt, D., Eds.; Industrial Relations Research Association, 1996.
3. Chandler, T.; Gely, R. Toward identifying the determinants of public-employee union's involvement in political activities. *Am. Rev. Public Adm.* **1996**, *26* (4), 417–438.
4. O'Brien, K.M. The impact of union political activities on public-sector pay, employment, and budgets. *Ind. Relat.* **1994**, *33* (3), 322–345 (July).
5. Kochan, T.A. The politics of interest arbitration. In *Public Sector Labor Relations: Analysis and Readings*; Lewin, D., Feuille, P., Kochan, T.A., Delaney, J.T., Eds.; Lexington Books, 1988; 76–85.
6. Freeman, R.B. Unionism comes to the public sector. *J. Econ. Lit.* **1986**, 41–53.
7. Anderson, J.C.; Delaney, J.T. The involvement of Canadian unions in political activities: an exploratory analysis. *J. Labor Res.* **1990**, *XI* (4), 361–379.
8. Hebdon, R. Contracting out in New York state: The story the lauder report chose not to tell. *Labor Stud. J.* **1995**, *20* (1), 3–29 (Spring).
9. Hebdon, R.; Stern, R. Do public sector strike ban laws really prevent conflict? *Ind. Relat.* **2003**, *42*, 493–512.
10. Anderson, J.C. Bargaining outcomes: an IR systems approach. *Ind. Relat.* **1979**, *18* (2), 127–143 (Spring).
11. Kearney, R.C. *Labor Relations in the Public Sector*, 2nd Ed.; Marcel Dekker: New York, 1992.
12. Johnston, P. *Success While Others Fail*; ILR Press: Ithaca, NY, 1994.

Community-Based Planning for HIV/AIDS

Bruce W. Kieler

Wharton County Junior College, Wharton, Texas, U.S.A.

Ishak Saporta

Faculty of Management, Tel Aviv University, Tel Aviv, Israel

INTRODUCTION

During the past decade, the devolution of responsibility for health care and social support services programs for persons infected and affected by human immunodeficiency virus (HIV)/acquired immune deficiency syndrome (AIDS), from the federal and state levels to the local level, was accompanied by an increasing utilization of community-based participatory planning entities. These entities became the primary means used for prioritizing services and allocating resources to meet the needs of persons living with AIDS, caused by the HIV, and for implementing HIV prevention programs throughout the United States. These participatory planning entities are also being used in other service areas including mental health, substance abuse counseling, housing, tobacco control, and disease prevention and education programs. They include such organizations as health services planning councils, health care consortia, disease prevention planning groups, tobacco control coalitions, and housing planning committees.

Despite their widespread usage, very little is known about community-based participatory planning entities, their formation, operations, development, and effectiveness over time. The study of community-based participatory planning entities can be based on a variety of theoretical perspectives including small group formation, authority in groups, procedural justice, group decision making, conflict resolution, organizational environment and form, institutionalization of new organizational forms, population ecology of organizations, organizational innovation and change, organizational alignment with the environment, management of strategic alliances, and coordination of services. The goals of this chapter are to examine several of the leading theoretical perspectives that are relevant to the study of the expanding organizational population of community-based participatory planning entities, and to suggest a methodology for conducting case studies of these entities.

BACKGROUND

This article is an update of Chapter 11, “Community-Based Planning,” by Kieler.^[1]

The rapidly changing health care industry in the United States suggests a dynamic and turbulent organizational environment. Into this complex and unpredictable environment, an unusual type of planning organization is being embedded, namely, the community-based participatory planning entity. This type of organization is variously called a community-based planning council, committee, consortium, coalition, or group. Establishment of these new organizations is mandated by both federal and state legislation. The premier examples of these new entities are the HIV health services planning councils mandated by Title I of the Ryan White Comprehensive AIDS Resources Emergency Act of 1990 (i.e., the CARE Act), under Public Law 101-381 as amended, and the HIV health care consortium mandated by Title II of the CARE Act. Another such planning entity is the HIV prevention community planning group (CPG) that is mandated by the Centers for Disease Control and Prevention.^[2] The HIV prevention CPG has responsibility for addressing the needs for education about and prevention of HIV infection in almost all public health jurisdictions in the United States.

Community-based participatory planning entities have many of the characteristics that define a coalition.^[3,4] Definitions of each of these entities, as well as enumeration of their respective roles, responsibilities, and mandates, are usually provided in the specific legislation or program announcements authorizing their use. Specific instructions to local governmental or health jurisdictions about establishing these entities are usually contained in the program guidance issued by the federal or state agency responsible for implementing the program. For example, Minkler^[5] provided guidance in community organization and community building for health, whereas the Academy for Educational Development^[6] and McKay^[7] both provided extensive discussions of the HIV prevention community planning process.

Several papers on Title I planning councils look at: 1) various issues and challenges that confront planning councils;^[8,9] 2) the initial stages of development of a typical planning council;^[10,11] 3) institutionalization and legitimation of a planning council process;^[11] 4) decision making by a planning council;^[11] and 5) organizational environment and organizational form of a planning council.^[11,12] Other papers on the Title I process focus on the types and availability of HIV/AIDS services before and after the allocation of Title I funds,^[13–16] as well as on the impacts of the various CARE Act titles.^[17–23] Several major unpublished manuscripts focus on the implementation of the various CARE Act titles, including a report on challenges facing Title I and Title II planning entities that was submitted to the AIDS Action Foundation in 1993^[24] and a report on the implementation of Title I in six major U.S. urban areas that was submitted to the Kaiser Family Foundation in 1994.^[25]

Funding for the various CARE Act titles^[26–28] and the effectiveness of services provided by these funds^[29–31] are explored in a number of reports and position papers. The general conclusion is that the CARE Act has resulted in substantial additional services and increased access to services for persons living with HIV disease. The recurring challenges that the CARE Act community-based planning entities must face in their decision-making processes are noted both in news reports and published research. These would include reports on contractual disputes, program management, program effectiveness, and use (and misuse) of funds.^[32–34]

In reporting on a multiyear observational study of the Oakland HIV health services planning council, Kieler et al.^[8] noted that, at the time the CARE Act was enacted by the U.S. Congress, the use of community-based participatory planning entities was popular with HIV/AIDS advocates, HIV/AIDS service organizations, and federal officials alike. This model of health care and support services planning provided a way for community-based organizations and public sector agencies, as well as individuals infected and affected by HIV disease, to play key roles in these newly initiated and federally mandated planning councils. Additionally, this model appealed to local agencies because it offered local control over the allocation of funds to meet local needs for health care and support services. It appealed to federal officials because this model placed AIDS constituency groups in the position of having to work collaboratively at the local level to achieve expanded availability of services, rather than constantly appealing to Congress and the federal executive branch in an ad hoc fashion for more funds.

The experiences of a number of planning councils in different parts of the country, most notably the

Oakland HIV health services planning council, suggest that this model of planning does work, but not in the collaborative, rational fashion envisioned in Title I of the CARE Act. As Kieler et al.^[8] reported, local control of the program in the Oakland area was accompanied by interorganizational conflict, accusations of personal and organizational misconduct, frequent challenges to the legitimacy of the process, and a general sense that local political factors played too great a role in priority setting, resource allocation decisions, and contracting for services. Not only in Oakland, but nationwide, local governmental entities and community-based groups have experienced challenges in grass roots decision making regarding implementation of Title I. In his study of planning council decision-making behavior, Slack^[35] applied a political model (the zero-sum model) and a bureaucratic model (the Herbert thesis) to identify the factors triggering conflictual zero-sum behaviors.

However, the Kieler et al.^[8] study found that, in spite of these difficulties, the Oakland planning council was able to distribute funds approximately on schedule to local service providers, and the availability of needed services was increased. The findings of the Oakland study suggest a number of ways that a planning council could be structured to increase its effectiveness while reducing interpersonal and interorganizational stress among service providers that the process induced in the past.

Additional work needs to be done to develop a theory-based literature that will document, explain, and analyze all aspects of the establishment, operation, development, and effectiveness of these newly mandated community-based participatory planning entities, and guide the establishment and development of future participatory planning entities. Such a literature might also serve as the basis for building organizational theory specific to participatory planning entities such as planning councils, consortia, CPGs, and coalitions.

AN INTEGRATED THEORETICAL MODEL

Federal and state programs mandating the use of community-based participatory planning entities in local health care and support services environments are only a part of the major changes that are resulting in the restructuring of health care and health-related services. Scott^[36] observed that the medical care field is a multifaceted and dynamic scene that it is complex and rapidly changing, that health care organizations in the United States have undergone a revolution during the past half century, and that particularly dramatic changes have occurred during the past two decades. "To throw some light on the current situation," Scott

described the nature of the changes that have occurred in health care organizations, particularly those in the medical care delivery system. Scott was convinced that “we will not make much headway in our understanding of complex societal systems until we begin to examine the ways in which institutional and technical environments, organizational fields, [organizational] populations, organizational sets, and individual organizations act and interact to constrain and change each other” (p. 296).

In describing the nature of the changes that have occurred in health care organizations and the medical care delivery system, Scott formulates a general theoretical framework within which to view, interpret, and explain such changes. This framework may provide us with a basis for studying the formation, operations, and organizational development of numerous new organizations, including community-based participatory planning entities, that are emerging in the health care and support services environment.

Scott begins by noting that the various organization theories (e.g., population ecology, strategic management, resource dependency, transactions costs, and institutional theory), which are used by analysts to account for one or another feature of the changing health care scene, “have only recently been developed (most appeared after 1975) and are typically treated by organizational analysts as offering contradictory or competing explanations” (p. 272). Scott feels, however, that each of these perspectives is limited and provides only a partial account of the complex phenomena being observed. Although development of unified theory is beyond the scope of his paper, Scott does propose that “the search for an improved, detailed understanding and for verifiable accounts will be advanced by the development of more integrative frameworks that seek to define where and when—to what types of phenomena and under what conditions—the various theory fragments apply” (p. 272). Scott proposes that the “effort to understand the medical care system should begin at the most comprehensive level—the institutional environment—and then proceed to examine more and more delimited systems and units” (p. 273).

The key features of the organizational analysis model that Scott proposes include three levels, namely, the organizational environments level, the organizational fields/populations level, and the organizational sets/organizations level. For each level, Scott identifies the appropriate theoretical perspective to use in discussing activities at that level.

Scott then identifies the major trends in the development of the U.S. health care sector. These include such features as the increased scale of the medical care system, increased concentration of medical resources devoted to the delivery of health services, and

increased specialization of both individual providers and medical care organizations. These trends also include greater diversification, such as in the range of services and types of clients served. Scott points out that other trends in the health care sector include increased linkages among provider organizations, increased governmental involvement in the health care system, increased privatization, increased managerial and reduced professional influence, and increased market orientation of the health care system.

Within this rapidly changing health care sector are embedded the newly established community-based participatory planning entities to which many must now turn for health care, support services, and prevention/education services.

ORGANIZATIONAL ENVIRONMENT AND FORM

Bidwell and Kasarda^[37] noted that organizational form is not only composed of interpersonal relations of members of the organization, but also contains properties that pertain to the organization conceived of as a collectivity with a unitary character. Some of these properties can be regarded as aggregates of interpersonal relations, such as the division of labor and hierarchy. Others may not be so regarded, such as the size and composition of an organization’s membership, its stock of technological and material resources, its own institutional characteristics (structure, bylaws, rules, policies, and processes), and the technological, physical, demographic, and institutional properties of its environment. The authors caution against using a behavioral approach to a theory of organizational form because of the likelihood that such an approach would yield a theory of “unmanageable complexity and intellectual difficulty.” Instead, they espouse a macrosocial approach that allows the treatment of aggregative properties of the organizational collectivity without appeal to the mediation of interpersonal ties or exchanges. They note that in “taking this step, other properties of the collective unit (such as technology, rules, and laws) are introduced into the web of systemic relationships at the level of the collectivity” (p. 25).

Bidwell and Kasarda next define organizational environment to include all external phenomena that affect or could affect an organization. They identify four aspects of the environment, namely, supplies of resources, actors who supply them or who in other ways may affect their supply, flows of resources to and among various populations of organizations within the environment, and relationships among the environmental actors that influence the flow and utilization of resources. The first two are compositional aspects of the environment and the third and fourth are relational aspects (p. 38).

Bidwell and Kasarda divide the organizational environment into an “internal” and an “external” environment. The internal environment is synonymous with the structure of the organization itself. They posit that organizational structure is the locus of opportunities for and constraints on the organization’s further morphological evolution. The existence of the external environment, in its relational aspects (i.e., the surrounding social and normative order), is another principal source of these opportunities and constraints (p. 39).

In viewing community-based participatory planning entities, such as the CARE Act Title I planning councils and HIV prevention CPGs, as organizations, one finds that this particular population of organizations is continually expanding as additional metropolitan and rural areas turn to this type of organization to address local needs for health care, support services, and disease prevention/education. The specific organizational form that characterizes all community-based participatory planning entities is not a unique form. Although at first glance an argument might be made that these entities have the characteristics of “synthetic” organizations, as defined in Thompson,^[38] their apparently increasing permanence and institutionalization in the local health care environment suggest something more. All such entities appear to have many of the characteristics of what is known in the organizational behavior literature as a “minimalist” organization.

The minimalist organizational form is identified by Halliday et al.^[39,40] in a study of the vital events observed in the organizational population of state bar associations. Furthermore, Aldrich et al.^[41] used a similar approach in their study of U.S. trade associations. Both of these studies focus on minimalist organizations in the private sector and use the population ecology of organizations perspective. (For comprehensive discussions of population ecology of organizations, see Hannan and Freeman,^[42–44] Hannan and Carroll,^[45] Singh,^[46] and Tucker et al.)^[47]

The studies of minimalist organizations observe that such organizations are structurally flexible, frequently exist in relatively non-competitive environments, and have long life spans. These organizations may not demonstrate the patterns of foundings and failures characteristic of most business organizations, and especially not the liability of newness. They are called “minimalist” because they require minimal resources for founding and sustenance.

The differences between minimalist and non-minimalist organizations can be stated in terms of four core dimensions. Halliday et al.^[39] identified these dimensions as follows:

1. **Initial Costs.** Many organizations require extensive capital investments and labor commitments for birth, but minimalist organizations can be

founded with very limited labor commitments and capital. In the case of Title I planning councils, the establishment of a council is mandated by federal law. To qualify for funding, an eligible metropolitan area (EMA) must use its own resources to form a council, the costs of which are not reimbursable. Following a metropolitan area’s qualification for the program, application can be made for Title I formula and supplemental grant funding, and a small part of the total grant funds can be utilized for council support and development.

2. **Maintenance Costs.** Non-minimalist organizations require resource abundant environments for survival, but minimalist organizations can subsist in substantially poor resource environments. In the case of Title I planning councils, even when federal program funds are reduced or withheld, the work of the planning council continues.
3. **Reserve Infrastructures.** Most non-minimalist organizations have few shadow organizational structures or external resources to fall back on in times of hardship, but many minimalist organizations can supplement organizational resources with administrative and other infrastructures to be called on when necessary. In the case of Title I planning councils, numerous reserve infrastructures may exist, including assistance and staffing from the local public health jurisdiction, members who are only nominally or minimally reimbursed for their participation, agencies that maintain their employees on the payroll while they serve on the council, and various incidentals, such as travel time, out-of-pocket expenditures, and use of personal equipment and time for council business, all of which are usually not reimbursed by the Title I grant.
4. **Adaptiveness.** High sunk costs in capital and labor commitments impart structural inertia to many organizations, but low sunk costs enable minimalists to adapt readily to changes in their environments. Minimalists may also gain an advantage from normative flexibility: their conception of what the organization is about can more readily be altered. In other words, the minimalist organization can easily adapt or evolve to fit the particular environment in which it must function. In the case of Title I planning councils, there is an expectation that individual councils will develop in response to the constraints found in each of their local health care and services environments.

In addition to these four critical dimensions, Halliday et al.^[39] identified a subclass of minimalist

organizations that have a unique distribution of vital events because they evidence a high standing on two additional dimensions (p. 457). These dimensions include the following:

5. **Niche Definition.** Most organizations have overlapping or poorly defined niches that encourage competition, but many minimalist organizations have well-defined niches and segmented competitive environments that require minimal defense. This dimension pertains, in some instances, to government-sponsored corporations. In the case of Title I planning councils, a specific niche has been defined by the CARE Act, namely, the planning, prioritizing, and allocation of resources to meet the needs of persons with HIV disease for health care and support services in a specific geographical area. Originally, HIV prevention and education were specifically excluded from the purview of Title I councils. In the case of the HIV prevention CPGs, the Centers for Disease Control and Prevention defined the niche to include HIV prevention planning in a specific geographical area, usually a health jurisdiction or a combination of jurisdictions.
6. **Norms of Competition.** Competition is tolerated in many organizational populations, but a number of minimalist organizations tend to discourage competition. In most cases of this type of minimalist organization, a specific niche has been preserved for a particular minimalist organization (i.e., the minimalist organization has a “regional monopoly” in a clearly delineated territory and no other similar organization is permitted to function within the confines of that specific niche).

In the case of Title I planning councils, the boundaries of each respective EMA define the area in which a particular planning council has the responsibility and authority to plan, prioritize, and allocate resources for health care and support services for persons with HIV disease. Except for the required periodic submissions of supplemental grant applications, through which the metropolitan areas participating in the Title I program receive additional grant funds based on the merits of their application, there is supposedly no competition among the organizational population of Title I planning councils or the EMAs in which they operate. The same is said for each regional HIV prevention CPG. There is no competition among CPGs for funding.

Kieler et al.^[12] posited that the HIV health services planning council, as mandated by Title I of the CARE Act, have all four of the core dimensions manifested by minimalist organizations as well as the dimensions of

the subclass of minimalist organizations enjoying a monopoly in a particular environmental niche. They conclude that the Title I planning councils represent a new organizational population of mandated minimalist organizations in the public sector. Use of the population ecology of organizations perspective in the study of these planning councils could contribute to a greater understanding of this particular organizational population as well as other organizational populations of community-based participatory planning entities such as consortia, CPGs, and coalitions.

COORDINATION OF SERVICES

The responsibility for developing a regional comprehensive plan for the delivery of health care and support services for persons with HIV disease makes the HIV health services planning council mandated by Title I of the CARE Act a key element in the coordination of such services at the local level. Therefore it can be expected that the planning council could face many of the same problems and barriers that Aiken et al.^[48] found to be impeding efforts in the planning and implementation of integrated and coordinated mental health service delivery systems. These barriers include fragmentation of services, inaccessibility of services, lack of accountability of service delivery agencies, discontinuities in services, dispersal of services, wastefulness of resources, ineffectiveness of services, short-term commitments, and multiple local governments. Barriers to coordination are linked to aspects of the service delivery system including organizational autonomy, professional ideologies, conflicts among various client interest groups, and conflicts over who is to control the resources (p. 4). Many of these same barriers to coordination and issues of ideology that Aiken et al. observed in the 1970s were also observed by Kieler et al.^[8] during the Oakland planning council case study in 1992–1995, or 20 years after they were first identified in a similar study of mental health service delivery systems.

Aiken et al. noted that professional ideologies often prevent professionals in one field from wanting to cooperate with professionals in another field. They observe that competing client interest groups may work at crosspurposes, cancel out each other's efforts, and present a less than united front in the community. They also note that service organizations frequently put their own survival and prestige ahead of the needs of the clients. They point out various studies which show that acceptance of clients by service organizations depends on social, cultural, and historical factors, and not just on the needs of the clients; that agencies refer clients to places that profit the agency, rather than to places good for clients; and that agencies like

to have the “right” clients rather than those with the most pressing problems.

Aiken et al. suggested that one way to conceive of a coordinated delivery system was to view it as a change process having several stages. These stages of development are identical to those noted by Hernandez and Kaluzny^[49] and include awareness, initiation of effort, implementation, and routinization or institutionalization. In each stage, specific critical problems emerge:

1. Awareness Stage. The coordination effort usually has an initial period of increasing agitation by community groups and awareness by professional groups that treatment requires specialized services and new programs (p. 22).
2. Initiation of Effort Stage. Problems of gaining power, legitimacy, and funding usually occur in the initiation stage of the coordination effort (p. 22). Kieler^[11] elucidated this stage of development in the Oakland planning council study.
3. Implementation Stage. The problems that arise during implementation usually stem from the choice of the organizational structure for the service delivery system, internal conflicts among key participants, and lack of effective control over other organizations (i.e., resistance to implementation by some of the participants). Another problem that might arise is the transformation of the goals of the change agent. As failures in achieving objectives occur, goals of the program may become displaced (i.e., the change agent might scale down its objectives and begin to concentrate on particular goals that reflect its inherent interests and values) (p. 23). Again, Kieler^[11] found evidence of this stage in the Oakland planning council study.
4. Routinization Stage. The primary problem that might emerge lies with the resource controllers. Without their continued support, even the best of programs or service delivery systems would be jeopardized (p. 23). In the study of HIV planning councils, one recent example of this stands out, namely, the suspension of health care and support services for a person with HIV disease in Norfolk, Virginia, because of a contract dispute between the resource controllers (i.e., the city manager of Norfolk) and the service provider (i.e., Eastern Virginia Medical School Clinics).^[32]

Not only could the Title I planning council process in a particular metropolitan area face the various problems associated with the first three stages noted by Aiken et al.,^[48] it could also face the uncertainty of continued funding. Beginning in spring 1995, efforts

were initiated in the U.S. Congress to secure reauthorization of the Ryan White CARE Act for an additional 5-year period. As of January 1996, the fate of the CARE Act had not yet been decided, nor had any funds been appropriated for the act's various titles for the fiscal year to begin in fall 1996. Non-reauthorization of the CARE Act would have resulted in major changes in the organizational environment of all Title I planning councils, and possibly in their demise as a regional planning and coordinating process for health care and support services for persons with HIV disease. However, Fortunately, final reauthorization was approved by Congress in March 1996. Since that time, Congress has periodically reauthorized the CARE Act.

IMPLICATIONS OF THE OAKLAND STUDY FOR PRACTITIONERS

Analysis of the data pertaining to the Oakland metropolitan area's HIV health services planning council includes a systematic examination, discussion, and critique of several key aspects of the Ryan White CARE Act, the organizational form of the planning council, and the organizational environment in which a planning council must function. It also includes a description of the organization of the Title I process in the Oakland EMA and a discussion on the major issues that the planning council successfully dealt with during the period 1992–1994. It applies a process model of organizational development and change to the planning council's efforts to reform its organizational structure, bylaws, processes, and procedures, and it examines the planning council's efforts to assure adherence to the planning council's own bylaws and Robert's Rules of Order, and to assure the compliance by all planning council members with the conflict-of-interest requirements.

The Oakland study also presents the implications for practitioners of the issues that arose to challenge the planning council during the period of observation. The primary issues and challenges that the Oakland HIV health services planning council faced are probably typical of all such community-based participatory planning entities. These include multiple complex interorganizational relationships, conflicts of interest, preexisting societal tensions, factionalism and competition for influence by the major social groups in the area, changing trends in the epidemic, consumers' conflicting demands for services, competition among service providers for funding, membership burnout, accountability, compliance with bylaws and rules of order, dealing with financially troubled service providers who were understandably reluctant to relinquish grant funding, and competition with other

metropolitan areas for a fair share of the Title I funds. Kieler et al.^[8,12] and Kieler^[11] presented discussions on several of these major issues challenging community-based planning entities.

In mandating the establishment of the HIV health services planning council as the mechanism for assessing needs, setting priorities, and allocating Title I funds to health care and support service providers in an EMA, the CARE Act mandated the creation of a population of new minimalist organizations in the public sector. The CARE Act embedded this new organizational population in local HIV/AIDS-related health care and support services environments that are characterized by uncertainties and limitations of multiple complex interorganizational relationships, rivalries between services providers competing for limited funds, and conflicting demands of various populations that the planning council is mandated by Title I to serve. In other words, the CARE Act set the planning council into a multifaceted, diverse, and dynamic environment (i.e., a turbulent environment). Such an environment can have unpredictable impacts on the membership, functioning, credibility, and viability of the planning council process.

Any new organization, including community-based participatory planning entities, may face a liability of newness. A new planning entity could conceivably face an enormous burden in establishing its organizational legitimacy, in settling on an appropriate organizational structure to facilitate the accomplishment of its legislative mandates, and in adopting a standard operating procedure that would accommodate and satisfy its various stakeholders. The effort to achieve organizational legitimacy could consume an inordinate amount of time and effort during the first years of such an organization's existence.

The Oakland data suggest that a Title I planning council, even though it is a government-mandated organization, can have serious organizational legitimacy problems. These problems with legitimacy were reflected both in the various letters of complaint and in the opinions of providers and consumers about the planning council's prioritizations, allocations of funds, request-for-proposals process, appeals process, and contracting process. Additionally, the periodic efforts to change the planning council's organizational structure, prompted by internal opinion as well as by concerns expressed by federal program monitors, indicated a minimalist organization that was attempting to structure itself in response to environmental challenges to its legitimacy. The challenges to the legitimacy of the Oakland planning council stemmed primarily from the planning council's involvement in the direct allocation of Title I funds to local-level service providers, a role that more appropriately should have belonged to the grantee. However, removal of

the planning council in the mid-1990s from involvement in provider-specific allocations served to increase the perceived legitimacy of the planning council process in the Oakland metropolitan area.

If the organizational ecology concept of failure to survive could be defined to include abrupt and significant changes (i.e., massive changes) in a planning council's membership, organizational structure, bylaws, policies, processes, and/or interorganizational relationships (i.e., its core features), then strong evidence of failure to survive by certain members of this new organizational population could be expected and probably be at a much higher rate than is found in the event histories of other populations of minimalist organizations. However, it should be noted that an initial high rate of "organizational death" typifies many organizational populations, but that this initial high rate declines over time as the population of organizations is legitimated and as the population's members successfully compete with each other for limited resources.

Because it is almost inconceivable that the chief elected official of an EMA and the various stakeholders in the Title I planning council process would condone a non-functional or dysfunctional planning council for very long, massive changes in a planning council's core features might suggest that the planning council is in the process of rapidly adapting itself—it is being adapted—to fit the specific social and organizational environment in which it must function. That is to say, the planning council is evolving to fit the local environment, and its evolutionary track is characterized by punctuated patterns of morphological change. In other words, it is experiencing periods in which changes are unusually significant when compared to its prior state.

However, given the event history of the Title I planning council in the Oakland metropolitan area through 1995, the case for evolution seems weak. The Oakland data from this period suggest an alternative hypothesis concerning massive changes in a particular planning council's core features, namely, that a replacement of an existing planning council by a successor planning council occurred. Furthermore, it appears that replacement of a planning council by a successor planning council is highly feasible, basically without cost, and might even serve to enhance the acceptability, credibility, and viability of the Title I process in an EMA. Swift replacement of an existing planning council, a dysfunctional planning council, or a collapsed planning council does not appear to jeopardize either the area's Title I status and eligibility, or its Title I-funded health and social services delivery system. However, the ramifications of the replacement procedure on perceptions of empowerment of the HIV/AIDS community are yet to be articulated.

CONCLUSION

An examination of the event histories of other Title I planning councils, or other similar community-based participatory planning entities, might provide additional evidence of the occurrence of replacement of an established planning entity by a successor planning entity. Quite possibly, it might be found that, in certain metropolitan areas and jurisdictions, there have been repeated replacements of the local participatory planning entity. Additionally, elaboration and analysis of the phenomenon of replacement of a planning council by a successor council may provide additional support for the hypothesis that environmental selection is occurring in this new population of public sector minimalist organizations.

ACKNOWLEDGMENTS

Preparation of this manuscript was begun during the Oakland HIV health services planning council study and was supported, in part, by a grant from the Henry J. Kaiser Family Foundation (Menlo Park, CA). Additional support came from the U.S. Public Health Service and the University of California's AIDS Research Program. Technical support was provided by the Institute of Industrial Relations at the University of California at Berkeley.

REFERENCES

- Kieler, B.W.; Saporta, I.; El'Amin, Z. Community-based planning. In *Handbook of Health Administration and Policy*, 1st Ed.; Kilpatrick, A.O., Johnson, J.A., Eds.; Marcel Dekker, Inc.: New York, 1999; 195–206.
- Centers for disease control and prevention. *Supplemental Guidance on HIV Prevention Community Planning for Noncompeting Continuation of Cooperative Agreements for HIV Prevention Projects*; Centers for Disease Control and Prevention: Atlanta, GA, 1994.
- Brown, C. *The Art of Coalition Building: A Guide for Community Leaders*; The American Jewish Committee: New York, 1984.
- Feighery, E.; Rogers, T. *Building and Maintaining Effective Coalitions*; Stanford Health Promotion Resource Center: Palo Alto, CA, 1989.
- Minkler, M., Ed.; *Community Organizing and Community Building for Health*; Rutgers University Press: New Brunswick, NJ, 1997.
- Academy for educational development. *Handbook for HIV Prevention Community Planning*; Centers for Disease Control and Prevention: Atlanta, GA, 1994.
- McKay, E.G. *Do's and Don'ts for an Inclusive HIV Prevention Community Planning Process: A Self-Help Guide*; National Council of La Raza/Center for Health Promotion: Washington, DC, 1994.
- Kieler, B.W.; Rundall, T.G.; Saporta, I.; Sussman, P.C.; Keilch, R.; Warren, N.; Black, S.; Brinkley, B.; Barney, L. Challenges faced by the HIV health services planning council in Oakland, California, 1991–1994. *Am. J. Prev. Med.* **1996**, *12* (4), 26–32.
- Bowen, G.S.; Marconi, K.; Kohn, S.; Bailey, D.M.; Goosby, E.P.; Shorter, S.; Niemcryk, S. First year of AIDS services delivery under title I of the Ryan White CARE Act. *Public Health Rep.* **1992**, *107* (5), 491–499.
- Kachur, S.P.; Sonnega, A.J.; Cintron, R.; Farup, C.; Silbersiepe, K.; Celentano, D.D.; Kwait, J. An analysis of the greater Baltimore HIV services planning council. *AIDS Public Policy J.* **1992**, *7* (4), 238–246.
- Kieler, B.W. *The Oakland CARE Act Title I HIV/AIDS Planning Council: A Minimalist Organization Functioning in a Turbulent Environment*; University Microfilms, Inc.: Ann Arbor, MI, 1994.
- Kieler, B.W.; Rundall, T.G.; Saporta, I. The Oakland HIV/AIDS planning council: its organizational form and environment. *Int. J. Public Adm.* **1996**, *19* (7), 1203–1219.
- Marconi, K.; Rundall, T.; Gentry, D.; Kwait, J.; Celentano, D.; Stolley, P. The organization and availability of HIV-related services in Baltimore, Maryland, and Oakland, California. *AIDS Public Policy J.* **1994**, *9* (4), 173–181.
- Gentry, D.; Rundall, T.G. Staffing in AIDS service organizations: the volunteer contribution. *J. Health Hum. Serv. Adm.* **1995**, *18* (2), 190–204.
- Rundall, T.G.; Kwait, J.; Marconi, K.; Bender-Kitz, S.; Celentano, D. Impact of the Ryan White CARE Act on the availability of HIV/AIDS services. *Policy Stud. J.* **1999**, *27* (4), 826–839.
- Kwait, J.; Marconi, K.; Helitzer, D.; Rodieck, M.; Rundall, T.G.; Celentano, D. Ryan White CARE Act Title I funding priorities and unmet needs in Baltimore, Maryland. *Policy Stud. J.* **1999**, *27* (4), 855–871.
- Fleishman, J.A.; Mor, V.; Piette, J.D.; Allen, S.M. Organizing AIDS service consortia: lead agency identity and consortium cohesion. *Soc. Sci. Rev.* **1992**, *66* (4), 547–570.
- Penner, S.J. *A Study of Coalitions Among Voluntary AIDS/HIV Service Organizations in California*; University Microfilms, Inc.: Ann Arbor, MI, 1992.
- Penner, S.J. Problems with planning for the HIV epidemic. *AIDS Public Policy J.* **1992**, *7* (2), 120–127.
- McKinney, M.M. Consortium approaches to the delivery of HIV services under the Ryan White CARE Act. *AIDS Public Policy J.* **1993**, *8* (3), 115–125.
- McKinney, M.M.; Wieland, M.K.; Bowen, G.S.; Goosby, E.P.; Marconi, K.M. States' responses to Title II of the Ryan White CARE Act. *Public Health Rep.* **1993**, *108* (1), 4–11.
- Mor, V.; Fleishman, J.A.; Piette, J.D.; Allen, S.M. Developing AIDS community service consortia. *Health Aff.* **1993**, *12* (1), 186–199.
- Mor, V.; Fleishman, J.A.; Allen, S.M.; Piette, J.D. Consortium structure and operation. In *Networking AIDS Services*; Mor, V., Fleishman, J.A., Allen, S.M., Piette, J.D., Eds.; Health Administration Press: Ann Arbor, MI, 1994; 55–74.

24. Doughty, R. *Lessons From the First Two Years: Issues Arising in the Implementation of Ryan White CARE Act Titles I and II*; AIDS Action Foundation: Washington, DC, 1993. (Unpublished report).
25. Ryan White Study Group; Rundall, T.G.; Principal Investigator. *Implementation of Title I of the Ryan White CARE Act of 1990: A Report to the Kaiser Family Foundation*; The Kaiser Family Foundation: Menlo Park, CA, 1994. (Unpublished report).
26. HHS awards 1 billion in Ryan White grants. *AIDS Policy Law* **2003**, 18 (8), 1.
27. CAEAR coalition. *Ryan White CARE Act Title I and Title III FY2003 Funding Needs*; Communities Advocating Emergency AIDS Relief Coalition: Washington, DC, (no date).
28. CAEAR coalition. *Ryan White CARE Act Title I and Title III FY2004 Funding Needs*; Communities Advocating Emergency AIDS Relief Coalition: Washington, DC, (no date).
29. Oversight of spending for Ryan White CARE Act to receive review at HHS. *AIDS Policy Law* **2001**, 16 (14), 1–2.
30. McKinney, M.M.; Marconi, K.M. Delivering HIV services to vulnerable populations: a review of CARE Act-funded research. *Public Health Rep.* **2002**, 117 (2), 99–113.
31. Buchanan, R.J. Ryan White CARE Act and eligible metropolitan areas. *Health Care Finance Rev.* **2002**, 23 (4), 149–157.
32. Lesson from AIDS flap: don't mess with success. In *The Virginian-Pilot*; Pilot Media Companies: Norfolk, VA, July 9, 2003.
33. Los angeles county board of supervisors releases results of investigation into county AIDS office. In *Kaiser Daily HIV/AIDS Report*; The Henry J. Kaiser Family Foundation: Menlo Park, CA, December 12, 2002.
34. New York mayor Bloomberg's decision to transfer AIDS groups to health department sparks 'bitter power struggle' over AIDS funds. In *Kaiser Daily HIV/AIDS Report*; The Henry J. Kaiser Family Foundation: Menlo Park, CA, April 29, 2003.
35. Slack, J. Zero-sum politics, the Herbert thesis, and the Ryan White CARE Act: lessons learned from the local side of AIDS. *J. Health Human Serv. Adm.* **2001**, 24 (1), 80–102.
36. Scott, W.R. The organization of medical care services: toward an integrated theoretical model. *Med. Care Rev.* **1993**, 50 (3), 271–303.
37. Bidwell, C.E.; Kasarda, J.D. *The Organization and Its Ecosystem: A Theory of Structuring in Organizations*; JAI Press, Inc.: Greenwich, CT, 1985.
38. Thompson, J.D. The synthetic organization. In *Organizations in Action: Social Science Bases of Administrative Theory*; Thompson, J.D., Ed.; McGraw-Hill Book Company: New York, 1967; 52–54.
39. Halliday, T.C.; Powell, M.J.; Granfors, M.W. Minimalist organizations: vital events in state bar associations, 1870–1930. *Am. Sociol. Rev.* **1987**, 52 (4), 456–471.
40. Halliday, T.C.; Powell, M.J.; Granfors, M.W. After minimalism: transformations of state bar associations from market dependence to state reliance, 1918 to 1950. *Am. Sociol. Rev.* **1993**, 58 (4), 515–535.
41. Aldrich, H.; Staber, U.; Zimmer, C.; Beggs, J.J. Minimalism and organizational mortality: patterns of disbanding among U.S. trade associations, 1900–1983. In *Organizational Evolution: New Directions*; Singh, J.V., Ed.; Sage Publications, Inc.: Newbury Park, CA, 1990; 21–51.
42. Hannan, M.T.; Freeman, J. The population ecology of organizations. *Am. J. Sociol.* **1977**, 83, 929–964.
43. Hannan, M.T.; Freeman, J. Structural inertia and organizational change. *Am. Sociol. Rev.* **1984**, 49 (2), 149–164.
44. Hannan, M.T.; Freeman, J. *Organizational Ecology*; Harvard University Press: Cambridge, MA, 1989.
45. Hannan, M.T.; Carroll, G.R. *Dynamics of Organizational Populations: Density, Legitimation, and Competition*; Oxford University Press: New York, 1992.
46. Singh, J.V., Ed.; *Organizational Evolution: New Directions*; Sage Publications, Inc.: Newbury Park, CA, 1990.
47. Tucker, D.J.; Singh, J.V.; Meinhard, A.G. Organizational form, population dynamics, and institutional change: the founding patterns of voluntary organizations. *Acad. Manage. J.* **1990**, 33 (1), 151–178.
48. Aiken, M.; Dewar, R.; DiTomaso, N.; Hage, J.; Zeitz, G. *Coordinating Human Services*; Jossey-Bass, Inc., Publishers: San Francisco, CA, 1975.
49. Hernandez, S.R.; Kaluzny, A.D. Organizational innovation and change. In *Health Care Management: Organization Design and Behavior*; Shortell, S.M., Kaluzny, A.D., Eds.; Delmar Publishers, Inc.: Albany, NY, 1994; 294–315.

Compensation: Pay and Benefits

James A. Buford Jr.

Ellis-Harper Management, Auburn, Alabama, U.S.A.

INTRODUCTION

Compensation is one of the most challenging areas of human resource management. The compensation system, consisting of pay and benefits, should attract and retain a competent workforce, build motivation and performance, and control costs.^[1] Most organizations in the public sector have job-based classification systems, grant annual pay increases or “step raises,” and offer conventional health insurance and defined benefit retirement programs. In recent years, however, emerging approaches in delivering pay and benefits have begun to gain acceptance. These include skill and competency-based pay, structures with wide grades or bands, merit pay, group incentive pay, managed care, medical savings accounts, and defined contribution retirement programs. Traditional and emerging approaches can be useful, depending on what the organization wishes to accomplish. In any case, clearly articulated strategic objectives are necessary if the compensation system is to match the unique characteristics and culture and effectively contribute to implementing the mission of the organization.^[2]

against classification descriptions arrayed in increasing order of difficulty and responsibility. In the point method, jobs are evaluated based on the degree to which they require factors or subfactors that measure the universal factors mentioned above. The point method is the technique of choice in the private sector and is used increasingly by public organizations such as municipalities and counties.

In evaluating jobs, organizations may use a plan containing a set of class descriptions or factors designed to cover all jobs in a single pay structure or separate plans for each job family, such as clerical, labor and trades, administrative and professional, and so on, and develop multiple pay structures. While placing all jobs in one structure implements the principle of equal pay for work of equal value (or comparable worth), many compensation professionals contend that the work content is too varied to be compared on the same factors. For example, factors that distinguish among levels of equipment operators include manual dexterity, physical effort, and working conditions. These factors rarely apply in clerical and administrative support work; rather such factors as knowledge of office technology, public contacts, and access to confidential data are more important.

JOB-BASED PAY STRUCTURES

Most public organizations use pay systems that focus on work performed in specific jobs. Once pay policies have been determined, the development of a pay system includes determining the internal value of jobs and establishing grades or classes containing jobs having approximately the same value.^[3]

Evaluating Jobs

Organizations first develop an internal hierarchy of job value through a process known as job evaluation. All job evaluation techniques are based, implicitly or explicitly, on the “universal” factors of skill, effort, responsibility, and working conditions. These factors are also set forth in the Equal Pay Act of 1963 (discussed in a following section) as the legal standard for equal work. The most common job evaluation techniques are classification and the point method. Classification, which originated in the public sector, is a whole-job approach that involves matching jobs

Determining Grades

Jobs of approximately equal internal value are then grouped into grades (or classes) that are assigned pay ranges (discussed in a following section). The actual number of grades in a pay structure is a policy question. Specialization and upward mobility are facilitated by a relatively large number of narrow grades. Wider grades promote teamwork and lateral mobility. Some public sector pay structures contain up to 50 grades, while others contain as few as 10. The trend is to have fewer grades in all types of structures. The General Schedule of the U.S. Office of Personnel Management classifies thousands of jobs into 15 grades, which might be considered a reasonable upper limit for most public organizations.

ALTERNATIVES TO JOB-BASED STRUCTURES

While job-based structures are still the major approach to pay in public organizations, they have come under

increasing criticism in recent years. Rather than build participation and contribution to objectives, these structures are seen to be rigid and bureaucratic and to promote individual agendas and entitlement. Alternatives to traditional structures include broadbanding and skill-based pay.

Broadbanding

In broadbanding, a number of grades are consolidated into a single band with a maximum and minimum rate of pay. Most organizations use four to eight bands set at logical breaks in skill or competency levels, such as clerical, technical, administrative/professional, managerial, and executive. Administration of pay to individual employees (discussed in a following section) is typically delegated to operating departments rather than centralized in the human resources department. The contrast between pay grades and bands are illustrated in Fig. 1.

Broadbanding provides greater flexibility to line managers in responding to market pressures and organizing work; however, the culture and work flow of an organization is a major factor in determining if this technique is appropriate. It is most effective with delayed organizations that de-emphasize job boundaries and focus on rewarding entrepreneurship and individual work contributions. In organizations that are hierarchical and job based the technique tends to disconnect pay rates from job families and can create pay inequities and cause legal exposure. Finally, studies have shown that organizations using broadbands have higher labor costs.^[4]

Skill- and Competency-Based Pay Structures

Skill-based pay structures essentially do away with job boundaries. In these systems, the employee is paid for skills mastered and applied in a work situation,

regardless of the tasks assigned at a particular time. Structures typically contain a block of foundation skills, a number of blocks of required or “core” skills, and additional blocks of optional skills. For example, an equipment operator might receive a pay increase for learning a required number of core skills, such as finish grading, even though most of the work consists of less difficult earthmoving tasks such as trenching, rough grading, and loading. A skill-based structure for equipment operators is shown in Fig. 2.

While skill-based structures are used mainly in hourly trade and craft work, competency-based pay extends the concept to administrative, technical, and professional work. However, the term competency covers a wider range of employee attributes, such as a competency to “make effective and timely decisions.” Accordingly, competencies are considerably more difficult to operationally define and build into a pay structure.^[5]

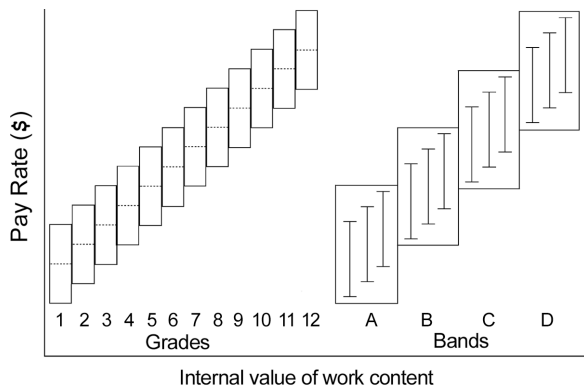


Fig. 1 Pay grades vs. bands.

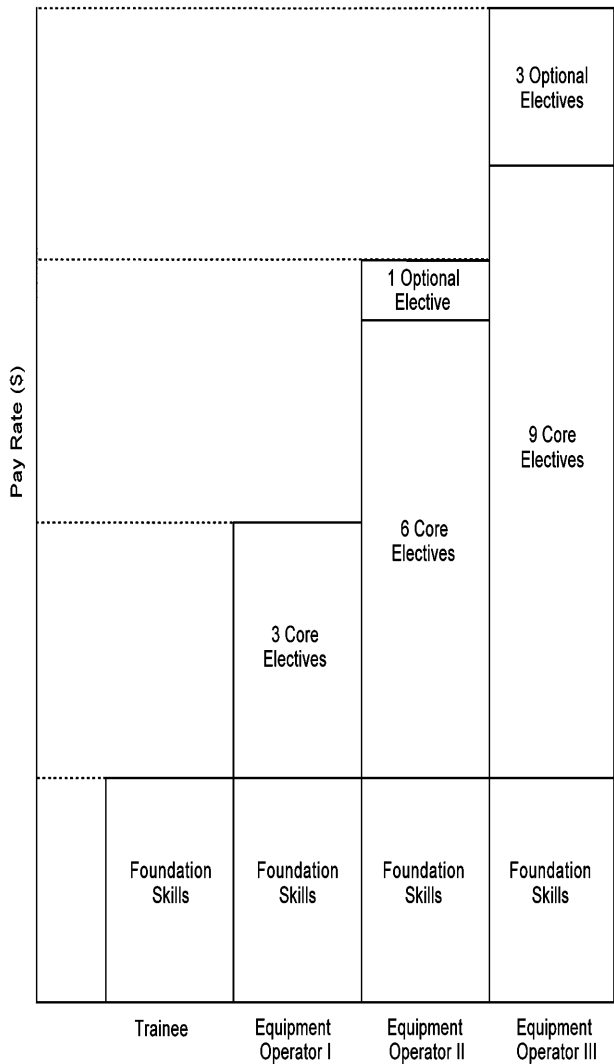


Fig. 2 Skill-based pay structure.

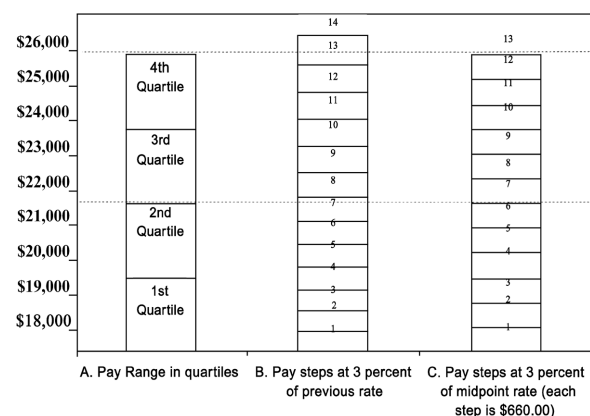
The next activity is to price the structure by establishing a pay policy line as a basis for pay ranges with minimum and maximum rates of pay. In establishing ranges, the organization must consider what other employers in the relevant labor market pay for similar jobs. This information may be obtained by conducting a compensation survey or by using secondary data. In pricing the structure, organizations may establish grade midpoints that lag, match, or lead the market, depending on funding levels, strategic objectives, and policy guidance. In skill-based structures, the pay must be further allocated to various skill blocks. Because most public organizations have attractive benefits programs (discussed in a following section), the policy is typically to lag or match the market. Moreover, organizations such as units of government would encounter considerable political resistance in setting pay higher than market levels.

Employee compensation is heavily influenced by laws and regulations that emphasize fairness, sufficiency, and protection. The Fair Labor Standards Act (FLSA) establishes the minimum wage, defines maximum hours for various work periods (typically the 40-hour week), provides for overtime pay, and restricts child labor. The FLSA is enforced by the Wage and Hour Division of the U.S. Department of Labor. The Equal Pay Act (EPA) requires employers to provide equal pay for equal work, defined as that which requires equal skill, equal effort, equal responsibility, and is performed under similar working conditions. If jobs are equal, males and females may be compensated differently only if pay is based on seniority, merit, quality or quantity of production (discussed in the following section), or other factors based on sex. Title VII of the Civil Rights Act prohibits disparate treatment in pay based on race, sex, color, religion, and national origin, even if jobs are not equal, and addresses discrimination in the provision of benefits. The Age Discrimination in Employment Act (ADEA) extends protection to persons 40 years and older. The EPA, Title VII, and the ADEA are enforced by the Equal Employment Opportunity Commission. While these are the major laws, other federal laws and regulations apply in specific situations, and many states also have laws applying to compensation.

In job-based structures, pay for employees is administered within the maximum and minimum for the

In Panel A, the grade is divided into quartiles. In Panel B, the grade is divided into 13 steps, with each step representing a 3% increase from the previous step. In Panel C (12 steps), each step represents 3% of the midpoint rate. Most public organizations follow a step approach in administering pay within grades. Pay adjustments are designed to set pay at a level appropriate to the employee's length of service, job performance, and qualifications. The grade midpoint represents the target rate of pay for a fully trained employee performing at a satisfactory level. Recall that midpoints were set previously at or near the market rate of pay. When bands are used, organizations often establish guidelines or "shadow grades" within bands as shown in [Fig. 1](#). Where this is the case, within-band pay adjustments are handled similarly to within-grade adjustments.

Seniority-based pay systems are common in the public sector. In these systems, employees are placed in and advance through rate ranges based on length of service (the familiar annual step increase). In recent years, however, various forms of performance-based pay for both individuals and groups (discussed in following section) have gained increased acceptance.^[6] Merit pay is an addition to an employee's base wage or salary in recognition of *past* contributions. A major disadvantage of merit pay is that increases are "locked in" and have a compounding effect even if performance declines. This problem is overcome by one-time incentives or bonuses, offered as an inducement for *future* performance. In this approach, pay reverts to the base



© 2008 by Taylor & Francis Group, LLC

rate during any period that performance does not exceed the standard, and the incentive amount is therefore “at risk.” Individual incentive plans such as piecework and performance bonuses are found mainly in the private sector, although a few public organizations are beginning to implement this approach.

In skill- and competency-based structures, pay is increased when an employee has mastered a skill or competency. Often, work sample tests are used for this purpose. For example, to demonstrate the skill of finish grading, an equipment operator might be required to operate a motor grader, manipulating controls, adjusting the height and angle of the blade, and positioning the wheels to raise or lower terrain to grade marked on reference stakes.

Rewarding Group Performance

Various types of incentive plans are used to reward group or team performance. In the approach known as “gainsharing,” a productivity standard based on allowable labor costs is established. For example, the labor cost standard in a solid waste collection department may be \$100,000 for a given month. If actual costs are \$95,000 during the month, there is a “gain” of \$5000. In distributing the gain, 50% is typically allocated to the work group, 25% is retained by the organization, and 25% is set aside for periods in which labor costs exceed the standard during a period. Note that in this approach, pay is at risk, as was the case with individual incentives. The procedures discussed above follow the Scanlon Plan, which is the gainsharing approach that has gained the most acceptance in the public sector.

DETERMINING EMPLOYEE BENEFITS AND SERVICES

The creation of a benefits package is the final step in designing the compensation program. As was the case with pay, strategic objectives and policies are important. Because public organizations do not typically lead the market in pay, a reasonably attractive benefits package is necessary to attract and retain employees. Certain benefits are required by law. These include workers’ compensation, unemployment compensation, and Social Security (some public jurisdictions elected not to participate and are required to offer a comparable retirement benefit). Optional benefits include retirement and savings plans, health and medical plans, payment for time not worked (paid leave and holidays), and miscellaneous benefits. Controlling the escalating costs of retirement and health benefits is a major

challenge and public organizations are beginning to adopt strategies and approaches similar to those used in the private sector.

Retirement and Savings Plans

Most public organizations, especially units of government, offer defined benefit retirement plans or pensions that provide lifetime annuities of 50–70% of employees’ pre-retirement earnings. It is well documented that these plans cost more to operate than defined contribution plans, similar to 401 (K) plans, which predominate in the private sector. Although few public organizations offer these plans, the number is expected to grow as plan administrators attempt to control costs.^[7]

Medical and Health Benefits

In the area of medical and health benefits, the traditional approach is to provide insurance coverage where the employees can select a health-care provider (physician, hospital, clinic, etc.). As is the case with defined benefit retirement programs the growing cost of these plans is a major concern and organizations are considering less expensive alternatives. A small but growing number of public organizations have begun to implement managed care plans through health maintenance organizations or HMOs. These organizations typically negotiate a flat rate and have an incentive to promote wellness rather than respond to costly medical emergencies. Under managed care, patients must use health-care providers designated by the plan. Moreover, the provider has the responsibility to provide care based on cost effectiveness, not necessarily on the needs of the patient. These and other issues have led to legislation at the state and federal levels establishing “bills of rights” for employees covered by managed care plans. Other approaches to controlling include preferred provider organizations and medical savings plans, which may be used in combination with traditional coverage and managed care.^[8]

Other Benefits and Services

In regard to payment for time not worked, public organizations tend to give more paid time off than their counterparts in the private sector. Practically, all organizations observe national holidays and others with a regional or statewide significance. Most also provide various forms of paid leave, including vacation, sick, military, jury, and so forth.

Other benefits and services often provided by public organizations include group life insurance, employee assistance programs, tuition refund programs, and credit unions.

SUMMARY AND CONCLUSION

Compensation practices vary widely in organizations. In pay systems work performed is exchanged for monetary rewards. Such work is typically organized into jobs, although it may be structured in other formats such as skill- or competency- based systems or replacing grades with broad bands that ignore or deemphasize job boundaries. In either case the work must be evaluated based on its content, the knowledge and skills needed to perform the work must be priced in the external market and employees must be rewarded for their individual contributions. Basing pay adjustments on seniority and past performance (merit) is common practice, but there is increasing interest in one-time individual and group incentive systems. These systems place pay "at risk" and are an inducement to future performance.

In non-monetary compensation, benefits are provided for membership in the organization. Certain benefits such as Social Security mandatory and must be administered in accordance with applicable laws and regulations. Other benefits are discretionary; however organizations usually offer retirement and medical benefits. Most public organizations offer defined benefit (pension) retirement plans and traditional health insurance plans, but escalating costs have resulted in a growing trend toward defined contribution plans similar to 401(k) plans in the private sector and managed care approaches including health maintenance

organizations (HMOs) and related programs emphasizing pre-paid medical care and wellness.

All aspects of compensation must be administered in accordance with legal requirements such as the Fair Labor Standards Act, The Equal Pay Act and Title VII of the Civil Rights Act as well as additional federal and state laws and regulations. The challenge for all organizations is to use compensation dollars wisely and to develop a system that is legally defensible, that attracts and retains well-qualified employees and that motivates these employees to work efficiently and effectively in achieving strategic objectives.

REFERENCES

1. Buford, J.A., Jr.; Lindner, J.R. *Human Resource Management in Local Government*; South-Western: Cincinnati, 2002.
2. Scott, D.; Sperling, R.S.; McMullen, T.D.; Wallace, M. Linking compensation policies and programs to organizational effectiveness. *WaW J.* **2003**, *12* (4), 35–44.
3. Milkovich, G.T.; Newman, J.M. *Compensation*, 9th Ed.; Irwin McGraw-Hill: Boston, 2008.
4. Fay, C.; Schultz, E.; Gross, S.E.; Van De Voort, D. Broadbanding, pay ranges and labor costs. *WaW J.* **2004**, *13* (2), 8–22.
5. Zingheim, P.; Schuster, J.R. Reassessing the value of skill-based pay. *WaW J.* **2002**, *11* (3), 72–77.
6. Van Neck, K.H.; Smilko, J.E. Variable pay plans: creating a financial partnership with the work force. *WaW J.* **2002**, *11* (4), 74–79.
7. Salisbury, D.; McDonnell, K.J. Benefit cost comparisons between state and local governments and private sector employees. *PPM* **2005**, *34* (4), 321–327.
8. Chiapetta, T.O. Managing healthcare costs. *PPM* **2005**, *34* (4), 313–320.

Comprehensive Annual Financial Report

Robert S. Kravchuk

Indiana University, Bloomington, Indiana, U.S.A.

INTRODUCTION

A comprehensive annual financial report (hereafter, “CAFR”) is a general-purpose financial report that is comprehensive in the depth of its reporting detail, and which provides a full disclosure of the financial affairs of the reporting entity, beyond the requirements of generally accepted accounting principles (hereafter, “GAAP”), and other applicable legal and contractual requirements. Pursuant to the 2003 *Codification of Governmental Accounting and Financial Reporting Standards* (hereafter, “Codification”), every government “should prepare and publish a CAFR covering all funds and activities of the primary government (including its blended component units) and providing an overview of all discretely presented component units of the reporting entity.”^[1]

A CAFR is recommended but not required under GAAP. However, the minimum requirements for fair presentation in connection with general-purpose external financial reporting are mandatory, and are specified in detail in the *Codification*. The CAFR is to be a general-purpose report, which means that its contents are intended to meet the needs of a broad range of users, including citizens, investors, credit rating agencies, other governments, oversight authorities, and other interested parties. The CAFR also provides supporting schedules and other information necessary to demonstrate compliance with finance-related legal requirements and contractual provisions, and statistical data pertaining to the reporting entity.^[2]

REPORTING ENTITY

The financial reporting entity covered by the CAFR includes: 1) the primary government; 2) organizations for which the primary government is financially accountable; and 3) other organizations of such nature or significance that their exclusion would cause the basic financial statements of the primary government to be incomplete or misleading.^[3] The nucleus of the reporting entity is the primary government, having as its foundation a separately elected governing body, which is legally separate from other governing bodies, and which is also fiscally independent from them.^[4]

A primary government is any state or local government that meets these criteria.

The primary government consists of all funds (defined as budgetary and accounting entities having a self-balancing group of accounts), departments, organizations, agencies, institutions, and other offices that are not legally separable from any part of the primary government. Certain other units may be required to be included with the primary government for financial reporting purposes. Such component units of the primary government are legally separable organizations for which the elected officials of the primary government are: 1) financially accountable; 2) able to control through appointment of the component unit’s officials; 3) otherwise able to impose their will on that organization; 4) able to access or employ that organization’s resources to the benefit (or burden) of the primary government; or 5) obligated in some manner for that organization’s debt.^[5]

Component units may therefore include other governmental organizations, not-for-profit corporations, and for-profit corporations. In any case, “component units are to include all other organizations for which the nature and significance of their relationship with a primary government are such that exclusion would cause the reporting entity’s financial statements to be misleading or incomplete.”^[6] Due to the closeness of the financial relations between the primary government and its component units, the latter should be blended as though they are a part of the primary government.^[7] Blending involves reporting transactions and account balances in a manner similar to those of the primary government itself. However, most component units should be discretely presented as well. Discrete presentation entails reporting the component unit’s financial detail on the financial statements in columns and rows separate from those of the primary government.

INTRODUCTORY SECTION

The CAFR should contain: 1) an introductory section; 2) a financial section; and 3) a statistical section. The purpose of the introductory section is to familiarize the user of the CAFR with the structure of the government, the nature and scope of its activities and the services it provides, and certain relevant aspects of its

legal environment. The introductory section includes a letter of transmittal from the government's chief financial officer conveying the report to the legislative body. The letter of transmittal includes a brief profile of the government and any information that CAFR users may find useful in assessing the entity's financial condition, such as developments in the local economy, long-term financial plans, cash management and investments, risk management, and pension obligations (together with any other post-employment benefits). If the government has been awarded a Certificate of Achievement for Excellence in Financial Reporting by the Government Finance Officers' Association, the current certificate is included after the letter of transmittal. This section also includes the names of the entity's principal officials, an organization chart, and any other material deemed appropriate by management.^[8]

FINANCIAL SECTION

The financial section is to present a top-down, comprehensive overview of the entity's financial operations. It contains the government's basic financial statements, along with all required supplementary information (hereafter, "RSI"), and the report of an independent auditor. Also included should be management's discussion and analysis (hereafter, "MD&A"), all RSI other than MD&A, and combining and individual fund statements and supplementary information. The financial section also contains any schedules necessary to demonstrate compliance with legal and contractual commitments, and other schedules in support of individual financial statements (e.g., lists of long-term debt obligations, investments, cash receipts, etc.).

Independent Auditor's Report

The report of the independent auditor appears first in the financial section. It is widely regarded as the most important document contained within the financial section. The report offers a professional opinion on the fair presentation of the financial position and results of operations of the entity for the period covered. Consequently, it is to be read carefully.

Management's Discussion and Analysis

The independent auditor's report must be immediately followed by the MD&A. The MD&A is always RSI. The MD&A is a relatively new requirement, which was introduced in June 1999. An MD&A has been a GAAP requirement of private firms for decades. The purpose

of its inclusion in governmental financial reports is to permit management to communicate its interpretation of the entity's financial performance during the period covered by the financial statements, compared to prior periods.^[9]

MD&A should be written in an easily understandable manner, and may include such tables, charts, and graphs as may enhance its readability. The focus of MD&A is on the primary government; however, component units may also be discussed. GAAP specifies certain minimum requirements for MD&A.^[10] However, it must be stressed that the standards are designed to permit management flexibility to present relevant data at whatever level of detail that it may deem appropriate. Insofar as MD&A is RSI, a government may not go beyond the topics required by GAAP for inclusion in MD&A. This means that MD&A is not to repeat information presented in the letter of transmittal, and is to avoid a "boilerplate" approach to the presentation of relevant financial information.

Requirements for MD&A include: 1) a brief discussion of the financial statements; 2) condensed financial information from the government-wide financial statements (i.e., statement of net assets and statement of activities) in comparison to prior years; 3) analysis of the government's overall financial position that aids users in determining whether the financial position has improved or deteriorated; 4) analysis of significant changes in fund balances; 5) analysis of significant variations between the original budget, amended final budget, and actual results; 6) a description of significant long-term liabilities and debt activity; 7) a discussion of infrastructure assets, if the modified approach is used to report some or all of a government's infrastructure;^[11] and 8) a statement of currently known facts, decisions, or conditions that are expected to have a significant effect on the financial position or results of operations.

Basic Financial Statements

MD&A is followed by the government's basic financial statements. Together with the accompanying notes to the financial statements, the basic financial statements are at the core of the CAFR financial section. Depending on the extent of a particular government's operations, some nine basic financial statements may be required:

- *Government-wide financial statements:*
 1. Statement of net assets (i.e., balance sheet)
 2. Statement of activities (i.e., operating statement)
- *Fund financial statements:*
 3. Balance sheet—governmental funds

4. Statement of revenues, expenditures, and changes in fund balance—governmental funds
5. Statement of net assets—proprietary funds
6. Statement of revenues, expenses, and changes in net assets—proprietary funds
7. Statement of cash flows—proprietary funds
8. Statement of fiduciary net assets
9. Statement of changes in fiduciary net assets

In addition, the financial section contains separate subsections for each fund type (i.e., governmental, proprietary, and fiduciary), including combining statements by fund type, along with physically separate individual funds statements, including a balance sheet, operating statement, budget-to-actual comparison schedules, and prior-year comparative data. Normally, fund data for discretely presented component units are not provided, in that interested parties can find such detailed financial information in the separately reported financial statements of such units.

Required Supplementary Information, other than MD&A

The basic financial statements are followed by RSI other than MD&A. The specific presentation depends on a government's specific circumstances. It is possible that four categories of RSI other than MD&A should have to be reported: budgetary comparison schedules showing budget vs. actual comparisons for the year covered by the financial report for all funds that have annually appropriated budgets;^[12] infrastructure condition and maintenance data (for entities that use the modified approach for infrastructure reporting); pension trend data for certain pension plans (including a schedule of funding progress and a schedule of employer contributions); and revenue and claims development trend data for governments participating in public-entity risk pools.^[12] Under a recently adopted GAAP requirement, the budgetary comparisons are to be based on the fund, organization, or program structure that the entity uses for its legally adopted budgets.^[13]

STATISTICAL SECTION

The purpose of the statistical section is to provide CAFR users with a broad range of information permitting a more complete understanding of government operations and financial trends than would be possible from the financial section alone. The statistical section therefore provides trend data and non-financial

information that may be useful in assessing the reporting entity's financial condition. These data may be of particular use to investors, creditors, and bond rating agencies in identifying non-financial trends that may affect the entity's creditworthiness.

In May 2004, the Governmental Accounting Standards Board (hereafter, "GASB") issued new GAAP requirements in order to improve the understandability and usefulness of supplementary information presented in the CAFR statistical section.^[14] While much of the contents of the statistical section required since 1980 remain, the new standards provide for improved consistency and comparability in reporting among governments. The new guidelines also call for appropriate narrative explanations for certain information presented in the statistical tables.

Statistical Section Contents

The revised statistical section is to provide information in five categories:

1. Ten years' worth of financial trend information, in order to be able to assess how an entity's financial position has changed over time (including net assets and fund balances, and changes therein).
2. Revenue capacity information, which pertains to a government's ability to generate own-source revenues (including information on revenue base, tax rates, principal taxpayers, property tax levies and collections).
3. Debt capacity information, to assist users to understand and assess a government's debt burden, and its ability to issue additional debt (including 10 years' worth of data on outstanding debt per capita and as a percentage of personal income, general bonded debt to the value of taxable property and per capita, current year direct and overlapping debt, legal debt margin, the applicable debt ceilings, and 10 years' worth of data on revenue-backed debt).
4. Demographic and economic information, to assist users to understand an entity's socioeconomic environment (such as population trends, personal income, and unemployment rates, and information on the principal employers in the community).
5. Operating information, to provide contextual information about a government's operations and resources (including numbers of government employees, performance and activity measures, such as crime rates, police and fire calls, miles of roadway paved, visitors to city parks, etc.), and information about the volume, usage,

or nature of capital assets (such as roads, bridges, sewers, buildings, etc.).

New Requirement for Narrative Explanations

In addition, governments are encouraged to use narrative explanations to enhance the understandability and usefulness of the quantitative information provided in the statistical section. Such narrative should address the objectives of reporting statistical information in general; identify and explain relationships among the information contained in various tables, and in other sections of the CAFR; and account for any anomalies or unusual trends in the data that CAFR users would not otherwise be able to understand. The new reporting guidelines for the statistical section are applicable for financial reporting periods beginning after June 30, 2005.

ADDITIONAL (OPTIONAL) CAFR SECTIONS

Governments are free to present other sections in the CAFR in addition to (but not instead of) the three basic sections.^[15] The three most common additional sections are:

1. *Investment section.* This may be presented since investing activities may be a significant activity in governments with large pension programs or that sponsor external cash and investment pools.
2. *Actuarial section.* This section would be devoted to the asset and liability calculations of actuaries that are used to establish pension plan funding levels, and the required annual employer contribution.
3. *Single audit section.* Single audits are required under the terms of federal grants and are conducted according to audit standards established by the U.S. Government Accountability Office (formerly the General Accounting Office). The single audit report is typically issued under separate cover, and is included in the CAFR only by reference. However, a reporting entity may elect to include this information in a separate section in the CAFR.

MINIMUM REQUIREMENTS FOR FAIR PRESENTATION

If prepared and published by a government, a CAFR must meet the minimum GAAP requirements for fair presentation. These apply exclusively to the contents of the financial section, and are summarized in Table 1.

FUTURE DEVELOPMENTS AFFECTING THE CAFR

The GASB in June 2004 issued an exposure draft of a proposed new Statement of Governmental Accounting

Table 1 Comprehensive annual financial report: minimum requirements for fair presentation in the financial section

Government-wide financial statements	Fund financial statements		
	Governmental funds	Proprietary funds	Fiduciary funds
<i>Management's discussion and analysis (MD&A)^a</i>			
Statement of net assets Separated into governmental and business-type activities Three-part categorization of net assets	Balance sheet	Statement of net assets; or balance sheet (optional)	Statement of fiduciary net assets
Statement of activities Net cost format Program vs. general costs Program vs. general revenues	Statement of revenues, expenditures, and changes in fund balances	Statement of revenues, expenses, and changes in net assets	[Not applicable]
[A government-wide statement of cash flows is not required]	Budgetary comparison statement (optional to place here or in the Notes)	Statement of cash flows	Statement of changes in fiduciary net assets
<i>Notes to the financial statements</i>			
<i>Budgetary comparisons^b</i>			
<i>Infrastructure condition data^c</i>			
<i>Other required supplementary information (RSI)^d</i>			

^aMD&A is always RSI, which must be presented in accordance with GAAP, but which is unaudited.

^bA budgetary comparison statement (subject to audit) may be used as an alternative to presentation as RSI (which is not subject to audit).

^cRequired if the modified approach is used, and depreciation of infrastructure is not reported.

^dRequired of pension trust funds and public-entity risk pools; the latter are classified as proprietary funds.

Concepts No. 3, *Communication Methods*. As a concept statement, if adopted, the document would not set GAAP requirements per se, but would establish the basis for setting standards in the future. The exposure draft proposes a framework for selecting communications methods to present information in financial reports. The effect of the proposed guidelines would be to more clearly identify which sorts of information should be presented on the face of the financial statements (i.e., displayed there), included in the notes to the financial statements (i.e., disclosed there), or provided as supplementary information, either in the form of RSI, or freely presented as supplementary information. If adopted, the proposed concept statement is expected to be issued towards the end of the second quarter of 2005.

REFERENCES

1. Governmental Accounting Standards Board. In *Codification of Governmental Accounting and Financial Reporting Standards*; Governmental Accounting Standards Board: Norwalk, Connecticut; June 30, 2003; Section 2200.101, pp. 133–134.
2. Codification, Section 2200.102, p. 135.
3. Codification, Section 2100.111, p. 96.
4. Codification, Sections 2100.112–117, pp. 96–98.
5. Codification, Sections 2100.119–136, pp. 98–104.
6. Codification, Section 2100.119, pp. 98–99.
7. Codification, Section 2100.142, p. 106; Section 2600.107–114, pp. 353–355.
8. Codification, Section 2200.105, p. 138.
9. Governmental Accounting Standards Board. Statement No. 34. In *Basic Financial Statements—and Management's Discussion and Analysis—for State and Local*

Governments; Governmental Accounting Standards Board: Norwalk, Connecticut, June 1999; Paragraphs 8–11, 289–295, pp. 107–111; Codification, Sections 2200.107–109, pp. 139–141.

10. Codification, Section 2200.109, pp. 140–141.
11. Codification, Sections 1400.105–108, pp. 27–28.
12. Codification, Section 2200.178, pp. 160–161.
13. Governmental Accounting Standards Board. Statement No. 41. In *Budgetary Comparison Schedules—Perspective Differences*; Governmental Accounting Standards Board: Norwalk, Connecticut, May 2003; Paragraph 3, pp. 1–2; Codification, Sections 2200.179–180, pp. 161–162.
14. Governmental Accounting Standards Board. Statement No. 44. In *Economic Condition Reporting: The Statistical Section*; Governmental Accounting Standards Board: Norwalk, Connecticut, May 2004.
15. Gauthier, S.J. *Governmental Accounting, Auditing and Financial Reporting Using the GASB 34 Model*; Government Finance Officers Association, 2001; 268 pp.

BIBLIOGRAPHY

1. Engstrom, J.H.; Hay, L.E. *Essentials of Accounting for Governmental and Not-for-Profit Organizations*, 4th Ed.; McGraw-Hill Irwin, 1997.
2. Gauthier, S.J. *Governmental Accounting, Auditing and Financial Reporting Using the GASB 34 Model*; Government Finance Officers Association, 2001.
3. Ives, M.; Razek, J.R.; Hosch, G.A. *Introduction to Governmental and Not-for-Profit Accounting*, 5th Ed.; Pearson Prentice Hall, 2004.
4. Wilson, E.R.; Kattelus, S.C. *Accounting for Governmental and Nonprofit Entities*, 13th Ed.; McGraw-Hill Irwin, 2004.

Conflict of Interest

Russell L. Williams

*Askew School of Public Administration and Policy, Florida State University,
Tallahassee, Florida, U.S.A.*

INTRODUCTION

The phrase “conflict of interest” may appear so self-explanatory that it does not require further examination. However, thoughtful scrutiny reveals that many of the pitfalls and failures associated with public service at any level grow out of legal or ethical conflicts of interest. Some consider conflicts of interest to occur only when specific conflict of interest rules or laws are broken. Many of those laws are aimed at specific issues, such as financial disclosure, nepotism, improper use of public resources, or influence peddling. However, in the larger sense, if an individual (or group) functioning under governmental auspices, responsible for making decisions or taking actions affecting the public, places self-interest ahead of public interest in exercising authority, then a conflict of interest exists.

A LONG-STANDING PROBLEM

Arguably, the necessity to address conflict of interest problems was the engine driving the civil service reform movement that started in the late 1800s. Woodrow Wilson’s canonized “The Study of Administration,” written in 1887, clearly recognizes that conflicts of interest form perhaps the most intractable challenge to truly improving and professionalizing public service:

The question for us is, how shall our series of governments within governments be so administered that it shall always be to the interest of the public officer to serve, not his superior alone but the community also, with the best efforts of his talents and the soberest service of his conscience? How shall such service be made to his commonest interest by contributing abundantly to his sustenance, to his dearest interest by furthering his ambition, and to his highest interest by advancing his honor and establishing his character? And how shall this be done alike for the local part and for the national whole? If we solve this problem we shall again pilot the world.^[1]

As obvious as “this problem” was to Wilson, it is also obvious that it was not then, and has not been to

the present, wholly solved. Yet, the modern American civil service system was built on the expectation that government employees’ actions are to be motivated by non-partisan, non-self-interested behavior directed toward the benefit of the public. This expectation was manifested in the Pendleton Act of 1883. Although limited in scope, it was the first substantive effort toward ridding the federal civil service of the blatant conflicts of interest created when the norm was to hire government employees routinely on the basis of their political loyalties. Ironically, however obvious the wisdom of such a reform was to many of the people of the time, it took the impetus created by the assassination of President James A. Garfield by a disappointed office seeker who expected a reward for political service to lead to the passage of that initial template for eliminating conflicts of interest.^[2]

Since the Pendleton Act, efforts have continued to eliminate situations wherein conflicts of interest can occur. Much of what is identified as reforms in “government ethics,” a rather general term, have come to be related to preventing conflicts of interest. This holds true even in the state and local level “ethics in government” initiatives that occurred in the wake of the Watergate government corruption scandal of the 1970s. Many state government ethics commissions or offices focus largely on monitoring financial disclosure statements from public servants, the presumption being that “following the money” reveals whether the potential for conflict of interest problems exists.

CONFLICT OF INTEREST AS A STATUTORY ISSUE

In a country devoted to the rule of law, there has been a logical effort to use laws, rules, regulations, codes of ethics, and even presidential executive orders to define and prevent conflicts of interest by those in public service. Section 201 of Presidential Executive Order 11222, signed by Lyndon B. Johnson in 1965, while not the first attempt at addressing conflict of interest issues by government employees,^[3] established a standard of conduct that not only devoted three of its six provisions to matters concerning conflicts of interest, but also called attention to the need to avoid even

the appearance of such conflicts (the so-called “appearance standard”):

It is the intent of this section that employees avoid any action . . . which might result in, or *create the appearance of*—

1. *Using public office for private gain.*
2. *Giving preferential treatment to any organization or person.*
3. Impeding government efficiency or economy.
4. *Losing complete independence or impartiality of action.*
5. Making a government decision outside official channels; or
6. Affecting adversely the confidence of the public in the integrity of the Government (emphasis added).^[4]

The Ethics in Government Act of 1978, passed during the administration of President Jimmy Carter, reinforced the identification of government ethics with conflict of interest concerns. It too focused on financial disclosure as a means to ameliorate the problem. The act also established the Office of Government Ethics, giving it the responsibility to examine the financial disclosure statements of federal officials. The Ethics in Government Reform Act of 1989 served to strengthen the financial disclosure system, again with the assumption that this was crucial to reducing the potential for conflicts of interest. This approach continues to be echoed in many state and local governments by the use of laws, as well as codes of ethics or conduct.

CONFLICT OF INTEREST AS A PROFESSIONAL ISSUE

Beyond the violation of laws and rules per se, there is another aspect related to conflicts of interest that cannot be overlooked. Intrinsic to the “modern” approach to public administration espoused by Woodrow Wilson is the concept of professionalism (for a discussion of what this might entail, see Ref.^[5]). That is, those who seek to serve the public must treat what they do as more than a mere means to trade labor for wages. They must aspire to levels of professional standards consistent with those expected of doctors and lawyers. Recognizing this, professional associations within the field of public administration also use a statutory (i.e., rule-based) approach regarding conflicts of interest. The codes of ethics of both the American Society for Public Administration (ASPA) and the International City/County Management Association (ICMA) specifically mention the need for government workers to avoid conflicts of interest or

their appearance. ICMA goes so far as to have punitive procedures in place for those found in violation of its code, including those tenets associated with conflicts of interest. In essence, avoiding conflicts of interest is instrumental to meeting the professional standards that define the practice of public administration.

CONFLICT OF INTEREST AS A MORAL ISSUE

Upon his inauguration on January 20, 2001, President George W. Bush repromulgated to all executive branch personnel 5 C.F.R. 2635, *Standards of Ethical Conduct for Employees of the Executive Branch*. Section 2635.101, titled, “Basic obligation of public service,” that contains the following statement:

Public service is a public trust. Each employee has a responsibility to the United States Government and its citizens to place loyalty to the Constitution, laws and *ethical principles* above private gain (emphasis added).^[6]

This is similar in tone to words found in the *Code of Ethics for Government Service*, found in Congressional Resolution 175, adopted by both houses of the 85th Congress on July 11, 1958 and still in effect:

Put loyalty to the *highest moral principals* and to country above loyalty to Government persons, party, or department . . . Uphold these principles, ever conscious that public office is a public trust (emphasis added).^[7]

There is within the above excerpts a clear expectation that public service requires action that is not “merely” legally, but also morally, principled.^a This is couched in terms of where the civil servant’s loyalties must lie. While the legally defined aspects of conflicts of interest can vary from one government entity to another, the ethical aspects do not. While road maps (written laws and rules) exist that, if followed, may allow someone to claim that a specific kind of conflict of interest does not exist, it is usually not so easy to justify actions that, looked at dispassionately, show a lack of regard for the public good. In ethical terms, the government employee is expected to act deontologically.

^aCarol Lewis highlights a classic example of the ramifications of that expectation in her book, *The Ethics Challenge in Public Service* (Jossey-Bass, 1991). She relates the succinct response of George Washington to a friend who was seeking a job: “As George Washington, I would do anything in my power for you. As President, I can do nothing” (Bailey, 1964; 241 pp.).” The root causes and pressures of, as well as the underlying appropriate response to, conflict of interest situations have remained consistent from Washington’s time until now.

There is an underlying moral obligation to do what is best for the public as a whole. This does not leave room for the erstwhile public servant to act egoistically, placing the needs of self, whether for money, power, position, or political advantage, ahead of the public's needs. To do so is antithetical to the concept of American public service, but it is the essence of what constitutes a conflict of interest in the public sector.

THE MOST INTRACTABLE CHALLENGE

In 1887, Woodrow Wilson did not have an answer to the problems stemming from conflicts of interest. Well over a century has passed. While much has changed with the advent of special laws and codes of ethics, somehow much is still the same. Now, as then, it is easy to see why, at the appointive and elective levels of public office, conflict of interest issues create a constant and inevitable tension. No one whose position is so clearly dependent upon the support of specific individuals or groups can be wholly insulated from the pressures that will come from those expecting special treatment in return for that support. However, the danger for the "average," unelected, unappointed public servant is no less considerable, despite, for many, the protection afforded by a merit-based civil service system. Anytime an individual is faced with a situation, no matter how trivial, that places perceived personal gain at odds with the public good, there will be conflicting interests. It may be as simple as favoring a somewhat less-qualified friend over a better-qualified

applicant in the hiring process for a public service job or withholding needed information from another agency in order to make oneself look good at a meeting. The ultimate intractability of the conflict of interest problem is not a function of being able to identify when such a problem exists. Imperfect as they may be, various laws, rules, and codes of ethics provide some assistance in that matter.^b It is in the moral realm, to have or find the strength to act in a manner that is not consistent with one's self-interest, wherein the solution to Wilson's problem is found.

REFERENCES

1. Wilson, W. The study of administration. *Polit. Sci. Q.* **1887**, 2 (2), 197–222.
2. Ingraham, P.W. *The Foundation of Merit: Public Service in American Democracy*; Johns Hopkins, 1995.
3. McCampbell, A.S.; Rood, T.L. Ethics in government: a survey of misuse of position for personal gain and its implications for developing acquisition strategy. *J. Bus. Ethics* **1997**, 16, 1107–1116.
4. Executive Order 11222 of May 8, 1965, 30 F.R. 6469, 3 C.F.R. (1964–1965).
5. Kettl, D.F.; Ingraham, P.W.; Sanders, R.P.; Horner, C. *Civil Service Reform: Building a Government That Works*; Brookings Institution, 1996.
6. *Standard of Ethical Conduct for Employees of the Executive Branch*; 1999; 2635, 5 C.F.R.
7. Code of Ethics for Government Service, Para. 10, H. Con. Res. 175, 85th Cong., 2d Sess., 72 Stat., pt. 2, B12 (adopted July 11, 1958).

^bThe following websites contain many examples and gateways to various ethics laws, rules, and codes that seek to clarify what remains a complex subject:

U.S. Office of Government Ethics at <http://www.usoge.gov/> (accessed December 2001).

Professions Codes of Ethics Online: Government & Military webpage of the Illinois Institute of Technology Center for the Study of Ethics at <http://www.iit.edu/departments/csep/PublicWWW/codes/gov.html> (accessed December 2001).

Los Angeles City Ethics Commission at <http://www.cityofla.org/ETH> (accessed December 2001).

Consolidated Governments

Dan Durning

*Carl Vinson Institute of Government, University of Georgia,
Athens, Georgia, U.S.A.*

INTRODUCTION

The idea of consolidating city and county governments is attractive to a variety of would-be reformers. Since the end of World War II, governmental consolidation proposals have been put forward, studied, or debated every year in dozens of counties and urban areas throughout the United States. However, few proposals have gone to voters, and only about 20% of those have received the needed majority votes.

In this article, the issue of city-county governmental consolidation is examined to identify where and how frequently the consolidation efforts occur, the arguments for and against it, and the evidence about its impacts. A review of the issue shows that consolidation has created heated controversy wherever it has been proposed. It also has generated lively disputes among academic researchers and experts, and a review of the empirical research shows that the debate about the impacts of governmental consolidation is far from settled.

CONSOLIDATED CITY-COUNTY GOVERNMENTS IN THE UNITED STATES

For decades, many researchers and citizens have believed that the fragmentation of local governments—the distribution of governmental responsibilities and powers among multiple governments in a particular location—contributes to pathologies that harm metropolitan areas.^[1,2] Other researchers and observers reject the assertions that governmental fragmentation causes harm. In fact, they argue that it contributes to governmental efficiency.^[3,4]

To remedy the putative problems created by fragmentation, “reformers” have proposed the consolidation of city and county governments. Efforts to consolidate local governments have had limited success: Out of the 3034 county governments in the United States, 39 have consolidated or unified with city governments.

Consolidated governments fall into two groups: those created before World War II and those that came after the war. The pre-World War II group consists of the eight consolidations that date back to 1805, when New Orleans and Orleans Parish were

consolidated. The group includes: Nantucket Town–Nantucket County, MA (1821); Boston–Suffolk Co., MA (1821); Philadelphia–Philadelphia Co., PA (1854); San Francisco–San Francisco Co., CA (1856); New York City and five boroughs, NY (1890s); Denver–Denver Co., CO (1902); and Honolulu–Honolulu Co., HI (1907). These consolidations were created by state legislation.

The post-World War II era of consolidation began in 1947, when Baton Rouge and East Baton Rouge Parish (LA) were consolidated. Since then, 30 other city-county consolidations or unifications have occurred; all but three were approved in local referenda. The Indianapolis–Marion County (IN) consolidated government was formed by an act of the state legislature. The city and county of Broomfield (CO) came through a constitutional amendment approved in 1998 by the voters of Colorado. The city and borough of Yakutat (AK) were incorporated as a consolidated government.

A list of the post-World War II consolidated governments is presented in [Table 1](#). This list shows that most mergers since 1969 have been in smaller metropolitan counties and rural counties, with the exception of the Louisville–Jefferson County (KY) consolidation. Among the recent developments is the case of Camden County (NC) where voters in 2006 approved the creation of a “unified government” in which the county government was given the additional powers of a city government. When the vote took place, the county contained part of Elizabeth City, but had no other cities. As a result of this vote, no cities can incorporate in Camden County and Elizabeth City cannot annex additional land there.

Since 1947, at least 145 votes have been held on proposals to consolidate city and county governments. Voters approved slightly fewer than one in five of the proposals (see [Table 2](#)). In the 1970s and 1980s, only 9 of 79 consolidation referenda (11%) resulted in a consolidated government. However, in the 2000s, 6 of 14 consolidation or unification referenda have been successful.

From 1990 to 2006, referenda on city-county consolidation were held in 13 states. In addition to the 10 mergers approved by voters in eight states (as shown

Table 1 Consolidated city-county governments in the United States

Year	City/county/state	2000 population
1947	Baton Rouge + East Baton Rouge, Louisiana	412,852
1952	Hampton + Elizabeth City County, Virginia	146,437
1957	Newport News + Warwick County, Virginia	180,150
1960	Chesapeake + South Norfolk Norfolk Co., Virginia	199,184
1962	Virginia Beach + Princess Anne County, Virginia	425,257
1962	Nashville + Davidson County, Tennessee	569,891
1967	Jacksonville + Duval County, Florida	778,879
1969	Juneau + Greater Juneau Borough, Alaska	30,711
1969	Carson City + Ormsby County, Nevada	52,457
1969	Indianapolis + Marion County, Indiana	860,454
1970	Columbus + Muscogee County, Georgia	186,291
1971	Sitka + Greater Sitka Borough, Alaska	8,835
1972	Lexington + Fayette County, Kentucky	260,512
1972	Suffolk + Nansemond County, Virginia	63,677
1975	Anchorage + Greater Anchorage Borough, Alaska	260,283
1976	Anaconda + Deer Lodge County, Missouri	9,417
1977	Butte + Silver Bow County, Montana	34,606
1984	Houma + Terrebone Parish, Louisiana	104,503
1988	Lynchburg + Moore County, Tennessee	5,740
1990	Athens + Clarke County, Georgia	101,489
1992	Yakutat + Yakutat Borough, Alaska	808
1992	Lafayette + Lafayette Parish, Louisiana	190,503
1995	August + Richmond County, Georgia	199,775
1997	Kansas City + Wyandotte County, Kansas	157,882
1998	Broomfield + Broomfield County, Colorado	38,272
2000	Louisville + Jefferson County, Kentucky	693,604
2000	Hartsville + Trousdale County, Tennessee	7,259
2002	Haines + Haines Borough, Alaska	2,392
2003	Cusseta + Chattahoochee, Georgia	14,882
2006	Camden County, North Carolina	6,885
2006	Georgetown + Quitman, Georgia	2,598

Note: Except for Indianapolis-Marion County and Broomfield-Broomfield County, the years are when the referendum on consolidation took place. For Indianapolis-Marion County, the date is the year when legislation creating the consolidated government was enacted. For Broomfield-Broomfield County, the date is the year that the creation of the combined city-county of Broomfield was approved in a state referendum. Camden County, North Carolina contains only part of one city. Its vote was to create a "unified government," giving its county government the status of both a city and county government, thereby precluding the incorporation of any new cities or annexation of any land to the existing city.

in Table 1), the proposal was rejected by voters in 27 referenda in 12 states: Georgia (6 referenda), Florida (3), Kentucky (3), Alaska (3), Iowa (2), North Carolina (2), Tennessee (2), New Mexico (2), Virginia (1), Washington (1), California (1), and Kansas (1).

Beyond the 145 referenda on city-county consolidation since World War II, many other efforts to consolidate cities and counties have been undertaken. For example, advocates promoted city-county

consolidation during the past decade in Texas (San Antonio-Bexar Co.), New York (Syracuse-Onondaga Co. and Buffalo-Erie Co.), Pennsylvania (Pittsburgh-Alleghany County), Wisconsin (Milwaukee-Milwaukee Co.), and Nebraska (Omaha-Douglas Co.); however, no votes on consolidation were held in those states.

Consolidation referenda are usually hard fought. Typically, support for consolidation comes from local newspapers, business owners, chambers of

Table 2 Total number of referenda on city-county consolidations and the number of referenda that passed by decade

Decade	Number of referenda	Number and percent of referenda that passed
1940–1949	3	1 (33)
1950–1959	9	2 (22)
1960–1969	20	6 (31)
1970–1979	49	7 (14)
1980–1989	27	2 (8)
1990–1999	23	4 (17)
2000–2006	14	6 (43)
Totals	145	28 (19)

Source: List of referenda on city-county consolidation compiled by the author from different sources. The most complete list of referenda was published by the National Association of Counties.

commerce, and “good government” advocates influenced by the ideas of progressive movement. Opposition is most likely to come from government employees, unions, suburban dwellers who do not want to be part of a larger local government, and racial minorities.^[5]

Recent research by Suzanne Leland and Kurt Thurmaier suggests that successful efforts to consolidate city and county governments are the result of the proponent’s ability to frame the issue as one of economic development.^[6] They wrote:

...successful consolidation attempts involve civic elites who craft a vision of the community’s economic development, craft a restructured local government organization to effectively implement the economic development plan, and convince enough average citizens that the current structure is unable to achieve the economic vision they believe is key to their own economic future. (On page 481 of Ref.^[7].)

When referenda are held, double majorities are usually required for the consolidation to pass. That is, most merger proposal must be approved by a majority of voters living in each city to be consolidated and by a majority of voters who live outside the city or cities to be consolidated with the county government. The double majority requirement has been a difficult one for consolidation advocates to overcome.

IS GOVERNMENTAL CONSOLIDATION A GOOD IDEA?

The advocates and opponents of city-county consolidation disagree about its value. Often, they turn to

research on the topic and to experts to bolster their stances. When they do, they find substantial disagreement about the impacts of consolidation.

The Simple Arguments About Efficiency and Knowledgeable Citizens

For most of the twentieth century, pro-consolidation views dominated the academic discussion of the topic. Early on, proponents cited the founders of public administration, including Woodrow Wilson and Frank Goodnow, in favor of this reform.^[8] Later, in the 1930s and 1940s, leading public administration academics formulated and refined arguments for consolidation. These views reflected the progressive agenda that advocated such things as fewer elected officials, shorter ballots, and the professional city management.

The public administrationists and progressives argued that fragmentation of local governments is counterproductive. They maintained that it causes nearby local governments to be inefficient because they duplicate services and lose economies of size in service delivery. Also, they asserted that fragmentation causes confusion among citizens about which governments are providing them with different services, thus reducing their ability to monitor and influence policy makers. In addition, they contended that governmental fragmentation makes impossible the area-wide planning and action needed to address some of the most important programs facing a metropolitan area. The influence of this perspective can be seen in the recommendations for governmental consolidation and regionalism that came in the 1960s and early 1970s from the Committee for Economic Development, and the U.S. Advisory Commission on Intergovernmental Relations (ACIR).^[1,9]

Opposition to the public administration orthodoxy developed in the late 1960s. It was led by political economists with a public choice perspective on consolidation.^[4,10] The growing influence of this perspective was evident in 1987 when the ACIR issued a report heavily influenced by it.^[3] The perspective of the political economists is now the dominant one among academic researchers addressing consolidation, and it has been termed the “new orthodoxy” of consolidation.^[11]

Political economists reject the idea that local government fragmentation is harmful. They argue that the existence of many proximate local governments contributes to governmental efficiency. They maintain that fragmentation creates a type of market in which local governments compete to attract citizens by offering different packages of services and taxes.^[4,10] In this situation, people are better able to maximize their utility by moving to the local jurisdiction that has the

service and tax combination that best reflects what they want. Because different people have different preferences, they are better able to find a closer match for their preference when many competing local-government tax and service packages are available.

According to this perspective, competition among local governments for residents makes them more efficient than one large consolidated government, which functions as a monopolist influenced by budget-maximizing bureaucrats. Among other things, these multiple competing governments provide citizens with valuable information on the costs of services. Because people can compare the costs per unit of services in their cities with those of neighboring cities, they can make it more difficult for their local government to overcharge for them.

The More Complex Arguments About the Value of Fragmentation and Consolidation

In the 1990s, a group of “new consolidationists” shifted the debate from concern about the efficiency of local governments to concern about equity and economic development issues. They pointed out the social and racial problems of central cities and suggested consolidation as a good cure for them. These “new consolidationists” challenged the “new orthodoxy” of the public choice perspective of consolidation.

A leading proponent and pioneer of the new pro-consolidation perspective is David Rusk, a former mayor of Albuquerque (NM). He argues that inelastic cities contribute to problems that are at the heart of central city problems: poverty; economic, racial, and social segregation; and fiscal disparities.^[2] He noted that poor and disadvantaged citizens are confined to core cities and suggested the creation of merged city-county governments as one way to make metropolitan areas elastic, thereby better able to address the central city problems. Other “New Regionalists” also prescribed city-county consolidation as a way to address the social and economic development problems of metropolitan areas.

In another challenge to the “new orthodoxy,” Lowery’s critique found it weak in many regards, and he formulated a theory underlying the new case for consolidation.^[11] His critique suggested that Tiebout sorting includes not only choosing the best service and tax package but also maximizing race and class proximity. Thus, people tend to locate within governmental boundaries with people of similar race and class. As residents of small cities, they define their self-interest according to what happens within city boundaries, not the boundaries of the metropolitan areas. Thus, they view any effort to use taxes of their

city to help the ailing central city as redistribution, which they oppose.^[11]

Lowery suggested that consolidation is a way to broaden the self-interest perspectives of citizens and to turn decisions to assist central cities into allocation issues, which have lower transaction costs than reallocation decisions. With consolidation, according to Lowery, it is much easier to “enhance equity of social opportunities and promote economic development.”

Political economists have also been refining their arguments against city-county consolidation. They rely less on the Tiebout model of citizen movement among competing local governments to achieve efficiency in fragmented areas. Because Tiebout’s theory includes the assumption that moves within urban areas are costless, it has been subject to criticism. According to Parks and Oakerson, Tiebout’s sorting theory is not necessary to explain why fragmentation creates economic benefits.^[12] The benefits can be derived through a combination of citizen voice and public entrepreneurship. They wrote that voice is amplified and access is increased when the ratio of citizens to elected officials is low. Also, they maintained that by having a multiplicity of local government, many sources of entrepreneurial initiative exist.

Also, public choice advocates have rejected the idea that metropolitan areas are a bewildering mix of unrelated governments. They view metropolitan areas as well-functioning urban economies in which complex inter-relationships have been formed to insure that services are provided at the proper (that is, most efficient) levels. According to Parks and Oakerson, “Instead of textbook fragmentation, real metropolitan areas exhibit varied and complex organizations: a variety of small-, medium-, and large-provision units are linked to a variety of production units.”^[12] They point to intergovernmental compacts, privatization, contracting for service, special districts, and functional consolidations as creating a complex web of relationships that makes a metropolitan area a functioning entity.

What Is the Empirical Evidence About the Impacts of Consolidation?

Many anecdotal accounts of city-county consolidation and some research have found that city-county consolidations have contributed to governments that are more efficient. Nevertheless, the preponderance of research has supported the conclusion that greater fragmentation is associated with lower government costs in large metropolitan areas.^[4,5,13] However, these results do not necessarily extend to smaller and non-metropolitan areas, as shown by the research by Chicoine and Walzer, who found weak evidence that

having fewer governments in a county resulted in lower per capita costs.^[14]

Similarly, anecdotal evidence and limited research have supported the view of proconsolidation advocates that consolidation contributes to more effective government, improved economic development, and greater governmental equity. These assertions and limited research findings have been contested by several other researchers, including political economists. Some of this research has suggested that consolidation does not contribute to economic development or increased citizen participation.^[15]

CONCLUSION: THE FUTURE OF GOVERNMENTAL CONSOLIDATIONS

The recent success of consolidation efforts in Kansas City–Wyandotte County and Louisville–Jefferson County may encourage consolidation advocates elsewhere to try against long odds to consolidate their city and county governments. Even without those successes, it is likely that attempts to consolidate local governments would have continued.

The history of city–county consolidation has shown it to be an “evergreen issue.” In cities that have voted on city–county consolidation in the past, the issue may submerge, but it rarely departs. For example, the 1990 vote to merge Athens and Clarke County (Georgia) was the fourth referendum on the issue and the 2000 referendum in Louisville (Kentucky) was the third vote there.

Just as referenda are rarely the last word on city–county consolidation, the issue of whether governmental consolidation is good public policy is unlikely to be definitively resolved soon. Starting from different assumptions and holding different views of the world, public administration and political economy researchers will likely continue to conduct research with results that support opposing views of city–county consolidation.

REFERENCES

1. Committee for Economic Development. *Modernizing Local Governments to Secure a Balanced Federalism*; Committee for Economic Development: New York, 1966.
2. Rusk, D. *Cities Without Suburbs*; Woodrow Wilson Center Press: Washington, DC, 1993.
3. U.S. Advisory Commission On Intergovernmental Relations. *The Organization of Local Public Economies*; Government Printing Office: Washington, DC, 1987.
4. McGinnis, M. *Polycentricity and Local Public Economies*; University of Michigan Press: Ann Arbor, 1999.
5. Carr, J.B., Feiock, R.C., Eds.; *City–County Consolidation and Its Alternatives*; M.E. Sharpe: Armonk, New York, 2004.
6. Leland, S., Thurmaier, K., Eds.; *Case Studies of City–County Consolidation*; M.E. Sharpe: Armonk, New York, 2004.
7. Leland, S.; Thurmaier, K. When efficiency is unbelievable: normative lessons from 30 years of city–county consolidation. *Public Admin. Rev.* **2005**, *65* (4), 475–489.
8. Stephens, G.R.; Wikstrom, N. *Metropolitan Government and Governance*; Oxford University Press: New York, 2000.
9. U.S. Advisory Commission On Intergovernmental Relations. *Trends in Metropolitan Government*; Government Printing Office: Washington, DC, 1977.
10. Bish, R.; Ostrom, V. *Understanding Urban Government*; American Enterprise Institute: Washington, DC, 1973.
11. Lowery, D. A transactions costs model of metropolitan governance: allocation versus redistribution in urban America. *J. Public Admin. Res. Theory* **2000**, *10* (1), 49–78.
12. Parks, R.; Oakerson, R. Regionalism, localism, and metropolitan governance: suggestions from the research program on local public economies. *State Local Gov. Rev.* **2000**, *32* (3), 169–179.
13. Boyne, G.A. Local government structure and performance: lessons from America? *Public Admin.* **1992**, *70*, 333–357.
14. Chicoine, D.; Walzer, N. *Government Structure and Local Public Finance*; Oelgeschlager, Gunn, and Hain: Boston, 1985.
15. Feiock, R.C.; Carr, J.B. A reassessment of city–county consolidation: economic development impacts. *State Local Gov. Rev.* **1997**, *29* (3), 166–171.

Constitutional Constraints on Administrative Behavior in the United States

David H. Rosenbloom

*Department of Public Administration and Policy, American University,
Washington, District of Columbia, U.S.A.*

Helena K. Rene

School of Public Affairs, American University, Washington, District of Columbia, U.S.A.

INTRODUCTION

U.S. constitutional law currently places broad and substantial constraints on public administrative behavior at all levels of government. Since the 1950s, public administrators' interactions with clients and customers, coworkers, contractors, patients confined to public mental health facilities, prisoners, and individuals involved in street-level administrative regulation have become bounded by constitutional law.^[1-3] Federal court decisions based on a mix of statutory, constitutional, and common law have also made public administrators vulnerable to "constitutional tort" suits, that is, civil suits for money damages to redress their violations of individuals' constitutional rights. This entry explains the legal doctrines that established and defined present-day constitutional constraints on administrative behavior.

RELATIONSHIPS WITH CLIENTS OR CUSTOMERS

Prior to the 1950s, there were few effective constitutional constraints on public administrators' relationships with clients or customers. Recipients of governmental benefits were effectively unable to challenge the constitutionality of administrative conditions attached to welfare payments, public education, occupational licenses, or similar governmental "gratuities" or "largesse." Neither, for the most part, could they successfully use the Constitution's Equal Protection Clause to challenge the distribution of such benefits. Client and customer relationships were governed by the historic "doctrine of privilege." It held that the government had no constitutional duty or obligation to supply benefits, the client or customer accepted them voluntarily and without governmental coercion, the government was, therefore, largely free to establish the conditions on which the benefits could be received, and benefits might even be distributed in a discriminatory fashion under some circumstances.

The key to this approach was a distinction between rights and privileges. Rights are viewed as predating government and protected by the Constitution. In the Declaration of Independence's terms, they are bestowed by humans' "creator" and "unalienable." Privileges, by contrast, are interests established by the government, essentially on whatever terms it chooses. Because clients or customers have no constitutional right to the privileges offered by government, the Constitution was essentially irrelevant to the denial or withdrawal of benefits, no matter how dependent on them recipients may be. Famous examples included the denial of such benefits as public university education to conscientious objectors who would not join the Reserve Officers Training Corps, old-age assistance to a man who refused to live in conventional housing, a license to sell liquor to a man who exercised his privilege against self-incrimination, participation in extracurricular public high-school activities by married students, bartenders' licenses to women (other than the wives and daughters of male bar owners), and licenses to practice law and medicine on account of one's political beliefs.

The doctrine of privilege predated the development of the large-scale U.S. administrative state. In the 1950s and 1960s, it came under increasing criticism as a threat to the constitutional rights of an ever-growing number of clients and customers. By the 1970s, three doctrinal developments had rendered the distinction between rights and privileges meaningless: 1) modern equal protection doctrine; 2) "new property" theory; and 3) reinvigoration of unconstitutional conditions doctrine.

Equal Protection

First, the federal courts reinterpreted the Constitution's Equal Protection Clause. Logically, the doctrine of privilege should not have trumped the guarantee of equal protection of the laws. The fact that there is no right to government benefits does not support the conclusion that they may be offered to members of

one race but not of another. Modern equal protection doctrine, beginning most clearly with public education cases culminating in *Brown v. Board of Education*, prohibits the government from offering a privilege to some but not others, except in a fashion that is consonant with equal protection.^[4]

Today, equal protection has a three-tiered structure. The threshold issue is whether a government policy classifies people by some characteristic, such as race, gender, wealth, residency, or age. The classification may be implicit in a policy or its implementation, rather than explicitly stated in legislation or formal regulations. However, if there is no classification, there is no equal protection issue.

The top tier: Strict scrutiny

Classifications based on race or ethnicity are considered “suspect” because, in view of the nation’s history, it is likely that they will be used to disadvantage minorities. They are subject to strict judicial scrutiny and are constitutional only when narrowly tailored to promote a compelling governmental interest. A heavy burden of persuasion rests on the government, and the courts will not be deferential to its claims. These requirements apply whether the purpose of the classification is benign, as in the case of affirmative action, or invidious, as in the case of racial segregation.

The lower tier: Rational basis

Classifications based on factors that are not immutable, such as wealth and residency, are constitutional when they are rationally related to the achievement of a legitimate governmental purpose. The burden of persuasion is typically on the challenger, and the courts grant considerable deference to the government’s claims. This is called “mere rationality” or “rational basis” analysis. It also applies to classifications based on age. An exception occurs if a non-suspect classification directly interferes with the exercise of a protected constitutional right. For example, a residency requirement for welfare eligibility might abridge the constitutional right of indigents to travel and relocate across state lines. In such a case, the classification is subject to strict judicial scrutiny and must serve a compelling governmental interest in a way that is narrowly tailored or, possibly, by the means that are least restrictive of the protected constitutional rights involved.

The middle tier: Intermediate scrutiny

Classifications based on gender are “quasi-suspect” and face an intermediate test. They must be substantially related to the achievement of important governmental objectives. The courts may ask the

government to provide an “exceedingly persuasive” justification for such classifications, many of which are legacies of earlier practices and beliefs about gender differences that society no longer supports.

Contemporary equal protection doctrine constrains administrative behavior with respect to clients and customers in a wide range of circumstances. Most obviously, it prohibits racial and ethnic discrimination in public education, housing, welfare, health, recreation, occupational licensing, and other programs, except where the government can meet the compelling governmental interest and narrow tailoring tests under strict judicial scrutiny. Invidious discrimination is likely to be justified only under the most extraordinary circumstances, if ever. Benign or remedial racial and ethnic classifications, as in affirmative action, face the same test. To date, the Supreme Court has ruled that diversity in public university legal education is a compelling governmental interest that can be advanced in a narrowly tailored way by making race and ethnicity significant, but not determinative, factors in admissions decisions.^[5] Even though classifications based on gender do not require a compelling governmental interest and narrow tailoring, it is highly unlikely that invidious discrimination against females or males will be constitutional in client–customer relationships. Theoretically, the weaker standard for gender classifications makes it easier to justify affirmative action programs for women and girls than for minorities. However, in practice, there may be no real difference. By contrast, classifications based on other factors, such as age, will most likely be constitutional unless they interfere with the exercise of another protected constitutional right.

Procedural Due Process: The “New Property”

A second doctrinal change contributing to the demise of the doctrine of privilege is called “new property” theory. This approach developed in the 1960s and 1970s. It treats individuals’ governmental benefits as their own property. As such, they are covered by the Due Process Clauses of the Fifth and Fourteenth Amendments, which are aimed at guaranteeing fair procedures and promoting rational decision making. Specifically, these clauses prohibit the deprivation of “life, liberty, or property, without due process of law.” Consequently, once they are defined as a form of property, welfare benefits, public housing, public education, various occupational and other licenses, and other benefits cannot be withheld or terminated (during the period for which they were offered) without a fair procedure.

Procedural due process has a fluid structure. It requires a balancing of three factors: 1) the individual’s

interest in the benefit and/or the severity of the deprivation at issue; 2) the risk that procedure in place, if any, will result in an erroneous decision *and* the probable value of additional procedures in reducing errors; and 3) the government's administrative and financial interests in not using additional or alternative procedures. The underlying assumption in this formula is that, in general, the more elaborate the procedures, the lower the risk of error. Once established by a statute or administrative regulation, a new property interest can be withheld or terminated only in consonance with constitutional due process.

The application of procedural due process is subjective. One must weigh both the individual and the governmental interests. Typically, two issues arise. First, whether the client is entitled to procedural due process, prior to the termination of a benefit during the term for which it was offered. Recipients of means-tested welfare benefits, which may be vital to their survival or ability to use post-termination procedures, generally enjoy a procedural due process right to some pretermination protection against erroneous administrative decisions. An exception exists when there is no factual dispute about their continuing eligibility.

The second issue is how elaborate the procedures required by due process must be. They can range from highly informal notice and opportunity to respond, to a formal response in writing only, to the rudiments of a non-jury judicial trial with a right to counsel, presentation of evidence and witnesses, confrontation and cross-examination of adverse witnesses, a written transcript, a decision by an impartial decision maker, and a right to appeal.^[6-8] In federal administrative law, an impartial decision maker is defined as one who does not have "an irrevocably closed mind." More generally, impartiality prevents those who investigated the case from making the final decision. Sometimes pre- and post-termination requirements are linked. For example, because elaborate post-termination procedures are available to social security disability recipients, only limited procedure is required before these non-means-tested benefits are cut off.

The impact of new property doctrine on administration depends on the specific context. Pre- or post-termination hearings became common in decisions regarding the termination of welfare and disability benefits and occupational licenses. Suspensions and expulsions from public schools are also controlled by procedural due process. It is important to remember that when benefits terminate because they were offered for a fixed term or the recipients have aged, moved, married, or otherwise changed their eligibility status, procedural due process may require nothing more than notice, if that. However, if there is a factual question about the recipient's continuing eligibility, more elaborate procedure may be required.

Unconstitutional Conditions

Reinvigoration of the "unconstitutional conditions" doctrine was the third constitutional change to undercut the doctrine of privilege. The conditions attached to governmental benefits sometimes interfere with clients' and customers' constitutional rights. For instance, a state might mandate that applicants for unemployment compensation be available for work on Saturday despite their Sabbatarian religious beliefs or that public school students recite prayers or the pledge of allegiance to the flag. The unconstitutional conditions doctrine seeks to limit the government's ability to use benefits as leverage for regulating behavior that it could not reach directly without violating individuals' constitutional rights. For instance, under the unconstitutional conditions doctrine, American governments, which lack the power to establish religion directly, cannot do so indirectly through the manipulation of benefits.

The problem public administrators and lawmakers face is knowing which conditions attached to benefits will be adjudged unconstitutional. Essentially, the courts apply a two-part test. As in the case of procedural due process, a good deal of subjectivity is involved. Specific requirements vary with the nature of the benefit and type of condition. Conditions that gratuitously infringe on protected rights, by failing to serve a significant governmental purpose or by prohibiting more than necessary to achieve a legitimate or compelling public interest, are likely to be unconstitutional. Narrowly tailored conditions involving freedom of speech, association, and religion must generally promote at least an important governmental interest. Conditions implicating Fourth Amendment privacy must be reasonable. Under the Fifth Amendment's "Takings Clause," conditions attached to building permits must be roughly proportional to the impact the proposed development will have on legitimate governmental interests, such as traffic congestion.

Taken together, contemporary equal protection, new property theory, and the unconstitutional conditions doctrine bring constitutional law directly into the relationships between public administrators and their clients and customers. They complicate public administration by broadening the set of values that administrators and programs must protect and, sometimes, by requiring elaborate procedure. Procedural due process may not be an efficient process. Narrow tailoring may add expense to programs. The application of constitutional constraints to the distribution of government benefits strengthens the role of courts and judges in this area of public administration, most notably with regard to remedying systematic violations of equal protection, as in desegregating public schools.

PUBLIC PERSONNEL ADMINISTRATION

The doctrine of privilege also applied to public employment. Until change began in the 1950s, a public sector job was considered a benefit or gratuity. Having voluntarily accepted it, public employees could not claim that conditions imposed violated their constitutional rights. As in the case of benefits to clients and customers, racial and gender discrimination in public employment was not controlled by the Equal Protection Clause. Until the 1950s, individuals could be denied public employment or dismissed from it essentially for any reason or no reason at all, including their speech, beliefs, religious practices, reading habits, race, gender, or membership in unions, civil rights, and other organizations.

Cases involving public employment contributed substantially to the demise of the doctrine of privilege. Constitutional constraints became important to public personnel administration as equal protection was applied to public employment, civil service jobs were considered part of the new property, and the unconditional conditions doctrine was invoked to protect public employees' First Amendment rights. However, the constitutional doctrine that constrains public sector human resources management differs from that pertaining to clients and customers.

Constitutional rights within the framework of public employment are now assessed through a "public service model." Its core premise is that the administrative values of efficiency and effectiveness are more important to governments when they act as employers than when they seek to regulate the behavior of ordinary citizens.^[9] Government, like any employer, needs greater authority over its employees than it can exercise over its clients, customers, or the general public. But there are also limits to that authority.

The public service model balances three considerations. First is the interest of the employees in exercising their constitutional rights, that is, in being free of governmental controls. Second is the government's interest in achieving some important purpose as an employer. Third is the public's interest in the way government and public administration operate. The key point in the public service model is that the public's interest may coincide with either that of the employee or that of the government. For example, in the realm of free speech, the public has a clear interest in affording strong constitutional protection to whistle-blowing by public employees. The public is better able to assess governmental performance when public employees can freely expose illegality, gross waste, fraud, abuse, and specific and immediate dangers to public health or safety. In contrast, the public has only a minimal interest, if any, in public employees' partisan campaign speeches, routine office

gossiping, or grouching about politicians and administrative higher-ups. Under the public service model, whistle-blower speech is strongly protected, whereas partisan speech and speech that is of minimal or no public concern are not.^[10-12]

The same kind of balancing applies to public employees' free exercise of religion and association, as well as to their Fourth Amendment privacy rights and guarantee of equal protection of the laws. In assessing constraints on personnel management in these contexts, one always considers the individual's interest in exercising or retaining protected rights, the government's interests that could justify abridgments, and the public's interest. It is important that public personnelists, managers, and employees bear in mind that the governmental interest, although often stated as serving the taxpayers, is not necessarily synonymous with the public interest. Even in the context of public employment, the government's interest in cost-effective administration may not justify infringements on constitutionally protected rights. If there is no clear public interest one way or the other, then the balance is simply between the interests of the employee and the government. The employee will sometimes win because, despite the importance of cost-effective administration, it is not always sufficient to justify interference with protected constitutional rights.

First Amendment Rights

Public employees enjoy broad First Amendment rights under the public service model. The major limitations on their freedom of speech are that, to be constitutionally protected, expressive activity must be on a matter of public concern and it cannot be part of a specific work product or assignment, such as a report or public address.^[13] Additionally, speech that is part of a partisan political campaign may be banned.^[11] Disruptive remarks on matters of private concern only are not protected because they serve no public interest. Restrictions on the partisan political activity of public employees are constitutionally justified by the public interests in administrative impartiality and efficiency, as well as by the protection they afford government workers against being coerced by political appointees and supervisors, to support political parties and candidates.

Public employees' speech and other expressive activity on matters of public concern are most vulnerable to adverse actions under two conditions. First, speech made pursuant to an employee's official duties is not constitutionally protected.^[13] Second, speech may trigger discipline when it impairs discipline or harmony in the workplace, it jeopardizes close working relationships, interferes with normal operations,

or detracts from employees' ability to do their jobs. In balancing the interests of the government, employee, and public, attention must be paid to the nature of the employee's position in the organization. Identical remarks by those at the top and bottom of an organizational hierarchy may be treated differently. When disciplining employees for their off-the-job remarks or other protected activities, the government should be able to show a nexus between its significant interests as an employer and the conduct involved.

The public service model also broadly protects employees' freedom of association and religious freedom. Restrictions on these freedoms have to be justified by the balance among the interests of the employee, public employer, and public. Constitutional law extends extensive protection to employee's freedom of association on the premise that membership alone is not a proxy for behavior. Consequently, employees may join labor unions, employee associations, offensive groups such as the Ku Klux Klan, and even subversive organizations. Exceptions are permissible for categories of employees whose membership in particular organizations would significantly harm the employer's or public's interests. For instance, supervisors and managers can be banned from joining labor unions representing rank-and-file employees, because such membership might conflict with their hierarchical responsibilities. Employees also have the right not to join organizations and, consequently, union shop and maintenance of union membership labor agreements are unconstitutional in the public sector. Since the late 1970s, public employees' freedom of association has been interpreted to make job actions, based on political partisanship, unconstitutional unless the government can show that party affiliation is related to effective job performance.

The parameters of public employees' religious freedom are less clear. The governmental employer cannot coerce employees to engage in prayer or other religious activity. Neither can it gratuitously restrict their religious expression or behavior. Because discrimination based on religion is prohibited by federal equal employment opportunity law, a public employee is ordinarily allowed to wear religious jewelry or headgear, even if other workers find it objectionable. By contrast, restrictions on religious activity might be justified, when overt religious proselytizing or vocal praying at work is disruptive or otherwise interferes with the employer's legitimate interests in workplace harmony and cost effectiveness.

Fourth Amendment Privacy

Public employees retain protections against unreasonable searches and seizures in the public workplace.

The threshold question is whether the employee has a reasonable expectation of privacy under the circumstances. That is, an expectation of privacy that society is prepared to support (according to federal judges). In the absence of such an expectation, public employers are free to search an employee's workspace, desk, files, computer, and so forth. Where there is a reasonable expectation of privacy, the governmental search must be reasonable in its inception and scope. Employees have a reasonable expectation of privacy in their body fluids, but suspicionless drug testing is constitutionally permissible when it reasonably promotes national security, law enforcement, or public safety.

Procedural Due Process

Although a counter trend developed in the 1990s and early 2000s, civil service systems typically provide public employees with new property rights in their positions. Consequently, procedural due process protects them against arbitrary, capricious, or abusive dismissal or demotion. Pretermination notice and opportunity to respond are generally required, as are more elaborate post-termination procedures. Depending on the circumstances, public employers may not be required to offer any procedural due process prior to suspending an employee without pay for wrongdoing. However, the employee should be offered a hearing of some kind within a reasonable time after the suspension takes effect. For the most part, the constitutional requirements for public employees' procedural due process are now written into civil service laws and regulations. However, once a new property right in public employment is created by the government, its termination or diminution is controlled by constitutional, rather than statutory law. Where public employees do not have civil service or tenured status, adverse actions may be "at will" only insofar as they do not violate First Amendment, Fourth Amendment, equal protection, or other constitutional rights.

Equal Protection

The three-tiered equal protection analysis previously discussed applies to public employees as well as to clients and customers. Racial and ethnic classifications, whether benign or invidious, are subject to strict scrutiny. Thus far, the Supreme Court has not held that promoting workforce diversity is a compelling governmental interest. It has permitted benign racial classifications only as a remedy for past, proven, and egregious violations of equal protection, though, logically, the court's language accepting diversity in higher education as a compelling governmental interest seems applicable to civil service jobs as well.^[5] All racial or

ethnic classifications must be narrowly tailored. That is, they must 1) be efficacious relative to alternative remedies that avoid such classifications (such as fining a jurisdiction for violations of equal employment opportunity); 2) have a clear and logical stopping point in terms of time elapsed or policy goal achieved; 3) be demographically proportionate; 4) contain waiver provisions so that the government is never forced to hire someone who is unqualified; 5) not unduly burden the employment interests of nonminority employees; and 6) provide individualized consideration of each qualified applicant.

Residency and age requirements are subject to rational basis analysis. They will be constitutional if they are rationally related to the achievement of a legitimate governmental purpose. Gender-based classifications face a higher standard—they must be substantially related to the achievement of important governmental purposes.

RELATIONSHIPS WITH CONTRACTORS

Governments also face constitutional constraints when dealing with contractors. Some of these are well settled. Governments cannot engage in invidious racial, ethnic, or gender discrimination when contracting out. Contracts that establish new property interests cannot be terminated without procedural due process. In the 1990s, the Supreme Court strengthened two additional types of constraints.

First, the court unequivocally held that benign racial and ethnic classifications in governmental contracting are subject to strict scrutiny. They must serve a compelling governmental interest in a narrowly tailored fashion. This restricted the use of minority business set aside and other preferences. Even though the constitutional test for gender classifications is weaker, preferences for women-owned businesses will probably face substantial limitations, if such litigation reaches the Supreme Court.

Second, the court made it clear that contractors and those engaged in pre-existing commercial relationships with governments, such as firms on a tow truck rotation list, retain First Amendment rights. Using unconstitutional conditions analysis, the court held that the Constitution protects contractors' public criticism of a government and opposition to its officials. Essentially, the same constitutional test that applies to public employees' speech is used to determine the free speech rights of contractors and others engaged in ongoing business relationships with government. A government can terminate its commercial relationship with such parties in response to their speech, only if it can show that, on balance, such action is justified

by the speech's detrimental impact on its ability to deliver public services.^[14]

THE MANAGEMENT OF PATIENTS IN PUBLIC MENTAL HEALTH FACILITIES

Since the early 1970s, constitutional law has required that patients confined to public mental health facilities be afforded treatment. Earlier, they were often "warehoused" indefinitely or otherwise inadequately treated. In *Wyatt v. Stickney*, a federal district court held that confining such patients without treatment constitutes a violation of the right to liberty under the Fourteenth Amendment.^[15] Although the decision applied directly only to the Alabama mental health system, its logic was soon adopted nationwide.

The constitutional right to treatment was eventually defined broadly to include reasonable care and safety, reasonably non-restrictive confinement conditions, and individualized treatment (or training for the mentally disabled). Implementation requires a humane physical and psychological environment, as well as adequate staffing. In practice, this may require far-reaching architectural and medical reforms. The *Wyatt* court eventually required Alabama either to release patients or to provide the following: protection of patients' right to wear their own clothes and to have physical exercise and freedom of religious worship; no more than six patients per room; no single room with less than 100 ft²; at least one toilet for each eight patients and one shower or tub for each 15 patients; not less than 50 ft² per person in the dayroom; at least 10 ft² per person in the dining room; a temperature range between 68°F and 83°F; and various per-patient staffing ratios for professionals, clerical workers, and other staff.

Although judicial involvement in public mental health care was sometimes characterized as interference, eventually the American Psychiatric Association supported the right to treatment. Along with other factors, including better psychoactive drugs, crowding was reduced through deinstitutionalization and shorter stays, and the overall conditions in public mental health facilities improved substantially. However, deinstitutionalization also fostered more homelessness among the mentally ill.^[16]

PRISON MANAGEMENT

Since the 1970s, the treatment of prisoners has also been thoroughly constitutionalized. The Eighth Amendment prohibition of cruel and unusual punishment was reinterpreted to apply to the conditions of confinement, rather than the sentence only. This required prison reforms in more than 40 states to

reduce overcrowding, provide adequate medical care and safety, and guarantee inmates at least a minimal level of nutrition, hygiene, and civilized life's other necessities. The scope of what is prohibited by the Eighth Amendment changes as society progresses and standards of decency evolve. For instance, in the 1990s, prisoners gained the right to challenge subjection to second-hand tobacco smoke, when it poses a risk of serious harm due to personal respiratory or other health problems.

In addition to Eighth Amendment protections, prisoners have a right to procedural due process, when proposed discipline creates extraordinary hardships within the framework of normal prison life. Similarly, within the confines of what incarceration requires in practice, they have the right to free exercise of religion and to marry. The general test is whether restrictions on the prisoners' constitutional rights serve legitimate penological interests, including their impact on guards and other inmates. Prison administration has become so thoroughly infused with constitutional concerns that some prison systems engage in "constitutional audits" to assure their compliance with the latest judicial decisions.

STREET-LEVEL REGULATORY ENCOUNTERS

Public employees whose jobs require constant face-to-face regulatory interaction with the public are often called "street-level bureaucrats." The term includes police, social workers, public school teachers, public health workers, health and safety inspectors, and others who tend to work in the absence of close proximate supervision, have a great deal of discretion in dealing with the public, are largely self-reporting, and are often in a position to have a great positive or negative impact on the individuals with whom they deal. It is also common for street-level bureaucrats to work with heavy caseloads and inadequate resources. Their work environment may be charged with physical and/or psychological threats.

The Fourth Amendment and Equal Protection Clause are the most relevant constitutional constraints on street-level bureaucrats' behavior. However, these controls are loose and subject to evasion.

Fourth Amendment Constraints

Fourth Amendment law is notoriously uneven and fact specific in its application. As noted earlier, it only applies to situations in which there is a reasonable expectation of privacy. The courts treat that expectation as a continuum. On one end, individuals who are in their own homes with the windows and doors

closed and curtains drawn have a great expectation of privacy. On the other, the same individuals would have a minimal expectation of privacy in their baggage when seeking to board a scheduled commercial airplane at a U.S. airport. In between, there are many gradations. In addition, some individuals have reduced Fourth Amendment protections due to their special relationship with an administrative function and its personnel. For example, public school administrators have greater leeway to search students because the schools have special custodial and tutelary responsibilities.

A distinction is usually drawn between law enforcement searches and administrative inspections. Whenever practicable, the Fourth Amendment requires law enforcement officers to seek a warrant from a judge, magistrate, or other judicial official. The warrant is issued on probable cause, which essentially means that there is specific evidence that a violation of law is impending or may exist. When time or other considerations make obtaining a warrant impractical, law enforcement agents can proceed on probable cause alone. Administrative inspections of non-pervasively regulated businesses may also require warrants. However, it is not necessary to show probable cause, as these may be issued when reasonably necessary for the enforcement of the regulatory statute involved. Administrative searches involving special relationships or needs, such as public school discipline, are generally governed by a reasonableness standard.

The constraints on street-level administration generated by this framework often appear inconsistent and, sometimes, close to non-existent. For instance, a federal Occupational Safety and Health Administration (OSHA) inspector may need a warrant to inspect a workplace, depending in part on how open the facility, or a section of it, is to the public. Similarly, a public hospital cannot engage in a law enforcement search of patients' bodily fluids without a warrant or their consent, although it apparently can draw the same fluids for diagnostic medical reasons without Fourth Amendment restriction.^[17] At the same time that an OSHA inspector and public health worker may need a warrant, an individual can be arrested for any traffic infraction, however minor, and consequently subject to a legitimate personal and vehicle search. The Supreme Court has eschewed all inquiry into whether traffic stops or arrests following them are merely pretexts used by the police to search individuals and their vehicles. Even arrest for a non-jailable offense, such as not wearing a seat belt, is permitted.^[18,19]

Equal Protection Constraints

Equal protection constraints on street-level bureaucrats are less complex than Fourth Amendment

requirements, but also subject to circumvention. The threshold question for equal protection analysis is whether a government or its employees classify individuals by some social characteristic such as race, age, or wealth. If there is no classification, there is no equal protection issue. A “classification of one” is cognizable under equal protection, but unfair treatment of particular individuals is typically litigated under procedural due process protection against arbitrary or capricious administrative decision making.^[20]

The main problem for individuals who would challenge street-level personnel’s behavior on equal protection grounds is that the burden of persuasion rests on the private party to demonstrate that a government or its employees have explicitly or implicitly created a classification. Explicit invidious racial, ethnic, or gender classifications are highly unusual in contemporary public administration. Implicit classifications, based on these social characteristics, may be inferred from patterns of administrative behavior. However, establishing their existence to a court’s satisfaction is difficult. This is why the Equal Protection Clause has been relatively ineffective in combating racial and ethnic profiling by law enforcement agents when there is no written or verbal evidence. Unless profiling is the only plausible explanation for the administrative behavior at issue, it may be impossible to show that an implicit classification exists. Of course, equal protection is a more effective constraint against street-level behavior that follows directly from racist, ethnically derogatory, or sexist statements.

PUBLIC ADMINISTRATORS’ LIABILITY FOR CONSTITUTIONAL TORTS

There is often a gap between formal constitutional requirements and actual administrative practice. Constitutional law may be unclear or poorly communicated to administrators. Sometimes, it is highly subjective or requires elaborate balancing of multiple factors. Specific constraints may seem counterintuitive to administrators or impractical due to scarce resources, including time. Furthermore, not everyone whose constitutional rights are violated has either the incentive or the wherewithal to sue the government or its employees for redress.

In a series of decisions crystallizing in the 1980s, the Supreme Court strengthened the application of constitutional constraints on administrative behavior, by reinvigorating the law of “constitutional torts.” The court made it far easier for aggrieved individuals to sue public administrators personally in federal court for money damages, for violations of their constitutional rights. It reasoned that facilitating constitutional tort suits serves both to deter breaches of

constitutional law and to compensate victims. The current standard for administrators’ liability for violating constitutional rights requires that they personally factor knowledge of constitutional constraints into their decision making.

Since 1982, most American public administrators potentially have been personally liable for money damages when their conduct violates “clearly established statutory or constitutional rights of which a reasonable person would have known.”^[21] Under this standard, knowledge of constitutional law becomes a matter of job competence. In the Supreme Court’s words: “[A] reasonably competent public official should know the law governing his conduct.”^[21] An exception is that public administrators cannot be sued for committing constitutional torts while engaged in adjudicatory or legislative functions. There are also some technical differences in how local, state, and federal employees may be sued. However, the vast majority of public administrators are subject to suit, and constitutional torts produce a considerable amount of litigation. Municipalities and their agencies are also vulnerable to such suits, when their policies directly cause violations of individuals’ constitutional rights. Under limited circumstances, they can also be sued for failure to train their employees to protect constitutional rights.

CONCLUSION

During the second half of the twentieth century, the federal courts, often led by the Supreme Court, revolutionized the application of constitutional constraints to administrative behavior in the United States. Constitutional law became central to administrative decisions and operations regarding clients or customers, public personnel systems, outsourcing, the treatment of patients in public mental health facilities, incarceration, and the exercise of regulatory authority in street-level encounters. The courts also vastly strengthened the opportunity for aggrieved individuals to vindicate their newfound rights through constitutional tort suits. Perhaps most important, the federal judiciary now expects public administrators at all levels of government to follow the development of constitutional law, to know how it bears upon their official functions, and to understand that obeying it is a central element of job competence.

REFERENCES

1. Barron, J.; Thomas, D.C. *Constitutional Law in a Nut Shell*; West: St. Paul, MN, 1999.

2. Rosenbloom, D.H.; O'Leary, R. *Public Administration and Law*; Marcel Dekker: New York, 1997.
3. Rosenbloom, D.H.; Carroll, J.; Carroll, J. *Constitutional Competence for Public Managers: Cases and Commentary*; F. E. Peacock: Itasca, IL, 2000.
4. *Brown v. Board of Education of Topeka*. 347 U.S. 483, 1954.
5. *Grutter v. Bollinger*. 539 U.S. 306, 2003.
6. *Goss v. Lopez*. 419 U.S. 565, 1975.
7. *Board of Curators of University of Missouri v. Horowitz*. 435 U.S. 78, 1978.
8. *Mathews v. Eldridge*. 419 U.S. 319, 1976.
9. *Waters v. Churchill*. 511 U.S. 661, 1994.
10. *Pickering v. Board of Education*. 391 U.S. 563, 1968.
11. *U.S. Civil Service Commission v. National Association of Letter Carriers*. 413 U.S. 548, 1973.
12. *San Diego v. Roe*. 543 U.S. 77, 2004.
13. *Garcetti v. Ceballos*. U.S. Supreme Court, No. 04-473, 2006.
14. *O'Hare Truck Service v. City of Northlake*. 518 U.S. 712, 1996.
15. *Wyatt v. Stickney*. 325 F. Supp. 781; 334 F. Supp. 1341; 1971.
16. Jencks, C. *Homelessness*; Harvard University Press: Cambridge, MA, 1994.
17. *Ferguson v. City of Charleston*. U.S. Supreme Court, No. 99-936, 2001.
18. *Atwater v. City of Lago Vista*. 532 U.S. 318; 2001.
19. *Whren v. United States*. 517 U.S. 806, 1996.
20. *Village of Willowbrook v. Olech*. 528 U.S. 562, 2000.
21. *Harlow v. Fitzgerald*. 457 U.S. 800, 1982; 819.

Constitutional Constraints on the Administrations of Public Policy

Matthew Woessner

*Department of Public Affairs, Pennsylvania State University at Harrisburg,
Middletown, Pennsylvania, U.S.A.*

INTRODUCTION

The Constitution of the United States, long celebrated as the greatest political accomplishment of the Enlightenment, is often regarded with less enthusiasm by proponents of strong and efficient public administration. Despite its many accomplishments, this political contract was primarily an instrument of decentralization, designed to frustrate those who might abuse the power of a strong central government. To the extent that the Founding Fathers constrained the new government, they also placed burdens upon servants of the public good by creating a complex chain of political command, establishing restrictive boundaries on federal power, and creating limits on government use of private property. Ultimately, the primary purpose of the Constitution was to establish domestic tranquility without unnecessarily depriving citizens of their basic liberties. The efficiency of public administration was secondary to concerns over a return to monarchy.

To allay concerns that a new central government would become too powerful and overrun the autonomy of the states, the Founders placed specific limitations on federal authority in a deliberate attempt to keep it at bay. From a conceptual standpoint, these limitations are often described in terms of *organizational*, *jurisdictional*, and *property constraints*.

From an organizational standpoint, administrators must know from whom they are taking orders. In the American context, the answer to this question turns out to be quite complex. Further, administrators must be aware of the jurisdictional limitations of their powers under a federal system of shared political authority. The fact that an organizational decree comes from the national government does not, in and of itself, guarantee that the directive is legitimate. The founders designed a layered system of authority in which the state and national governments are given jurisdiction over specific segments of American affairs. Understanding how this has been interpreted will influence how and when administrators choose to act. Finally, the power to act is often tempered by constitutional guarantees of due process and just compensation.

So even with clear directives and unambiguous jurisdiction, administrators must occasionally bow to the rights of property. Although these limitations have varied remarkably over the last century, they remain an important consideration for any scholar of public administration.

ORGANIZATIONAL CONSTRAINTS

Organizational constraints refer to the inherent difficulties of public administration, where the chief bureaucratic officers answer to more than one central authority. In contrast to the classic bureaucratic model, the heads of government agencies do not report exclusively to one individual. Rather, department heads are, themselves, responsible to a large number of different authorities, including but not limited to, the President of the United States, the Speaker of the House, the Majority Leader of the Senate, members of governing committees, subcommittees, and the Congressional membership in general. While it is true that the President has an inordinate degree of influence over the goals and direction of high-level public administrators, they are, nonetheless, compelled to consider the interests of legislators.^[1]

The myth of the presidential bureaucracy stems largely from the fact that department heads serve at the pleasure of the president. To obtain a post at the head of a large government agency, an aspiring cabinet secretary must demonstrate that his appointment will ultimately serve the interests of the chief executive. Given the president's power to terminate top-level appointments, the tenure of a department head will last only as long as he can continue to demonstrate his usefulness. However strong a motivating factor, the power to dismiss high-ranking administrators does not guarantee their loyalty. Beyond the power of appointment, administrators look to Congress to meet their most basic needs, including the confirmation of executive appointments, relevant statutory authorities, and agency funding.^[2] To the extent that public administrators are dependent upon the legislature, agency heads will look to Congress and the President for oversight.

From an organizational standpoint, the job of a public administrator is further complicated by the responsibility to act within the bounds established by the Federal Judiciary. Although the courts do not traditionally make demands on public administrators, the process of statutory interpretation and judicial review forces judges to examine the legal basis for the actions taken at the behest of the political branches. Whenever government takes actions deemed contrary to the letter of the law or the spirit of the Constitution, administrators alter their policies in such a way as to satisfy the concerns of the courts.^[3]

While useful in preventing despotism, the principles of “checks and balances” and the “separation of powers” are generally at odds with classical notions of bureaucratic organization. As a result, nimble public administrators must be capable of pursuing some public good without a single clearly-defined goal from any of the competing branches.

JURISDICTIONAL CONSTRAINTS

Jurisdictional constraints refer to the constitutional limitations placed on federal power in order to prevent the national government from infringing upon the authority traditionally left to states. Before exercising a power established by the legislature, a public administrator must know that the given authority does not conflict with an existing power granted exclusively to the state government or to the people. To do so might risk a lengthy legal battle and, ultimately, repudiation by the judiciary.

From an eighteenth century perspective, concerns over *jurisdictional constraints* were paramount in the minds of public administration, as the newly established national government was granted a narrow list of specific powers meant to provide for domestic security and to promote economic integration. In the execution of administrative power, bureaucrats looked carefully to the enumerated powers listed in Article I, Section 8, of the U.S. Constitution, as a roadmap of legitimate federal authority.

Early battles over the power of the federal government and the scope of its administrative control turned on the last line of Section 8, known commonly as the elastic clause. “The Congress shall have Power To . . . make all Laws which shall be necessary and proper for carrying into Execution the foregoing Powers, and all other Powers vested by this Constitution in the Government of the United States, or in any Department or Officer thereof.” For, while the Founders methodically listed the powers of the new central authority, this final directive would seem to imply that Congress could move beyond its narrowly defined

charter, so as to satisfy one or more of its aforementioned responsibilities.

Early on, the federal courts were receptive to arguments in favor of an expansive federal power. In *McCulloch v. Maryland*,^[4] the Supreme Court was asked to determine whether the charter of the Second Bank of the United States was constitutionally permissible under the powers outlined in Article I. According to its detractors, the federal government lacked the specific authority to charter banks and, therefore, had exceeded its constitutional mandate. Attorneys representing the United States argued that, under the elastic clause, the government was within its right to establish financial institutions for the purpose of strengthening commerce and promoting the general welfare. Writing for a unanimous court, Chief Justice John Marshall argued:

Among the enumerated powers, we do not find that of establishing a bank or creating a corporation. But there is no phrase in the instrument which, like the articles of confederation, excludes incidental or implied powers; and which requires everything granted shall be expressly and minutely detailed . . . After the most deliberate consideration, it is the unanimous and decided opinion of this court that the act to incorporate a bank of the United States is a law made in pursuance of the constitution and is part of the supreme law of the land.

Accordingly, in the exercise of federal authority, officials would, henceforth, not be limited to the enumerated powers listed in Article I, Section 8. Rather, Congress could expand the powers of the federal government to include any influence roughly tied to its responsibility, apart from those specifically forbidden by the Constitution.

Despite the Supreme Court’s recognition of implied federal power in *McCulloch v. Maryland*, the states managed to retain an enormous degree of autonomy, particularly in terms of their domestic affairs. Early attempts to regulate or outlaw child labor were met with stiff resistance by a court predisposed to favor a state’s right to self-determination. In *Hammer v. Dagenhart*,^[5] a private citizen challenged the right of the federal government to set minimum age standards under the Federal Child Labor Act of 1915. The Supreme Court ultimately upheld the decision of the district court, which struck down the act as a usurpation of state authority. Four years later, in *Bailey v. Drexel Furniture Co.* (1922), the Supreme Court would strike down attempts to regulate child labor, this time through taxation. While the court was willing to acknowledge Congress’s implied powers under the elastic clause, it was, as yet, unwilling to let loose that authority, so long as it clashed with the responsibilities

of the states. This, too, would change with the further integration of state commerce and the onset of a national economic emergency in the Great Depression.

The great turning point in the battle over the scope of federal power would begin in 1932 with the election of Franklin D. Roosevelt as president. With the onset of the Depression, the Roosevelt Administration carried a popular mandate to exercise federal authority over the faltering agriculture and manufacturing sectors in an effort to jump-start the national economy. However, the federal judiciary, still accustomed to seeing economic regulations as a state prerogative, struck down a vast majority of Roosevelt's New Deal programs, arguing that congressional regulations of state commerce must be substantially related to matters of commerce between the several states. With the appointment of sympathetic justices and a threat to pack the court, the judiciary gradually relented. Arguing in *National Labor Relations Board v. Jones & Laughlin Steel Corporation*,^[6] the Supreme Court conceded that, in matters of commerce, it is not sufficient that states be granted autonomy over domestic manufacturing, as disputes within those industries could affect the nation as a whole:

When industries organize themselves on a national scale, making their relation to interstate commerce the dominate factor in their activities, how can it be maintained that their industrial relations constitutes a forbidden field into which Congress many not enter when it is necessary to protect interstate commerce from paralyzing consequences of industrial war?

In the decisions that would follow, the Supreme Court gradually expanded the power of congressional administration, using its constitutional authority over interstate commerce as the justification for intervening in any number of state-related activities.

From a practical standpoint, the importance of federal jurisdictional considerations has declined quite dramatically with the growth of a modern interstate economy. Since the court battles of the mid-1930s, the judiciary has scarcely challenged the power of federal officials to regulate state functions under the interstate commerce clause. In one notable exception, *United States v. Lopez*,^[7] the Supreme Court upheld a lower court decision to strike down the Gun-Free School Zone Act of 1990. Under the act, individuals were prohibited from knowingly possessing firearms within a "school zone." Stemming from an earlier decision, the U.S. Congress had grown quite accustomed to asserting its authority in matters relating to the public good under the assertion that it was regulating commerce. This case, however, provided a more difficult situation.

Congress sought to regulate local crime and education, traditionally matters of state control. In addition, the act scarcely provided any evidence that gun possession on school property was in any way related to matters of interstate commerce. Although legal scholars debate the long-term implication of *Lopez*, in striking down the law, the Supreme Court did, if only modestly, reassert the independence of state power and limit Congress' authority to regulate its internal functions.

In the modern context, the constraints of federalism are far less potent than they were when the nation was founded. Nevertheless, under the American constitutional system, administrators of public policy must come to recognize that all federal power operates with specified legal boundaries. To the extent that Congress authorizes the use of federal power outside of long-established precedent, administrators will face a daunting string of court challenges, whether or not the policies enjoy broad public support.

PROPERTY CONSTRAINTS

Property constraints refer to constitutional limitations placed on the government's administration of private resources.

It has been argued that the preservation of property rights was a major motivating factor for the American Revolution. Consequently, safeguards against unreasonable intrusion on privately held resources would eventually become a hallmark of the newfound government; buried inconspicuously within the Constitution's Fifth Amendment:

No person shall be held to answer for a capital, or other-wise infamous crime, unless on a presentment or indictment of a Grand Jury, except in cases arising in the land or naval forces, or in the Militia, when in actual service in time of War or public danger; nor shall any person be subject for the same offence to be twice put in jeopardy of life or limb; nor shall be compelled in any criminal case to be a witness against himself, nor be deprived of life, liberty, or property, without due process of law; *nor shall private property be taken for public use, with-out just compensation.*

Originally, the protections granted under the Fifth Amendment were narrow in scope and application. The meaning, as determined by early American courts, was to prevent the outright confiscation of land by agents of the federal government. Early attempts to broaden the protection to include the mere devastation of property were met with stiff judicial resistance.^[8]

From an administrative perspective, these early constraints were of little consequence in the day-to-day

execution of government policies. The practice of eminent domain, while long established in common law, was rarely ever invoked. The true significance of the Constitution's *property constraints* would not become apparent until the Fifth Amendment protections were expanded beyond the physical occupation of privately held lands.

One of the first successful extensions of the Constitution's property protections came amidst preparation for the Second World War. In *United States v. Causby*,^[9] a family of chicken farmers brought suit against the federal government, arguing that a newly reconstituted Air Force base had destroyed their poultry business and ruined their living space. The vibrations from the airstrip and the sound from the low-flying bombers so upset the livestock that the Causby's could not continue to raise healthy chickens. Relying on traditional notions of eminent domain, the government argued that, absent the physical occupation of their land, the air traffic alone did not justify compensation. The Supreme Court disagreed, arguing that the deprivation of one's livelihood might, under the appropriate conditions, require government compensation. In this case, the restitution was appropriate.

More recently, the courts have come to view the Constitution's *property constraints* in even broader terms, extending compensation for some environmental and zoning regulations. In *Lucas v. South Carolina Coastal Council*,^[10] an oceanfront property owner brought suit against his state for its application of the Beachfront Management Act. Under the regulations, the development of his oceanfront properties would be prohibited. Arguing that the regulations were passed after he had acquired the properties, Lucas suggested that the state legislature had rendered his land worthless. Accordingly, if the state's interests in preserving the land were valid, he must be compensated. The Supreme Court ultimately agreed, ordering the state to provide for the substantial economic damage done to the property as a result of the restrictions.

The overall importance of the Constitution's property constraints is tempered by the prescribed remedy. Even in the most recent rulings, the judiciary has failed to question whether the alleged intrusion on private lands is within the government's valid police powers. Rather, the court has focused, instead, on the circumstances in which the valid use of government power is sufficiently destructive as to warrant compensation. From the perspective of the Constitution's *property constraints*, public administrators need not worry whether a given power is permissible, but rather, what are the public costs associated with any mandated compensation.

CONCLUSION

In assembling the American constitution, the Founding Fathers were forced to confront the tensions between private freedoms and the efficient administration of public resources. By fragmenting the chain of command and limiting the scope of governmental authority, they consciously elected to bias the system in favor of personal liberties.

The *organizational constraints* nested within the "separation of powers" have, to date, thwarted a dangerous consolidation of power. But these organizational safeguards serve to complicate the efficient execution of public policy by forcing administrators to accommodate conflicting political demands.

The Constitution's *jurisdictional constraints* further obscure the administration of public policy by forcing administrators to mind the complicated legal boundaries established between state and federal governments. Even with clear administrative directives, government officials are occasionally reined in by the judiciary for overstepping their constitutional responsibilities.

As if the execution of public policy were not sufficiently complicated, officials must be mindful of the *property constraints*, which compel government entities to provide compensation for the regulation of some privately held resources.

In each of the three examples, administrative efficiency is sacrificed to ensure the preservation of basic personal liberties. However problematic or politically inexpedient, their basis in constitutional law renders many notions of practical reform virtually impossible.

REFERENCES

1. Rosenbloom, D.H.; O'Leary, R. *Public Administration and Law*, 2nd Ed.; Marcel Dekker: New York, 1997.
2. Neustadt, R.E. *Presidential Power and the Modern Presidents: The Politics of Leadership from Roosevelt to Reagan*; The Free Press: New York, 1990.
3. Cooper, P.J. *Public Law and Public Administration*; Mayfield Publishing Company: Palo Alto, CA, 1983.
4. *McCulloch v. Maryland* 4 Wheat. 316 (1819).
5. *Hammer v. Dagenhart* 247 U.S. 251 (1918).
6. *National Labor Relations Board v. Jones & Laughlin Steel Corporation* 301 U.S. 1 (1937).
7. *United States v. Lopez* 514 U.S. 549 (1995).
8. *Barron v. Baltimore* 7 Pet. 243 (1833), 603.
9. *United States v. Causby* 328 U.S. 256 (1946).
10. *Lucas v. South Carolina Coastal Council* 505 U.S. 1003 (1992).

Constructive Change Proposals, Assessing the Validity of

Elisabeth Wright

*School of Business and Public Management, George Washington University,
Washington, District of Columbia, U.S.A.*

INTRODUCTION

The theory of changes is unique to public contracting. While the parties to a commercial contract may negotiate the right to make subsequent changes to the contract in public sector contracting, the public entity generally retains the right, through the contractual language of a changes' clause, to make unilateral changes within some predetermined parameters. This right seeks to preserve the government's fiduciary duty in the expenditure of public funds.

However, changes outside of the contractual right to make changes can occur. Such changes are commonly known as constructive changes, i.e., the changes outside of the authority of a changes' clause. This article examines the nature of the constructive change and offers an approach for examining the validity of claims based on the constructive change argument.

THEORY OF CHANGES

Contracts between commercial and public entities generally contain a clause that permits the public party (i.e., the buyer) to the contract to make changes in specific areas related to or arising from the contract. Such a right protects the expenditure of public funds; that is, work or other contractual matters can be redirected or redefined to ensure that what is actually necessary is accomplished. The ultimate objective is simple and important: wise expenditure of tax dollars in pursuit of satisfying a government need.^[1]

The changes' authority is specific and is limited to certain predefined actions that can be taken only by those who have actual delegated authority to make or direct changes. The authority and scope of changes is limited by design. Decentralized or informal authority to make changes to any contract would undoubtedly cause systemic havoc and chaos—from a contract performance perspective and from a financial management perspective.

Contract performance is framed by the boundaries of the written contract. The supplier is required to furnish the items required by the contract according to the terms and conditions of the written contract. Thus a fundamental requirement of changes is the ability to

identify work performed, but not required by the contract. Changes that have been properly ordered according to the terms and conditions of the contract (i.e., a changes' clause) are considered to be formal changes and are recognized as within the changes' authority of the contract.^[2] This clause gives the government the right to unilaterally change the contract after it has been awarded. The specific language of the changes' clause limits the extent and scope of changes. In consideration of this extraordinary right, the government agrees that it will compensate the supplier for the additional costs of the work as changed, will extend performance or delivery dates, if appropriate, or will compensate with money and time.

The contract requires the government to issue changes in a formal written manner. However, circumstances may arise in which the supplier is directed to perform outside of the formal written contract. For example, a government inspector may impose more stringent acceptance criteria on supplies tendered under the contract. In effect, the inspector will change the acceptance criteria of the written contract; this may cause the supplier to incur additional costs or to experience a schedule delay. When this manner of effecting a contract change occurs, the government can be liable for the costs the supplier incurs in complying with the change. In addition to a cost impact, a schedule impact can occur as the result of a contract change. Changes that take place outside of the explicit authority of the changes' clause are considered as "constructive changes." The supplier attempts a legitimate argument that a constructive change occurred and that compensation is appropriate. That compensation can be in the form of additional money, time, or both. Such compensation is referred to as equitable adjustment.

CONSTRUCTIVE CHANGE

The constructive change can be thought of as an "informal" change to the contract. A formal change is written direction from a government official with the specific authority to direct a change. The formal change is followed by the issuance of a "Change Order," generally in the form of a modification to

the contract. The informal change lacks the formality of the written change; that is, it is a verbal or written act by a government employee that causes a change to the existing contract.

The constructive change can arise from any number of actions or inactions on the part of the government. If the government inspector failed to appear at the agreed upon time and date to accept the supplies, then a constructive change argument might be appropriate. Of course, it is necessary in any change situation that entitlement and quantum be proven. That is to say, mere failure to appear on time to accept supplies does not entitle the supplier to additional compensation. If the supplies were not ready for inspection and the inspector failed to appear, then a constructive change argument would not prevail.

Some examples of circumstances that have been held by boards and courts to constitute constructive changes are the following:

1. Impossibility of performance. Time, money, or impracticality makes it impossible to perform. Drawings or specifications tendered by the government that are impossible to meet would pass the test of practical impossibility. (Recognize that with unlimited time and money, many things are possible, but not practically possible.)
2. Defective specifications. Ambiguous or defective specifications that prevent the work from being accomplished. The supplier may have spent time and money after contract award attempting to comply with the specification as provided by the government.
3. Over-inspection and delays. Actions or inactions by government representatives that may place overly stringent inspection and testing requirements on the supplier after contract award. The supplier may have experienced delay and disruption when supplies rejected should have met the contractual inspection and testing requirements.
4. Additional tasking or direction. Actions that cause the supplier to perform work in excess of the level of work anticipated by the contract.
5. Failure to provide timely and/or suitable government-furnished equipment, property, or information. When the contract calls for items and/or information to be provided by the government, timely availability of those items may be critical to preserving contract price and schedule. Additionally, property or equipment provided as "suitable for its intended use" is warranted for its purpose. Unsuitable equipment that cannot be used by the supplier may constitute a constructive change.

There are many other circumstances that can occur during performance of the contract that provide grounds for claims of constructive change. Government contracting officers must ensure that all government personnel involved in postaward activities comply with stated terms and conditions and are fully aware of the scope of their authority and the potential consequences of their actions.

As a consequence of a constructive change, suppliers prepare and submit a proposal for compensation based on the changed work. The quality of the change proposal is of utmost importance because the government must evaluate the validity of the claim. The claim and its supporting documentation must undergo careful scrutiny and analysis. To that end, evaluators must look for specificity in all information provided. A claim that fails to support entitlement should be returned to the supplier. If entitlement is justified, then quantum may be subject to negotiation.

CONTENTS OF THE CONSTRUCTIVE CHANGE PROPOSAL

At a minimum, the supplier's change proposal should consist of the following elements:

1. Statement of the alleged change. This statement should clearly and precisely outline the basis of the claim and the equitable adjustment requested.
2. Statement of the relevant contract requirement. The contract requirements must be clearly established. The supplier's interpretation of the relevant contract requirement may be included. Attention must be paid to this interpretation because it may be unreasonable or inconsistent with the contract language. For example, the use of the word "shall" in the work statement connotes a requirement while "may" suggests a permissive state. Information given in this section of the claim is a critical component of the claim and forms the foundation for continued review and consideration. It defines what the supplier exactly perceived to be his responsibilities under the contract at the time it was signed.
3. Statement of the government action or inaction that caused the performance of work outside the boundaries of the contract. This statement provides detailed information regarding the government's action or inaction that caused the change. It must support the argument that a gap exists between the contract language and the government's action or inaction. In the absence of a well-written document pointing

- to government “interference,” this statement becomes difficult to prove. For example, a letter rejecting all supplies tendered for failure to meet a tolerance of 1 in. compared to the contract requirement of a 1.5-in. tolerance makes entitlement easier to prove. Evaluators are reminded that government action needs only contribute substantially toward the changed work for a supplier to successfully argue a constructive change.
4. Detailed discussion of the additional work performed. This portion of the claim must include a detailed discussion of the excess work performed. For example, specific information regarding increased scope of duties, increased performance levels, higher personnel standards, etc., must be provided. It is not sufficient for the supplier to make a broad-sweeping statement that “delay occurred as a result of government action.” The link between cause and effect must be clearly established. For example, if an item required rework because of overly stringent testing requirements, then the claim should include amount of rework (hours, dates, etc.) and the detailed bill of material substantiating additional material costs.
 5. Detailed cost proposal supporting the additional work performed. The claim must also include a detailed cost proposal showing the actual costs of the changed work. While some suppliers may attempt to use the total cost method approach toward justifying quantum, evaluators are cautioned that using such a method does not provide a causal link for purposes of additional costs incurred. To avoid such a circumstance, contracts should include a change order accounting clause that requires suppliers to have an accounting system in place that segregates costs of changed work. The cost proposal should contain the direct and indirect costs strictly associated with the constructive change. Sometimes, constructive change determinations are retroactive making the total cost approach the logical choice. However, total cost approach suggests that the difference between the contract price and the new cost of the work (plus profit) represents the total new value of the work. Oftentimes, this is not the case. Inefficiencies and other matters unrelated to the change may find a way into the cost proposal when the total cost approach is used. Detailed cost data should be included as a supplement to the claim and should include all pertinent cost back-up information. Detailed records, such as timesheets, supplier’s invoices, and payroll records, should augment the cost proposal.

6. Compensation requested. The supplier should state the specific compensation requested, i.e., time, money, or both. Remember that the existence of a constructive change does not automatically confer rights of compensation. The supplier must demonstrate a direct link between the alleged cause of the change, the change, and the additional costs incurred.

In addition, the well-documented claim should include supporting information that augments or “backs up” the alleged claim. Information, such as records of conversations, technical write-ups, copies of relevant portions of the contract, data analysis, and basis of cost estimates, are useful in analysis of the claim. Photographs can also lend credibility to the claim. In reviewing the content of the request for equitable adjustment, the evaluator should be cognizant of information that is noticeably absent or of conflicting information either among the documents submitted or among documents not submitted. For example, the supplier’s initial technical proposal may address some aspects of the work in a manner that challenges the constructive change argument.

Evaluators must also be alert to the submission of superfluous information that has no relevance to the constructive change itself. Well-organized, clear, and unambiguous information should provide support to the claim. If necessary, additional information should be requested.

IDENTIFICATION AND INVESTIGATION OF THE ALLEGED CHANGE

As part of the review process, the government must conduct an independent review and analysis. This will include a thorough review of all contractual records not only those contained in the official contract file but any that might exist in other government files. Interviews and meetings with government personnel involved or alleged to be involved provide a valuable opportunity to understand the facts and circumstances surrounding the alleged change.

A detailed technical review of the alleged change must be conducted if technical matters are relevant. Experts in the field are best suited to conduct a review and analysis of the supplier’s technical arguments and presentation of the alleged facts. An independent comparison of the supplier’s information and the government’s information must be made to identify any areas that are not in agreement. Such a comparison may form the basis for the denial of the claim or for negotiation of the claim. The evaluation must focus on the extent to which a cost and/or schedule overrun resulted from the government’s action or inaction.

The government team must establish the baseline contract requirements. Reviewing the contract and any extrinsic evidence, such as precontractual correspondence and records, may point to discussions or clarifications with the supplier. Additionally, the supplier's proposal can provide important information about the supplier's intent and understanding when entering into the contract. Review of the supplier's proposal is a critical part of the investigation that cannot be overlooked. In many contracts, the supplier's proposal is incorporated into the contract by reference. This most often occurs when the government elects to pay a premium for a better technical solution. In such cases, the supplier's proposal forms the baseline for technical expectations.

Recreating the history of what happened and who was involved is important. This can be a difficult task because of personnel change and records transfer. File documentation may not adequately represent the events surrounding the constructive change making the recollections of government personnel important. In such cases, oral findings must be documented and added to the government record.

This investigation is important for several reasons:

1. It enables the government to assess the validity of the claim.
2. It allows the government to isolate the extent to which the government may have caused the constructive change.
3. It enables the government to reconstruct the events for purposes of establishing a monetary and/or time value.
4. It prepares the government to support its position in the event of litigation.

RECOVERABLE COSTS

If entitlement has been determined, then every conceivable type of cost analysis should be performed. However, recognize that the extent of resources used during this effort should be commensurate with the value of the claim. Solid cost analysis will place the government in a solid position when negotiating the quantum aspect of the constructive change claim.

It is generally recognized that in a firm fixed-price bid, a supplier does not include contingency costs. A firm fixed-price contract type indicates a well-defined work statement for which no contingency costs are necessary. Additionally, firm fixed-priced contracts are often awarded to the low-priced bidder. That being

the case, suppliers are not motivated to include "unnecessary costs" into their bid price. If it is possible to reconstruct the bid price using some fundamental assumptions, then analysis of various cost elements and cost-estimating relationships may help develop a good negotiation position. Richard Newman's *Supplier Price Analysis* contains useful information that may assist evaluators in developing solid assumptions about a supplier. Typical examples of recoverable costs include increased labor costs caused by productive time problems, additional man hours, wage increases due to extended performance periods, overtime costs; increased material costs, increased overhead costs and repair and alteration costs. The causal relationship must be established prior to any establishment of reasonableness.

Suppliers are not expected to add costs for anticipated changes to their bid price. The costs caused by such changes, if they occur, are expected to be paid by supplemental agreement to the contract. Although a supplier may identify and quantify all costs incurred that are associated with the change and a reasonable profit, such costs may not be an accurate representation of due compensation. In fact, federal procurement regulations caution that costs incurred are not presumed to be reasonable. For example, inefficiency on the part of the supplier may have contributed to the additional costs. Constructive changes does not provide a supplier the opportunity to recover acquired losses that are not related to the change.

CONCLUSION

Constructive changes present unique challenges for government contract managers. Great care must be taken to avoid situations that can result in constructive changes. However, if such a claim is made, then a thorough identification and investigation of the circumstances and facts surrounding the allegation will ensure that the government's findings and conclusions result in a justifiable position that leaves the parties whole.

REFERENCES

1. Sherman, S.N. *Government Procurement Management*; Wordcrafters Publications: Germantown, MD, 1995.
2. Cibinic, J.; Nash, R. *Administration of Government Contracts*; The George Washington University: Washington, DC, 1996.

Contracting for Social Services

Nancy McCarthy Snyder

*Hugo Wall School of Urban and Public Affairs, Wichita State University,
Wichita, Kansas, U.S.A.*

INTRODUCTION

This article addresses the issues associated with growth in government contracting for human services. Interest in the contracting relationship between governments and non-profit organizations has increased since the 1980s. Political pressure to reduce the size of government has led states and municipalities to become more dependent on non-profits to deliver public human services. This trend has been reinforced by deinstitutionalization of disadvantaged populations. Non-profit organizations have stepped forward to design and deliver community-based services for the elderly, the mentally ill, abused and neglected children, and the developmentally and physically disabled. As a result, non-profits have become increasingly dependent on government.

Growth in the non-profit sector and heavy reliance on tax-supported revenue have led to increased scrutiny and concerns about the management capacity of non-profit agencies, particularly in human services where executive directors are typically promoted from line positions and lack training in finance and management. Therefore, the trend toward contracting raises important challenges for governments. Public officials have less control over contracted services than they do over directly produced services. This article further discusses the need for government capacity to monitor the contracting relationship and to ensure accountability for public resources.

HISTORICAL CONTEXT

The primary challenge of public finance is always resource allocation, including how the level and content of government services fits in the overall mix of public, business, and non-profit sector production. The history of social welfare in the United States is a case study in federalism, as well as in the division of responsibilities between the public and private sectors. Relationships between governments and non-profit social service agencies have a long tradition in the United States. At the turn of the twentieth century, it was common for local governments to fund orphanages, hospitals, and relief organizations that were operated by private organizations.

The Social Security Act of 1935 brought the federal government into the social welfare policy arena. However, the federal role was limited, with the exception of retirement insurance, to small grants to states for a few welfare programs. During the 1960s and 1970s, the federal role expanded greatly with the creation of new programs (e.g., foodstamps, Medicaid, Medicare) and large increases in existing programs (e.g., Aid to Families with Dependent Children, child care, and social services). The 1980s was marked by retrenchment in many areas of social welfare. By fiscal year 1997, social service agencies received 20% less from the federal government (as a percent of gross domestic product) than they had in 1980.^[1] In addition to reducing relative spending, the federal government devolved operating responsibility for many programs to state and local governments. At the same time, major changes in policy toward services for the elderly, the mentally ill, and physically and developmentally disabled individuals resulted in the creation of community-based programs to replace institutional care. Thus, at the turn of the twenty-first century, state and local governments had considerably more responsibility for human services and turned increasingly to non-profit organizations to deliver services.

Two major factors have increased attention to the contracting relationship between governments and non-profits organizations. First is a general growing interest in the U.S. non-profit sector. The overall size of the sector increased dramatically during the second half of the twentieth century, from 50,000 to approximately 1.5 million organizations, over one-third of which are “charitable” (i.e., eligible to receive tax-deductible contributions; other categories of non-profit agencies are religious congregations, foundations, and mutual benefit organizations, such as labor unions and professional associations). Today, the non-profit sector, broadly defined, receives revenue equal to almost 9% of gross domestic product and employs roughly 11 million people or 7% paid employment.^[2] This success has led to increased scrutiny and concerns about the management capacity of non-profit agencies,^[3,4] particularly in human services where executive directors are typically promoted from line positions and lack training in finance and management. Attention to non-profit operations has also grown because human service non-profits have become increasingly

dependent on government. By 1980, over one-half of the revenue of these agencies came from the public sector.^[5]

The second trend that has stimulated interest in contracting is political pressure to reduce the size of government. For the most part, Americans have always been suspicious of the public sector, but the period from 1935 through 1980 was marked by significant expansion in all levels of government. Although the trend toward contracting with non-profit agencies for human services predates the strong privatization ideology of the last two decades, the movement has been accelerated by the desire to limit the size and scope of government that has been evident since 1980.

THE CONTRACTING DECISION

The contracting decision involves two major questions: what services should government produce itself, and what should it contract out? Second, when it does contract services, what mechanisms should be employed to ensure effective service delivery?

Governments contract with non-profit and for-profit agencies for services for various reasons—most often to save money, to limit the size of the public labor force, or to use specialized expertise (e.g., legal, architectural, or construction services), for which there is insufficient need for in-house production. In the case of human services, there are additional motives. Among them are to offer consumers greater choice in service providers, to minimize bureaucratic red tape that restricts or slows innovative policy, a belief that non-profit values are more suitable to human service delivery, and a desire to avoid political blame for the failures that inevitably occur. In some jurisdictions, contracting is simply a matter of history or habit. Relationships between governments and non-profit agencies are well established and are maintained with little thought given to their costs or advantages.

Most recently, the increase in contracting has been reinforced by ideological beliefs that markets and competition work better than governments. The traditional accountability tools employed in contracting with for-profit firms are unlikely, however, to be successful with human service non-profits. Contracting works best when market forces are present (i.e., a demand and a supply made up of many buyers and many sellers who compete with each other). To capture competitive efficiencies, the market should be more or less perfect: balanced and complete information, no externalities, increasing costs, homogeneous products, rival consumption, etc. The more imperfect the market, the lower the potential benefits to taxpayers.

The market for human services is characterized by imperfections on both the demand and supply sides.^[6] There is often a single buyer (a monopsonist), either

the state or local government, purchasing specialized services from relatively few suppliers. The problem of limited supply is particularly acute outside large urban areas. The presence of market imperfections means that government has limited alternatives, increasing its interdependence with private providers, exacerbating the problem of acquiring accurate information to make contracting decisions and raising the potential for conflicts of interest. The result is similar to a bilateral monopoly, the very antithesis of competitive efficiency. As Don Kettl pointed out: “a too enthusiastic embrace of the competitive prescription might produce the wrong medicine for a misdiagnosed disease.”^[6] Nevertheless, the trend toward expanded contracting is unlikely to be reversed in the foreseeable future. As a result, governments need effective tools to manage the contracting process.

ACCOUNTABILITY

When governments produce services directly, accountability to the public is assured through the electoral process and the presence of hierarchical bureaucracies that make line workers answerable to supervisors who are ultimately answerable to elected officials. When governments contract for services with for-profit firms, market competition enforces accountability for cost and service quality. Non-profit organizations lack both political and market feedback systems to ensure accountability.

Early financing relationships between government and non-profits involved grants with few strings attached. As government has come under increased scrutiny and demands for accountability, public officials have passed those demands on to non-profits. Increased contracting for human services creates challenges for both governments and non-profits to develop additional methods of assuring the public that its resources are being efficiently and effectively used. Contracts essentially substitute for markets and typically incorporate two different accountability models:^[3] standards/compliance and outcomes/performance.

Standards/Compliance

The standards/compliance approach is the most commonly used method of ensuring accountability. It compares agency operations with a set of established standards. It asks questions such as: How many clients were served? Were clients accurately screened to determine eligibility for services? What services and how much of each service did they receive? Are client records complete and current? Are procedures in place to protect client confidentiality? Do employees possess the appropriate professional licenses? Are clients made aware of

the agency's process for appealing a service decision? Does the agency have a governing board of at least seven members that meets at least four times per year? Are copies of last year's audit and IRS Form 990 on file?

The standards/compliance model is used extensively because it is the cheapest and most easily managed accountability tool. Standards are relatively easy to write into contracts and auditors can easily monitor and document compliance. The major disadvantage of this model is that it does not guarantee effective services. It is possible to meet all standards and not achieve program goals.

Outcomes/Performance

In recent years, funders of human services have placed greater emphasis on identifying program outcomes and measuring performance toward achievement of those outcomes. This outcomes/performance model of accountability addresses many of the weaknesses of standards/compliance techniques. It keeps agencies focused on desired outcomes, rather than on mere service levels. It starts with the underlying question of "what is this program trying to achieve?" It then asks questions such as: Is the client improving? Has the client gained knowledge or changed their behavior? Is the client avoiding negative outcomes (e.g., hospitalization, incarceration, truancy)? Is the client satisfied with the services received?

Outcomes/performance accountability has been difficult to implement. Non-profit agencies argue that much of the work in human services is inherently unmeasurable.^[7] The outcomes/performance model is also hampered by the fact that consistently effective treatments do not exist for some conditions. So even when measurable outcomes are known, dependable interventions may not be available to achieve desired outcomes. Contracting agencies are naturally resistant to being held accountable for outcomes that are not consistently achievable or when factors outside the contractors' control (e.g., judicial decisions) affect outcomes.

Because human service goals are hard to define, results hard to measure, and treatments uncertain, it becomes difficult to write meaningful quality specifications into requests for proposals (RFPs) and contracts. This creates a tendency to substitute client satisfaction or professional standards for quality. Although satisfaction is a legitimate concern, what happens if the client is satisfied, but does not appear to be improving?

Professional standards have their own set of problems, the greatest of which is cost. Although most national accrediting bodies would argue that they enforce minimal standards, they tend to have biased views of what is minimal or acceptable. Substituting professional standards for quality measures (e.g.,

client-staff ratios) can have the effect of forcing government to purchase a higher level of service than the public generally demands. Decision makers need to be able to distinguish a "Cadillac" from a "Chevy" in terms of service quality and allocate resources with knowledge of those distinctions. However, despite these weaknesses, outcomes/performance approaches show considerable promise for improving the effectiveness of human service programs.

Capacity Building

It has been convincingly argued^[4] that overemphasis on outcomes can create problems for non-profits and ultimately for society. As funders (United Ways and foundations, as well as governments) have become more outcomes conscious, they have moved more and more toward program funding (as opposed to general grant funding). The result is that it is increasingly difficult for non-profits to use "profitable" programs to subsidize unprofitable ones or to fund administrative overhead.^[7] Outcomes are more expensive for agencies to document and for funders to monitor. Social service agencies are generally resource poor relative to other non-profits. In 1996, human service organizations comprised 40% of active charities (those that file IRS Form 990), but expended only 13% of the resources.^[8] Many lack up-to-date computer technology, appropriate accounting systems, and generally accepted management practices.

Analysts of non-profit reform^[3,4] have begun to argue that far too little attention has been given to the potential of capacity-building in the non-profit sector as a way to ensure accountability. They claim that effective non-profit organizations are more important than effective programs in providing lasting social impacts. Programs do not stand alone and are a weak base for large scale social reform because client needs and the operating environment are always changing. A capacity-building model of accountability asks questions such as: Does the agency have the ability to identify changing community needs? Can the agency adapt to shifting needs and operating environments? Does the agency possess the information systems to accurately track finances and client outcomes? Does agency staff possess the managerial, financial, and technical skills to design, fund, and implement innovative programs?

The advantage of capacity building is that it institutionalizes responsiveness, learning, and innovation, and ultimately reduces the need for detailed contract monitoring. It is, however, a more expensive and long-term accountability method than either standards/compliance or outcomes/performance, requiring efforts and resources that may be beyond those of a single government.

GOVERNMENT CAPACITY

Contracting imposes significant transaction costs on both governments and non-profits. The latter must adapt to short lead times, voluminous reporting requirements in formats that vary among funders, uncertain political climates, limited discretion, divergence from declared mission to respond to government service priorities, fragmented planning and program structures, and cost shifting resulting from demands that they supplement contracts with their own resources.^[9] Both governments and non-profits need good information systems, the full cost of which is rarely covered in contracts.

Governments incur other transactions costs that are typically not calculated in the cost of contracting. It has become a truism that “contracts do not manage themselves.” It takes time and resources to effectively monitor the performance of non-profit service providers. Particularly when employing outcomes/performance accountability, government agencies need staffs who can write detailed and specific RFP and contract specifications, identify effective rewards and penalties, and who know how to advertise and solicit bids, review proposals, resolve technical disputes, analyze costs, negotiate final^[6] agreements, and monitor for service quality and costs. They need a strong mix of program expertise and contract management skills, as well as communication skills and sound judgment. Staff members with long histories of monitoring standards/compliance-type contracts often find it difficult to loosen the control necessary for successful outcomes/performance contracting.

Effective contract enforcement requires independent audits of reported outcomes. In addition to hiring individuals with the requisite skills, governments need to have sufficient internal resources to effectively manage the existing contract caseload in both program and legal departments. From a broader policy perspective, contracting makes it more difficult for government managers to develop a clear jurisdictionwide picture of services and beneficiaries, hindering effective policy planning.

The challenge for governments is to assist in capacity building in the non-profit sector, as well as to instill best practice in non-profit management and operations, while resisting the temptation to dictate, dismiss, duplicate, or micromanage the contracting agencies. In an era of privatization, government performance depends on contractor performance. “[G]overnment is no longer a buyer dealing at arm’s length with a seller, but a partner in a virtually seamless, mutually dependent interrelationship.”^[6]

The most effective partnerships incorporate joint planning in identification of need, service design, and specification of performance measures. Non-profits are more likely to embrace outcomes/performance

accountability if they participate in selection of the measures and targets. Partnerships are characterized by mutual respect for the values and demands of both parties. Public managers need to develop a deeper appreciation of the strengths of the non-profit sector^[10] while they work to teach non-profit managers to speak the language of performance, outcomes, and accountability. Successful contracts depend more on the personal relationships between contract monitor and program director than on legal specifications. Communication and trust can overcome many technical hurdles.

POLICY ISSUES

Contracts add organizational layers between taxpayers who finance human services and the non-profit workers who deliver the services. This inevitably increases administrative costs and hinders effective communication. It also complicates questions of accountability for public resources, and of who gets political credit for successes and blame for failures.

The recent push toward privatization is based less on clear evidence of its benefits than on ideology. Historically, human service non-profits have developed informal networks of like-minded agencies to serve disadvantaged populations. The advent of winner-take-all contracting and the forcing of competitive “business-like” practices on individuals and organizations that are naturally non-competitive appears to be breaking down the cooperative infrastructures that makes non-profits so appealing as service delivery agents (i.e., values of compassion and service, responsiveness and flexibility, disregard for profit motives, willingness to assume contracts at less than full cost of production, acceptance of below-market wages, openness to sharing information).

Public sector human service systems (e.g., public welfare, mental health, developmental and physical disability services, senior services) were created in the mid-twentieth century because the private social welfare system was unable to meet public demand for serving the most difficult client populations. State and local governments are now asking non-profit organizations to assume a larger portion of service delivery to the most needy clients. This raises questions about the sustainability of non-profits’ willingness or ability to assume public duties. It also begs the question of what is public and what is private.

CONCLUSION

Overall, contracting for human services is a public policy tool that can serve useful purposes. It is not a substitute for answering the underlying policy questions. What services should be contracted and with

whom? What benefits are achieved if a government service delivery monopoly is replaced with a non-profit monopoly? Are efforts to improve accountability forcing uniformity that undermines the diversity and flexibility of non-profit service delivery? With what justification do governments treat non-profit contractors different from for-profit contractors? When resources are inadequate to serve all those who are eligible, what allocation decision should be made (e.g., restricted eligibility, waiting lists)? Who should make those decisions? How thin can resources be spread before they become ineffective for clients? Is the public willing to accept the reality that there are some clients who are unlikely to improve no matter how many services are provided, or that in the best human service delivery systems, there will be occasional failure? Finally, who is ultimately responsible for caring for the hardest to serve, most vulnerable and needy populations?

REFERENCES

1. Abramson, A.J.; Salamon, L.M.; Streuerle, C.E. The non-profit sector and the federal budget: recent history and future directions. In *Non-profits and Government: Collaboration and Conflict*; Boris, E.T., Steuerle, C.E., Eds.; The Urban Institute Press: Washington, DC, 1999; 99–139.
2. Salamon, L.M. *America's Nonprofit Sector: A Primer*, 2nd Ed.; The Foundation Center: USA, 1999.
3. Light, P. *Making Nonprofits Work*; Brookings Institution Press: Washington, DC, 2000.
4. Letts, C.W.; Ryan, W.P.; Grossman, A. *High Performance Nonprofit Organizations: Managing Upstream for Greater Impact*; John Wiley & Sons: New York, 1999.
5. Hall, P.D. *A History of Nonprofit Boards in the United States*; National Center for Nonprofit Boards: Washington, DC, 1997.
6. Kettl, D.F. *Sharing Power*; The Brookings Institute: Washington, DC, 1993; 13, 201.
7. Smith, S.R.; Lipsky, M. *Nonprofits for Hire*; Harvard University Press: Cambridge, MA, 1993.
8. Boris, E.T. Nonprofit organizations in a democracy: varied roles and responsibilities. In *Nonprofits and Government: Collaboration and Conflict*; Boris, E.T., Steuerle, C.E., Eds.; The Urban Institute Press: Washington, DC, 1999; 3–29.
9. Gronbjerg, K.A. Transaction costs in social service contracting: lessons from the USA. In *The Contract Culture in Public Services: Studies from Britain, Europe and the USA*; Perri 6, Kendall, J., Eds.; Arena, 1997; 99–113.
10. Frumkin, P. *Balancing Public Accountability and Non-profit Autonomy: Milestone Contracting in Oklahoma*; Harvard University: Cambridge, MA, 2001; http://papers.ssrn.com/paper.taf?abstract_id=269361.

Control Systems

Lynda M. Dennis

Public Administration Department, University of Central Florida, Orlando, Florida, U.S.A.

INTRODUCTION

Control systems in governmental organizations vary greatly based on the type of organization as well as its size and service capacity. These systems help assure accountability with respect to the receipt and expenditure of public funds.^[1,2] Effective control systems require cooperation between and among officials, management, and staff at all levels of the organization.^[3] At the end of the day, elected officials are ultimately and directly accountable to citizens and stakeholders (external accountability). Management and staff of the governmental organization, however, are accountable and responsible for their actions day after day (internal accountability).^[4] Therefore, public policy efficacy is determined in large part by the quality of financial management information and its underlying control systems.^[5]

Safeguarding the assets of the governmental organization has been the traditional goal of most control systems and this was achieved through any number of methods, policies, and procedures. As such, these control systems focused on producing accurate and reliable financial information and encouraged adherence to the established policies and procedures.^[6] A secondary focus of the traditional control systems was to promote operational efficiency primarily by establishing methods, policies, and procedures^[6] based on their cost–benefit to the organization.^[2]

There are a number of characteristics unique to governmental organizations that influence the design, implementation, execution, and review of their respective control systems. Governmental organizations are established by, and operate under, a myriad of laws, regulations, statutes, and rules; and, as such they are subject to far more legal constraints than nongovernmental organizations. The political aspect of governmental organizations may result in elected officials that are not familiar with governmental activities, the impact of regulations, and the totality of external accountability.^[2] This coupled with finite terms of office and/or term limits may cause them to make short-sighted decisions and/or focus more on visible results rather than the strength of the underlying control systems.^[7] Senior management staff appointed by elected officials, acting alone or as a governing body, may find themselves subject to the same political

influences as the elected officials.^[2,7] The public aspect of governmental organizations often results in elected officials and management being held responsible by the press and constituents for insignificant matters. This same “publicness” also requires elected officials and senior management to respond to the results of audits and other regulatory reviews.^[2] The education and experience of an organization’s leadership, management, and staff influence whether control systems have a control, management, or planning orientation. In a control oriented organization control systems focus on completion of specific tasks and strict observance of policies and procedures. Efficiency and effectiveness in daily operations are hallmarks of a management orientation while a planning orientation focuses on organizational goals and how to achieve them.^[8] In small and medium governmental organizations, there may be a limited number of administrative personnel and/or a limited number of qualified personnel in administrative positions. The lack of qualified personnel in these governmental organizations may affect the orientation, scope, and effectiveness of its control systems.

Overview

This chapter discusses control systems in governmental organizations in the context of accountability from both a traditional and innovative perspective. Stakeholders insist that their governments provide services economically, efficiently, and effectively.^[1] They want to know that programs are not only operating as designed and achieving the desired results but that this is accomplished at a reasonable cost.^[3] As such, managers and elected officials are accountable to all stakeholders for the activities of their organizations and the related results.^[1] In this chapter, control systems, as a measure of accountability for public resources, are presumed vital to the governance process and an essential element for a healthy democracy.^[1]

CONTROL SYSTEMS AND ACCOUNTABILITY

Financial or performance audits performed by independent auditors external to the governmental

organization are considered the most objective measure of accountability for governmental organizations.^[8] External auditors performing audits of governmental organizations are usually certified public accountants (CPAs) that conduct their audits in accordance with professional auditing standards referred to as generally accepted auditing standards (GAAS). Auditing standards related to financial statement audits of nonpublicly traded companies are promulgated by the Auditing Standards Board of the American Institute of Certified Public Accountants (AICPA). All CPAs conducting audits of nonpublicly traded entities must abide by these auditing standards.^[2]

Auditors conducting financial or performance audits of governmental organizations may be required to also adhere to government auditing standards (GAGAS) which are promulgated by the Comptroller General of the United States and the Government Accountability Office (GAO). These additional government auditing standards apply to financial and performance audits of governmental organizations receiving government assistance (i.e., federal grants) at or above certain statutory amounts (currently \$500,000 in the aggregate).^[1] A number of states require that audits of local governments located in their states be performed in accordance with GAGAS (in addition to GAAS). Additionally, some state auditors promulgate standards, rules, or regulations that auditors must adhere to in financial or performance audits of governmental organizations. These requirements typically apply regardless of the level of federal government assistance received and/or if state financial assistance is received.^[1,2]

Auditing standards related to governmental organizations (GAAS and GAGAS) require the audit be planned to provide “reasonable assurance” the financial statements are free of material mis-statements due to illegal acts having a direct and material effect on financial statement amounts. Illegal acts include violations of laws, rules, regulations, or contract/grant provisions.^[1,2] In audits of governmental organizations where GAGAS are required, illegal acts also include instances of “abuse” (which is different from violations of laws, rules, etc.) if such instances could significantly affect financial statement amounts or program performance results. Abuse occurs when a program/entity falls short of reasonable and necessary businesses practices expected by a prudent person.^[1] The standard of “reasonable assurance” is also extended to material misstatements due to errors or fraud.^[2]

Inherent in both GAAS and GAGAS is the requirement for the auditor to review the governmental organization’s system of internal control.^[1,2] Because of this auditing standard, financial managers in a governmental organization may tend to be control oriented in their approach to the organization’s control system.

Unfortunately, managers in other functions within the organization may believe such controls to be onerous.

The auditing standard related to reviewing the system of internal control requires that the auditor obtain and evaluate evidential matter concerning the amounts reported by management in the financial statements and/or supporting books and records. To do this, the external independent auditor will examine such supporting documentation as vendor invoices, bank statements, billings to customers, etc. The purpose of this review is to determine the legitimacy of the transaction, accuracy of transaction amounts and coding, whether any unrecorded transactions exist, and/or the existence of any related rights and/or obligations. In addition to the detailed review of the internal control system, the external independent auditor will also review the control environment of the governmental organization.^[2] This may include ascertaining the following:

1. Management’s reputation in the community;
2. Management’s attitude, awareness, and actions concerning the control environment;
3. Effectiveness of overall financial controls related to operating within approved budgets and management’s ability to produce timely and accurate financial reports; and
4. Degree of centralization/decentralization in the accounting systems and whether or not it is warranted in the circumstances.^[2]

A financial or performance audit conducted by an independent auditor external to the governmental organization is a key component of providing accountability to stakeholders. In the case of audits performed under GAGAS, the auditor also reports on the governmental organization’s system of internal control noting any significant deficiencies in it found during the audit.^[1] The financial or performance audit allows stakeholders and other parties to have confidence that the reported financial and/or program operating information is reliable.

MANAGEMENT’S RESPONSIBILITY FOR CONTROL SYSTEMS

In a governmental organization, management as represented by elected officials and senior staff is responsible for its control systems. As such, management is concerned with the efficient, economic, and effective application of resources in order to legally achieve the purposes for which they were established.^[1,9] Management is responsible for ensuring its control systems comply with all applicable laws, regulations, statutes, and rules.^[1,2] Additionally, management is responsible

for providing the public, oversight agencies, and other stakeholders with appropriate reports showing accountability for resources used and program results.^[1,2] To summarize, control systems established and maintained by management should ensure the following:

1. Appropriate goals and objectives are met;
2. Resources are used efficiently, effectively, and economically;
3. Resources are safeguarded;
4. Applicable laws, regulations, statutes, and rules are followed; and,
5. Data produced is reliable and fairly disclosed.^[1,2]

Statement on Auditing Standards No. 99, *Consideration of Fraud in a Financial Statement Audit* (SAS No. 99), was issued by the AICPA in October of 2002. As defined in this statement, fraud is an intentional act that results in material misstatements in the financial statements that arises from fraudulent financial reporting and/or misappropriated assets. Generally, fraud occurs when three conditions are present: 1) incentive/pressure; 2) opportunity; and 3) rationalization/attitude.^[10] It is the responsibility of management to design, implement, and maintain control systems to deter, prevent, and detect fraud.^[9,10]

Public demands on elected officials and staff to provide more services with fewer resources may place undue pressure on staff and create situations conducive to fraud. Pressure put on staff by elected officials to report positive financial results may also create situations that encourage fraudulent financial reporting.^[9] Situations that auditors might look at as creating opportunities for fraud include the following:

1. No formal budgeting or planning process;
2. Budget not prepared by appropriate level of management or prepared without input from employees understanding the budgetary aspects of the organization's activities;
3. Uncontrolled, or ineffectively controlled, access to cash or cash equivalents;
4. Appropriate documentation for time worked, overtime, etc. is not required and/or properly documented; and,
5. Department/agency directors have authority to spend outside of budgetary process.^[2]

With respect to rationalization/attitude, auditors may be concerned if the governing body/management does not support public accountability or programmatic efficiencies. Likewise, a governing body/management displaying significant disregard for regulatory, legal, and/or oversight requirements of regulatory agencies would be of concern to the external

independent auditor.^[2] Under SAS No. 99, external independent auditors are required to consider opportunities management may have to override existing control systems as a fraud risk factor.^[10,11]

TRADITIONAL CONTROL SYSTEMS

Historically, the background of many of the chief financial officers of governmental organizations was strongly routed in accounting, auditing, or both. As such, they typically viewed governmental accountability as a function of strict internal controls primarily in the areas of spending, procurement, payroll processing, and hiring practices.^[8] Traditional control systems were principally concerned with the following:

1. Properly qualified personnel in appropriate positions;
2. Proper segregation of duties;
3. Separating operations from accounting (often created through centralization of all or most accounting functions);
4. Assignment of responsibility as appropriate in specific circumstances;
5. Maintaining control over bank accounts, cash receipts, and cash disbursements;
6. Recording transactions accurately and timely; and,
7. Safeguarding assets by limiting access to source records, accounting systems, etc.^[6,8]

Under the traditional concept of control systems, administrators used the budget as the primary financial control mechanism. The budget was utilized to prevent overspending, ensure budgetary balance, and ascertain funds were spent in accordance for approved budget purposes and amounts.^[3] Following the scientific administration theorists, early control systems focused on technical skills and work efforts. During this time, resources were focused on the efficient achievement of goals and objectives albeit using a scientific management orientation.^[12,13]

CONTROL SYSTEMS AS FINANCIAL MANAGEMENT CAPACITY

During the 1980s and 1990s, techniques such as budgeting for outcomes were developed to accommodate the delegation of decision making to lower levels and to increase accountability among public organizations.^[14] In an effort to link resource allocation decisions with performance and outcomes, alternative budgeting techniques and formats such as program, zero-based, target-based, outcome-oriented, and

performance budgeting were adopted by a number of governments.^[15,16] Decentralization of functions and moving from rules to guiding principles occurred in such areas as purchasing, budgeting, and personnel.^[17–19]

In the context of control systems, managing for results entails establishing controls that ensure the sound financial condition of a governmental organization. In other words, effective management is necessary for effective governance. As such, management capacity is not only personal qualities but organizational processes and activities as well. Without effective management, government organizations will never reach their optimum level of overall effectiveness. Management capacity is the crucial link between government resources and public services. Understanding how management capacity relates to results is the first step in improving performance and accountability in government organizations.^[12,20]

Financial management capacity is the fundamental ability of a governmental organization to develop, direct, and control its financial and economic resources. Similar to supply chain management, financial management capacity is having the right resources in the right place at the right time. Drivers of financial management capacity include: 1) characteristics of the financial management system; 2) nature and extent of support from top leadership; and 3) the degree of integration and alignment across other management systems.^[12] As a control system, financial management capacity includes policies and procedures related to procurement, accounting, cash and investment management, debt management, and budgetary execution and control. To be effective, financial management control systems should support managers while allowing them autonomy and holding them responsible for results.^[12,21]

The objective of a government's control system related to financial management capacity is to ensure and maintain sound financial condition without negatively affecting service levels and/or quality. For any government the challenge is identifying those components of an ideal financial management control system that can be effective in their specific circumstances. An effective financial management control system provides information related to how well the governmental organization used its available resources. It also provides quality information on a timely basis that is appropriate for the intended decision maker.^[21]

Applying management capacity theory, financial management control systems that simply measure inputs and outcomes without considering the context of, and capacity for, organizational performance are misleading. Therefore, financial management control systems should treat the traditional functions of analysis and measurement as neutral activities. Conversely, managers should be viewed as decision makers,

strategists, and analysts rather than neutral bureaucrats. This nontraditional focus of financial management control systems on organizational performance necessitates allocating the appropriate level of resources in accordance with strategic concerns. In so doing, public managers will spend the governmental organization's limited resources effectively and accountably.^[12]

Accurate revenue and expenditure forecasting is critical to the effectiveness of this nontraditional model of financial management control systems. Additionally, elected officials and staff must focus on the long term including planning, both financially and operationally, for contingencies. It is also necessary that all involved in this type of nontraditional financial management control system understand the relationship between cost, performance, and resource allocation. As with any control system focused on overall organizational performance, appropriate flexibility must be included in this nontraditional type of financial management control system.^[12]

COMMUNICATION, COORDINATION, AND INTEGRATION FOR FLEXIBILITY AND CONTROL

Various relationship networks exist within and across all levels of today's complex governmental organizations. These internal and external environments significantly impact the classic policy/performance equation (public resources to public management to public policy results) in any governmental organization. Effective implementation of a nontraditional financial management control system requires network-wide cooperation. As such, common objectives should exist that unite the various network components in essential ways and that contribute to the development of integration strategies.^[12]

The number and complexity of financial controls increased in many governmental organizations in response to stakeholder demands for increased accountability.^[3] It may prove difficult to integrate a nontraditional financial management control system in organizations where existing budget allocation and execution systems are mired in control oriented processes and procedures. In such organizations, elected officials and management may need to be reminded of the objectives of public expenditure management. Such objectives include not only fiscal discipline (through expenditure control) but also resource allocation consistent with policy priorities and good operational management.^[4]

Chances for successful integration of financial management control systems can be increased by combining managerial and political needs. This can be done

most effectively by using citizen articulated goals to combine the routine needs of departments with policy goals of elected officials.^[3] There still exists, however, a need to balance this strategic and citizen-driven process with effective financial controls. For example, decentralization typically reduces micromanagement and unnecessary oversight. Therefore, if certain processes are decentralized based on the mutual agreement of the affected networks, there should be an increased level of accountability for performance. Decentralization only works if managers possess the requisite skills related to the decentralized process. If funds are allocated using a strategic focus and based on projected results, managers should be held accountable if the projected results are not achieved.^[3]

Other areas where it may be possible to achieve effective communication, coordination, and integration of nontraditional financial management control systems include the following:

1. Increased budget and contract authority for senior management;
2. Increased spending authority at all levels of the organization as deemed appropriate in specific circumstances;
3. Allowing budget transfers at various administrative levels but within statutory limits;
4. Establishing discretionary spending accounts for use in unforeseen emergencies or in certain other qualifying circumstances; and,
5. Full integration of performance and/or program budget formats including linkage to budget allocations to increase effectiveness.^[3]

CONCLUSIONS

This chapter discussed control systems in governmental organizations in the context of accountability from both a traditional and innovative perspective. It was pointed out that stakeholders want their governments to provide services economically, efficiently, and effectively and that they want to know programs are operating as designed and achieving the desired results. This chapter focused on the idea that managers and elected officials are accountable to all stakeholders for the activities of their organizations and the related results. Key to this chapter was the point that as a measure of accountability, control systems are vital to the governance process and an essential element for a healthy democracy.

In this chapter, it was emphasized that an integrated financial management approach balances control with effectiveness. A number of ways this balance could be accomplished were discussed in this chapter. The chief premise of this chapter was that control systems should

be viewed in the context of management capacity theory. Control systems, as financial management capacity, expand the traditional concept of financial management (i.e., control systems) from a narrow accounting/budget focus to one of strategic issues and program results. As such, governmental effectiveness depends on the quality of financial management information and procedures.^[5]

For some governmental organizations, the quality of financial management information may be improved by a more comprehensive and/or consistent budget process and/or format. Better integration of accounting and budgeting data with operational results can also improve the quality of financial management information for some organizations. For others, financial management information may be improved by simply installing up to date information technology systems. Improvements to increase the quality of financial management information will be in vain unless there is increased accountability for costs and results.^[5]

For purposes of external and internal accountability, control systems are a necessary part of the operations of all governmental organizations. External independent auditors rely on an organization's control systems in assessing the effectiveness of the related controls. Moving from a traditional control oriented system to one focused on financial management capacity is not a short-term effort.^[5] It requires a paradigm shift in thinking at all levels of the organization as well as a high level of trust between citizens, elected officials, and staff.

REFERENCES

1. Comptroller General of the United States. In *Government Auditing Standards: 2003 Revision*. GAO-03-673G; United States Government Accountability Office (GAO): Washington, DC, 2003.
2. American Institute of Certified Public Accountants. In *State and Local Government Audit Guide*; American Institute of Certified Public Accountants: New York, NY, 2004.
3. Bland, R.; Rubin, I. *Budgeting: A Guide for Local Governments*; International City/County Management Association: Washington, DC, 1997.
4. Schiavo-Campo, S.; Tommasi, D. *Managing Government Expenditure*; Asian Development Bank: Manila, Philippines, 1999.
5. Bowsher, C. Governmental financial management at the crossroads: the choice is between reactive and proactive financial management. *Public Budget. Financ.* **1985**, 3, 9-22.
6. Mikesell, J. *Fiscal Administration: Analysis and Applications for the Public Sector*, 6th Ed.; Wadsworth/Thomson Learning: Belmont, CA, 2003.

7. Svava, J. The shifting boundary between elected officials and city managers in large council-manager cities. *Public Admin. Rev.* **1999**, *1*, 44–53.
8. Reed, B.; Swain, J. *Public Finance Administration*, 2nd Ed.; Sage Publications: Thousand Oaks, CA, 1997.
9. Lundelius, C. *Financial Reporting Fraud: A Practical Guide to Detection and Internal Control*; American Institute of Certified Public Accountants: New York, NY, 2003.
10. Statement on Auditing Standards No. 99. In *Consideration of Fraud in a Financial Statement Audit*; American Institute of Certified Public Accountants: New York, NY, 2002.
11. Ramos, M. *Fraud Detection in a GAAS Audit: SAS No. 99 Implementation Guide*; American Institute of Certified Public Accountants: New York, NY, 2003.
12. Ingraham, P.; Joyce, P.; Donahue, A. *Government Performance: Why Management Matters*; Johns Hopkins University Press: Baltimore, MD, 2003.
13. Gullick, L. Management is a science. *Acad. Mgmt. J.* **1965**, *000001*, 7–14.
14. Martin, L.L. Outcome budgeting: a new entrepreneurial approach to budgeting. *J. Public Budget. Account. Financ. Mgmt.* **1997**, *1*, 108–126.
15. Rubin, I.S. Budget reform and political reform: conclusions from six cities. *Public Admin. Rev.* **1992**, *5*, 454–476.
16. Wang, X. Conditions to implement outcome-oriented performance budgeting: some empirical evidence. *J. Public Budget. Account. Financ. Mgmt.* **1999**, *3*, 533–552.
17. Savas, E.S. It's time to privatize. *Gov. Union Rev. Public Policy Digest.* **1993**, *1*, 37–52.
18. Bartle, J.R.; Korosec, R.L. A review of state procurement and contracting. *J. Public Procurement* **2003**, *2*, 192–214.
19. Gianakis, G.A.; Wang, X. Decentralization of the purchasing function in municipal government: a national survey. *J. Public Budget. Account. Financ. Mgmt.* **2000**, *3*, 421–440.
20. Moynihan, D.P.; Ingraham, P.W. Look for the silver lining: when performance-based accountability systems work. *J. Public Admin. Res. Theory* **2003**, *4*, 469–491.
21. Ingraham, P.W.; Donahue, A.K. Dissecting the black box revisited: characterizing government management capacity. In *Governance and Performance: New Perspectives*; Heinrich, C., Lynn, L., Eds.; Georgetown University Press: Washington, DC, 2000.

Cooperative Purchasing

Alex Sekwat

*Institute of Government, Tennessee State University,
Nashville, Tennessee, U.S.A.*

INTRODUCTION

Cooperative purchasing refers to a variety of arrangements in which two or more entities buy goods and services under the same agreement or contract. The National Association of State Purchasing Officials (NASPO) (Ref.^[1], p. 94) defines cooperative purchasing as “buying through public solicitation of competitive bids or competitive proposal by two or more other public jurisdictions, authorities, or agencies.” According to NASPO (Ref.^[1], p. 94), cooperative purchasing “may include the resale or exchange of goods and services without competitive bidding and the shared use of facilities, procurement information, and procurement personnel.” Generally, cooperative purchasing arrangements require that participating entities sign a voluntary agreement or memorandum of understanding specifying, among other things, procedures for participation in contracts, warehousing, fee payments when applicable, contract dispute resolution, terms of agreement, and payment of invoices. Sponsorship of cooperative purchasing programs vary among states. However, all require participating jurisdictions to use best value in the procurement of goods and services.

OVERVIEW

There are many forms of cooperative purchasing agreements. The most common types include the join-bid contract, piggybacking, and joint administrative or consolidated purchasing. The join-bid method is popular among state and local government units. Under this technique, two or more jurisdictions take advantage of the benefits of centralized purchasing by reaching consensus on specifications, contract terms, and conditions for common products. The bidding requirements of the participating jurisdictions are pooled in a single invitation for bids to realize larger volume and better unit pricing. The piggybacking method requires a lead jurisdiction or large purchaser of product(s) to invite bids, enter into a contract, and arrange for other jurisdictions to purchase the same products under the same conditions.^[2] Local jurisdictions often piggyback

purchases on procurement done by state agencies especially of heavy equipment. The joint consolidated or administrative purchasing method is a formal contractual arrangement where several jurisdictions agree to set up an administrative agency responsible for partial or entire purchases for the participating entities. Under this method, each jurisdiction must give up its freedom to comply with terms of the formal contract. The lead agency is in charge of preparation of specifications, solicitation and evaluation of bids, monitoring participation, and administration of the contract. Participants share the administrative costs of the program. Joint purchasing is typically used by public agencies within a specific metropolitan area, agencies within a state, jurisdictions taking advantage of contract prices negotiated by a state, and educational institutions using contract prices negotiated by national educational associations.^[3]

Other variants of cooperative procurement, especially among local jurisdictions, include the joint use of facilities and exchange of technical information and personnel.^[2] Under the former method, two or more entities may agree to jointly use facilities such as a warehouse or a testing laboratory. Such an arrangement is intended to reduce procurement costs. In the event that jurisdictions are not in position to purchase cooperatively, “they at least can help one another by sharing information, loaning their staff, or letting others use their facilities or product specifications” (Ref.^[2], p. 352).

Cooperative purchasing is used by private, public, and non-profit sector entities. In the public sector, federal agencies, state agencies, and the approximately 87,000 local jurisdictions in the United States can engage in cooperative procurement. Non-profit agencies, such as hospitals, colleges, universities, and religious organizations, are also major participants in cooperative purchasing. Although jurisdictions or organizations may have operational differences, they buy basically the same types of goods and services. The most jointly procured goods and services by state and local jurisdictions include vehicles, computer systems, software, office supplies, gasoline, automotive and heavy equipment parts, furniture and office equipment, asphalt, tires, pharmaceuticals, janitorial supplies, training services, travel services, and telephone

services. In general, products or services, which are utilized by a large number of jurisdictions, are awarded by competitive bids.

HISTORICAL BACKGROUND

Cooperative purchasing is not new in the United States. The earliest known established practice in the public sector dates back to about 1930 in Alamosa, CO, when eight school districts formed a purchasing pool. Other early initiatives in cooperative purchasing included the formation of a joint procurement venture between the City of Cincinnati, OH, and the Hamilton County, OH, in 1931 and the purchase of fire hoses by the Michigan Municipal League for eight municipalities in 1938. The Cincinnati–Hamilton County plan was formed on a voluntary basis and managed by a coordinating committee of purchasing agents of Hamilton County composed of the purchasing agents of Hamilton County, the City of Cincinnati, the Cincinnati Board of Education, the Public Library of Cincinnati, and the University of Cincinnati.^[4] By the late 1970s, intergovernmental cooperative purchasing reached its maturity. The publication in 1978 of the Model Procurement Code marked a milestone in the efforts of state and local governments to adopt laws and ordinances that encouraged and facilitated intergovernmental purchasing cooperation. Today, the practice is widespread among states, counties, cities, towns, villages, school districts, special school districts, colleges, universities, agencies, authorities, commissions, and other bodies with the power to award public contracts.

BENEFITS

Cooperative or pool purchasing has several benefits. First, by buying identical or similar goods and services in large volumes, participants save money and time. Volume purchasing provides participants with economies of large-scale purchasing by lowering unit costs of products. Pooling of purchasing functions further lowers administrative costs by substantially reducing duplication of efforts associated with competitive procurement. Second, cooperative purchasing programs save participants time to research specifications, to identify vendors, and to prepare, advertise, and administer a bid proposal or maintain a contract. Third, pool purchasing offers program subscribers standardized specifications, better quality control, and broader selection of goods and services. Finally, cooperative purchasing facilitates information sharing and expertise, especially among smaller jurisdictions that lack resources to hire full-time professional purchasing

managers. Professional management of purchasing functions further benefits smaller jurisdictions in technical procurement functions such as preparation of specifications, training, economic analysis, development of buying strategies, testing and inspection of products, and disposition or exchange of surplus property.^[1]

DRAWBACKS

Despite the benefits highlighted above, cooperative purchasing has some major drawbacks. First, certain legal and political obstacles impede the growth and development of cooperative purchasing programs. The major political obstacles include “preference of local vendors, the fear of loss of autonomy, the difficulty of settling on standardized items to all, and the feeling of the larger participants that their savings will be less than those of the smaller units.”^[3] Experience indicates that over time, political barriers are more difficult to overcome than legal barriers.^[3] Second, sharing of administrative and technical overhead costs, especially under a join-bid arrangement, is problematic. Great effort is needed to assemble the participating jurisdictions, to draft and coordinate terms and specifications, to prepare bidders’ lists and draft solicitations, and to tabulate and evaluate responses.^[4] Last, if poorly designed, a cooperative purchasing program can increase the carrying costs for participating jurisdictions. Volume buying typically increases storage, insurance, coordination, and transportation costs. As noted by Reed and Swain,^[5] “when items are purchased in large amounts, those items will need to be stored since products must be stored for an extended period of time. They must be stored so that they do not deteriorate, which may require special facilities. . . . Clearly, considerable thought needs to be given to the costs of such enterprises.”

CONCLUSION

Cooperative purchasing promotes sound administrative values, including economy and efficiency, by maximizing the time and resources of participating units. It is cost-effective and provides participants with the economies of large-scale purchasing. Thus it renders high value for taxpayers dollars. Well-rounded cooperative purchasing programs further foster intergovernmental cooperation and reduce duplication of work. However, as noted by NASPO (Ref.^[1], p. 96), “successful programs need favorable laws, adequate implementing regulations, energetic leadership, and cooperation among the parties.”

REFERENCES

1. National Association of State Purchasing Officials, NASPO. *State and Local Government Purchasing*; The Council of State Governments: Lexington, KY, 1988; 94–96.
2. Gordon, S. Purchasing. In *Local Government Finance*; Petersen, J.E., Strachota, D.R., Eds.; Government Finance Officers Association: Chicago, IL, 1991; 339–353.
3. Zenz, G.J. *Purchasing and the Management of Materials*; John Wiley & Sons, Inc.: New York, 1994; 364–365.
4. Page, H.R. *Public Purchasing and Materials Management*; Lexington Books: Lexington, MA, 1980.
5. Reed, J.B.; Swain, J.W. *Public Finance Administration*; Prentice Hall: Englewood Cliffs, NJ, 1990; 195 pp.

Correctional Administration, The Challenges of

Betsy A. Matthews

*Department of Correctional and Juvenile Justice Studies, Eastern Kentucky University,
Richmond, Kentucky, U.S.A.*

INTRODUCTION

The American correctional system consists of four major components including jails, prisons, probation, and parole. These components are interdependent, working together to protect society and carry out the criminal sentence. Because corrections is a matter of public policy, its goals and activities are often dependent on which political ideology is dominant at the time.^[1] As such, correctional policy has vacillated between the liberal agenda of rehabilitation and the more conservative approaches of deterrence and incapacitation. For the past 25 years, the latter has won out with “get-tough” policies that have contributed to unmanageable growth in the correctional population. Most of a correctional administrator’s time is devoted to managing this growth and to developing practices that promote the objective and efficient allocation of resources. This article will first review the evolution of correctional policy and discuss the consequences of getting tough on crime. It will then describe four correctional trends that have emerged to manage the ever changing and growing population of offenders.

THE EVOLUTION OF CORRECTIONAL POLICY

In an effort to compete for limited resources and gain public support, correctional policies have adapted over time to reflect the prevailing political ideology. This chameleon-like shift in policy has resulted in two very different approaches to achieving the correctional goal of public safety.

The Rehabilitative Ideal

The Rehabilitative Ideal served as the basis for our criminal justice system from the early 1900s until the 1970s. It was an outgrowth of the positivist school of thought, which assumed that criminal behavior was determined by biological, sociological, and psychological factors over which people have little control. As such, the focus was on identifying and addressing these factors as a way to curb criminal behavior. The

Rehabilitative Ideal involved the Progressives’ plan for individualized treatment, which featured broad discretionary practices, and reflected their abiding faith in the state to act in the best interests of the offender.^[2,3] Offenders were sentenced to an indeterminate period of incarceration (e.g., 2–5 years) with their release based on their demonstrated reform.^[2] Probation was introduced as an alternative to incarceration for those who did not require incarceration, and parole was instituted to facilitate an offender’s reintegration to the community.

Several events throughout the 1960s and 1970s led to the dismantling of the rehabilitative agenda. In 1974, Robert Martinson published an article about what works in offender reform. In the article, Martinson reviewed 231 community- and institutional-based programs and concluded that there was no appreciable evidence that rehabilitation works to reduce the recidivism of offenders.^[4] While Martinson is often credited for undermining rehabilitation, several authors speculate that the lost faith in rehabilitation was more of a reflection of the times than a careful consideration of Martinson’s research.^[2,5] The 1960s were turbulent times with protests against Vietnam, Watergate, the deaths at Kent State, and the emergence of the Civil Rights movement. For conservatives, these times were signs of a disintegrating moral order; for liberals they were signs of a corrupted government.^[2] Both views prompted changes in the criminal justice system.

According to Cullen and Gilbert,^[2] conservatives viewed rehabilitation strategies as signs of a permissive society. They believed that the criminal justice system was coddling offenders and decreasing the costs of crime by releasing offenders early. They called for determinate sentences (i.e., a fixed period of incarceration) aimed at minimizing discretion and tougher sentencing practices designed to restore social order. Liberals, believing that the government was “using the mask of benevolence” to inflict suffering on offenders, agreed that discretion must be limited through the use of determinate sentences. However, they wanted to reduce the length of sentences to minimize the harm caused by the coercive prison environment. In the end, punishment won out,^[3] and the get-tough movement was set into motion.

The Get-Tough Movement

The get-tough movement is characterized by deterrence and incapacitative-based strategies. The concept of deterrence can be traced back to the 1700s and the classical school of thought.^[6,7] The classical school of thought assumes that people are rational beings who calculate the benefits of crime against the potential risks of getting caught and behave according to what will bring them the most pleasure and the least pain. Thus, according to deterrence theory, as the severity, swiftness, and certainty of punishment increases, crime will go down. Tough-on-crime policies such as mandatory sentences, truth in sentencing, and three-strikes legislation are designed to create special deterrence for the individual experiencing the punishment and general deterrence for others who observe the punishment.^[8]

Incapacitation, which emerged in the early 1980s, is designed to prevent crime through the physical constraint of individuals. Advocates of incapacitation make no claims of offender reform; they simply assert that offenders will not be able to commit crimes against society while incarcerated or otherwise restrained.^[9,10] Incapacitative measures are seen by many as the shortest path to public safety and thus, are popular in the American culture that is permeated by the principle of efficiency.^[11] Together, incapacitative and deterrent ideologies have contributed to unmanageable growth in the U.S. correctional population.

THE CONSEQUENCES OF GETTING TOUGH ON CRIME

Unmanageable Growth

The persistence of the get-tough movement over the past 30 years has contributed to an expanded net of social control in the United States. According to the Bureau of Justice Statistics, there were 6.5 million people, or 1 in every 32 adults, under some type of correctional supervision by the year-end of 2000.^[12] Of this correctional population, one-third, or 1,933,503 people, were in prison or jail, an annual increase of 5.6 percent since the year-end of 1990 and the continuation of a trend that began in 1971.

It is often assumed that the growth in the prison population reflects an increase in crime rates, but comparisons of trends have revealed that changes in the crime rates and changes in the prison populations are loosely matched over time.^[13,14] Other explanations offered for the growth in prison population include the introduction of tough legislation such as mandatory sentences, truth in sentencing, and three-strikes legislation,^[15-17] and the increased willingness by judges to imprison convicted felons.^[18] Regardless of

the cause, the increasing prison population has created overwhelming challenges for the American correctional system.

In an effort to keep pace with the growing population, states have added 351 correctional facilities in the past decade.^[19] Still, many state prisons are operating at 15% above their rated capacity,^[19] a situation that creates management nightmares for prison administrators. Prison crowding has been found to contribute to negative psychological and physical reactions among inmates^[20] and to increased rates of violence.^[21] It also limits the number and quality of vocational, educational, and counseling programs that are available to inmates,^[22,23] and thus, impedes the extent to which correctional facilities can achieve the goal of rehabilitation.

Recognizing that it is impossible to build their way out of the current correctional crisis, states have instituted a range of community-based correctional options that are designed to divert low-level offenders from prison. Two of the most popular programs are intensive supervision programs (ISP) and electronic monitoring, which are typically operated by probation and parole agencies. The defining characteristics of ISPs are small caseloads (e.g., 25-40 offenders) and an emphasis on controlling offenders in the community through increased contact, drug testing, and stringent responses to technical violations.^[24] Electronic monitoring involves the electronic surveillance of offenders to assure compliance with home confinement.^[25] Other, less popular, options include community service, house arrest, and day reporting centers. Despite their community placement, all of these correctional options (often referred to as intermediate sanctions or alternatives to incarceration) are tough programs designed to protect the public by providing close surveillance of offenders in the community and quick responses to violations or new criminal behavior. These tough new programs were well received by the public, politicians, and practitioners because they "shared the rhetoric of punishment but offered to accomplish crime control at a reduced cost."^[26] Their continuing popularity has contributed to an alarming growth in probation and parole populations.

By the end of 2000, approximately 4.6 million men and women were on probation or parole, representing an annual increase of 3.6% since 1990.^[12] Unfortunately, state budgets have not kept pace with the growing probation and parole populations. Despite the fact that three-fourths of the correctional population is under probation and parole supervision, only about one-tenth of the correctional budget is allocated to probation and parole agencies.^[27] This translates into high caseloads for officers: The national average for the number of cases per officer is 124 for probation and 67 for parole.^[28] As is the case with the prison

system, this type of unmanageable growth in probation and parole populations has limited the scope of services that can be provided to offenders. Without the type of meaningful treatment and supervision that is capable of promoting long-term behavioral change, significant reductions in recidivism are unlikely and the correctional system appears inadequate in its pursuit of public safety.

Lack of Credibility

Public opinion polls have repeatedly revealed that what the public wants from corrections is to be safe from crime.^[29–32] Thus, in the eyes of citizens and public officials, the value of correctional policies lies in their ability to reduce recidivism and crime rates. Repeated news reports of heinous crimes committed by offenders under correctional supervision have contributed to a lack of confidence in the correctional system. Unfortunately, the results of more objective outcome studies would do little to bolster this confidence.

Recidivism rates generally refer to the proportion of offenders under correctional supervision who are arrested or convicted for a new crime. A national study of adults on probation found that 43% of probationers were rearrested for a felony within three years of their probation sentences.^[33] Recidivism rates for parole are equally disheartening. According to a recent report by the Bureau of Justice Statistics, about 60% of parolees were returned to prison in 1999; 70% of those returned had been arrested or convicted for a new offense.^[34]

Studies of control-oriented, intermediate sanction programs have revealed that they are no better at reducing recidivism than the more traditional probation and parole programs. As compared to control groups of offenders placed in other correctional options, participating offenders had an increase in technical violations (i.e., violations of the rules of probation/parole) but no significant differences in the rates of new arrests.^[35–38] In fact, a meta-analysis of studies on these control-oriented intermediate sanctions revealed that they produced a slight increase in recidivism.^[39] As a result, more offenders in these specialized programs are returned to prison, thus limiting the extent to which any diversionary effect or cost savings can be achieved.

Recidivism rates and studies, such as those cited above, show that a substantial proportion of offenders (the people most directly affected by get-tough strategies) are not deterred from future crime. Crime rates offer a better reflection of whether or not these strategies are providing a general deterrent. The two primary methods of crime measurement are the Uniform Crime Reports (UCR), which reports on crimes reported to the police, and the National Crime Victimization

Survey (NCVS), which estimates crime rates based on interviews with members of a nationally representative sample of households in the United States about victimization they have experienced in the past year. According to the UCR, rates of serious crime in 2000 were 14% lower than in 1996 and 22% lower than in 1991.^[40] Statistics from the NCVS reveal a similar downward trend since 1993.^[41]

There are disparate views as to whether these statistics provide support for get-tough legislation. Langan, for example, sees the rising incarceration rates from 1973 to 1986 and the concomitant decline in crime rates as evidence that prisons were “responsible for sizable reductions in crime.”^[18] Similarly, a report by the National Center for Policy Analysis concludes that the recent reduction in crime can be attributed to an increase in “expected punishment” (i.e., an increase in the likelihood of going to prison and the median sentence length).^[42]

Other research disputes the role of imprisonment and other get-tough strategies in reducing crime. For example, given Langan’s contention regarding the benefits of prison, how would he explain the increase in violent crime in the wake of rising incarceration rates from 1986 to 1993?^[43] Reiss and Roth report that violent crime rates remained stable despite a tripling in the average prison time served per crime.^[44] Similarly, Visher examined several studies on the effect of incapacitation and concluded that despite a doubling of the prison population crime, reductions were limited to 10–30%.^[45] Spokespersons for the Campaign for an Effective Crime Policy summarize the debate in a 1995 report by concluding that “clearly, imprisonment has an incapacitative effect on the individual offender. The key issue, however, is whether increased imprisonment results in an overall decrease in crime and at what cost.”^[46]

A review of recidivism studies and crime rates reveals little support for get-tough policies. Despite this, they persist as the basis of the American correctional system leaving correctional administrators to deal with large populations, limited resources, and a lack of public confidence. The remainder of this article will focus on measures that are being taken to overcome these obstacles.

MOVING TOWARD RATIONAL CORRECTIONAL POLICY

Keeping pace with the growing correctional populations and managing constant change is no easy task. Proposed solutions often are nothing more than stopgap approaches designed to appease a particular audience. Four recent trends, however, represent a sincere effort to find long-term answers and to interject

rationality and accountability into the system. Following are brief discussions on how community justice initiatives, privatization, structured decision making, and the development of research-based intervention programs are improving services and decision making in the correctional systems.

Community Justice

Traditional correctional practices, whether rooted in rehabilitation or deterrence and incapacitation, have focused on things to be done *to* or *for* offenders.^[47] One of the most persistent trends in corrections is to shift from this offender-focused model of justice to a model of community justice that “places the community and victims at the center of justice activities and efforts.”^[47] Toward this end, correctional agencies are redefining the way they do business. Instead of working as a closed system, in isolation from the community, they are seeking community input and developing opportunities for citizen involvement.

A common approach to involving citizens is through the use of community advisory boards. Most states have some form of citizen advisory board although the composition and functions may differ. The role of the 21-member State Advisory Board in New Jersey, for example, is to advise the Supreme Court on matters related to probation. The Community Corrections Advisory Committee in Multnomah County, Oregon, is a statutory requirement of the Community Corrections Act adopted in 1977; it helps develop the county’s community corrections plans and oversee its operations. Advisory boards bring many benefits to corrections agencies including a sounding board and a set of allies for correctional personnel, diverse views and perspectives, access to private and public resources, accountability to the public, and a forum in which public officials and private citizens work together on developing opportunities and strategies for change.^[48]

Another common approach to involving citizens in corrections is through restorative justice initiatives. Restorative justice is a philosophy that views crime as a violation of people and interpersonal relationships.^[49] As such, it seeks to repair these relationships by: 1) providing victims with support, opportunities for input, and restitution; 2) holding offenders accountable for the harm they caused; and 3) empowering communities to develop constructive responses to crime that recognize our mutual responsibilities to one another.^[50] Two of the more popular restorative justice programs include victim offender mediation (VOM) and community sentencing panels. The VOM programs provide an opportunity for the victim of a crime to meet face-to-face with the offender and a volunteer mediator

to discuss how the crime has impacted their lives and to work out an agreement for restitution.^[51] Community sentencing panels involve citizen volunteers in the sentencing of offenders. For example, in Minnesota, volunteers serve on a community intervention team that communicates to offenders the impact of their behavior on the local community, sets special conditions of probation, provides support to offenders, and periodically reviews offender progress.^[52]

Although there is evidence to suggest that these types of programs lead to an increase in both victim satisfaction and citizen awareness of the justice process, there is little research available to ascertain the extent to which community and restorative justice initiatives are effective in controlling crime.^[51] Despite this lack of evidence, and unlike other correctional initiatives, community and restorative justice have gained acceptance from both conservatives and liberals.^[53] This widespread appeal and the pragmatic need to expand resources and support for correctional services is likely to sustain the movement toward a more open and inclusive system of justice.

Privatization

Complaints about ineffective practices, crowding, and rising correctional costs have led policy makers to turn to the private sector for solutions. Although private non-profit organizations have a history of involvement in the provision of diversion and drug treatment programs for offenders in the community, it has only been since the early 1980s that private for-profit corporations have engaged in the full-scale operation of prisons.^[11] There are several objections to the privatization of prisons. Most of the controversy, however, surrounds the issues of the profit motive and the right to punish and use force. It has been argued that the profit motive of private corporations will contribute to a lower quality of services in an attempt to cut corners and to lobbying for excessive use of incarceration because it is good for business.^[54,55] More fundamental is the concern about granting private corporations the power to punish and use force.^[56] Some authors claim that punishment is a government function and that granting this right to private corporations undermines the legitimacy of the state.^[57,58]

Despite these controversies, many states are moving ahead and authorizing private corporations to engage in the full-scale operation of prisons. By the end of 2000, 5.8% of all state inmates were housed in privately operated facilities.^[19] Several studies have reported that private prisons outperformed their public counterparts on several dimensions of quality (e.g., care, safety, security, order) and that they accomplished this at a reduced cost.^[59–61] Camp and Gates report that

these evaluations failed to statistically control for factors (e.g., age and race of inmates, staff-to-inmate ratios) that could influence the research outcomes and that their findings should be reviewed with caution.^[62] Based on a comprehensive review of the available research on privatization, the Bureau of Justice Assistance concludes that: privately operated prisons function at least as well as publicly operated prisons; private prisons can reduce costs in markets that have high benefits for public employees; there is evidence to suggest that private corporations can build prisons faster and cheaper than the public sector; and the presence of private prisons has prompted public facilities to develop cost-saving strategies.^[63] Although they are not overly supportive, these conclusions certainly favor the continued inclusion of the private sector in the operation of prisons.

Structured Decision Making

Shrinking resources and a demand for equitable and appropriate decision making have led to an increased interest in strategies that structure the discretion of judges and correctional personnel.^[64] Sentencing guidelines and case classification systems have emerged to guide decision making in a way that minimizes disparities and reserves scarce prison resources for the most dangerous offenders.

Sentencing guidelines

The introduction of sentencing guidelines in the late 1970s represents one of the most salient efforts to interject accountability into the criminal justice system. As of the end of 1999, 18 states had implemented sentencing guidelines, 4 states were considering proposals for guidelines, and 3 states had appointed commissions to study the approach.^[65] The initial aims of sentencing guidelines were to promote proportionality and reduce disparity in sentencing by prescribing a punishment that is based on the seriousness of the offense and the criminal history of the defendant rather than on extralegal factors like race and sex.^[66,67] Increasingly, however, states (e.g., Minnesota, North Carolina, Utah) have included mechanisms within their sentencing guidelines for linking sentences with correctional resources.^[68]

In 1994, North Carolina introduced sentencing guidelines that were based on projections of future crime and sentencing patterns, and matched sentences to the number of prison beds, probation slots, and other correctional resources. The guidelines called for shorter and community-based sentences for non-violent, first time offenders and longer sentences for violent and career offenders. Preliminary results suggest that the

plan is working with a greater percentage of serious felons serving longer than average prison terms and an increased use of non-prison punishments for less serious felons.^[69] The North Carolina Legislature has demonstrated its commitment to these “capacity-based guidelines” by requiring that a fiscal impact statement accompany any recommended revisions to the guidelines.

Case classification

Whereas sentencing guidelines attempt to control judicial discretion and the inputs to correctional systems, case classification systems attempt to guide the discretion of correctional personnel themselves. Classification systems are designed to categorize offenders based on specific behavioral or psychological dimensions and to facilitate the placement of offenders into appropriate programs or facilities.^[70] A well-designed classification system can minimize bias in correctional decision making and promote more efficient use of resources.

The dominant focus of case classification is on predicting offenders’ risk of recidivism or their risk to institutional security.^[71] The methods of risk prediction have evolved over the years from a more subjective form of clinical assessment based on professional judgment to actuarial risk assessment based on objective and standardized instruments.^[64] These instruments include items that are known to be associated with recidivism or institutional misconduct (e.g., age at first arrest, prior criminal history). Although variations exist, the basic process for actuarial risk assessment is the same: Correctional personnel use information gathered from the offender’s case file and/or a structured interview with the offender to assign a score on each item depending on the degree to which the risk factor is present; the individual item scores are then summed for an overall risk score; and this score is used to categorize offenders as low, medium, or high risk. Depending on the offender’s categorization, s/he is placed in a minimum, medium, or maximum level of community supervision or facility. Risk assessment can assist agencies in managing offenders’ risk through the application of control-oriented measures (e.g., surveillance, curfews, electronic monitoring), and, more importantly, in planning for more effective interventions to reduce the probability of recidivism or institutional misconduct.

Developing Research-Based Interventions

A quick perusal of the recidivism rates previously reported for probation and parole perpetuates the “nothing works” mentality propagated by Martinson. A closer examination of outcome studies, however,

provides a more optimistic perspective on the effectiveness of correctional programming.

Over the past two decades, numerous authors have conducted literature reviews and meta-analyses to examine the effectiveness of various correctional interventions. These reviews directly challenge Martinson's conclusion that "nothing works" in correctional programming: About 60% of the studies reviewed reported a reduction in recidivism with the average reduction being 10%.^[72] Even more promising, programs that met certain principles of effective intervention were shown to reduce recidivism by 30 to 50%.

The programs that have lead to the largest reductions in recidivism possess several common characteristics.^[73-76] The most effective programs were conducted in the community,^[75-78] included multimodal programming,^[72,76,79,80] and involved the family in the offender's treatment.^[79-81] Other more specific characteristics that have been identified have been referred to as "the principles of effective intervention."^[73,74,82] They include:

1. Effective interventions are behavioral in nature. A well-designed behavioral program combines a system of reinforcement with modeling by the treatment provider to teach and motivate offenders to perform prosocial behaviors. Additionally, problem solving and self-instructional training may be used to change the offender's cognitions, attitudes, and values that maintain antisocial behavior.
2. Levels of service should be matched to the risk level of the offender. Intensive services are necessary for a significant reduction of recidivism among high-risk offenders, but when applied to low risk offenders, intensive services produce a minimal or negative effect.
3. Offenders should be matched to services designed to improve their specific criminogenic needs such as antisocial attitudes, substance abuse, family communication, and peer associations. Improvements in these areas will contribute to a reduced likelihood of recidivism.
4. Treatment approaches and service providers are matched to the learning style or personality of the offender. For example, high anxiety offenders do not generally respond well to confrontation,^[83] and offenders with below-average intellectual abilities do not respond to cognitive skills programs as well as offenders with above average or high intellectual abilities.^[84]
5. Services for high-risk offenders should be intensive, occupying 40 to 70% of offenders' time over a 3- to 9-month period.
6. The program is highly structured, and contingencies are enforced in a firm but fair way: Staff

design, maintain, and enforce contingencies; internal controls are established to detect possible antisocial activities; and program activities disrupt the criminal network and prevent negative peers from taking over the program.

7. Staff relate to offenders in interpersonally sensitive and constructive ways and are trained and supervised appropriately.
8. Staff monitor offender change on intermediate targets of treatment.
9. Relapse prevention is employed in the community to monitor and anticipate problem situations, and to train offenders to rehearse alternative behavior.
10. High levels of advocacy and brokerage occur if community services are appropriate.

The research on "what works" in correctional interventions provides a powerful agenda for correctional programming, one that holds agencies accountable for achieving the bottom-line results that the public expects—a reduction in recidivism.

CONCLUSION

Years of politically motivated policies have strained the correctional system beyond its capacity. Prison crowding and large probation and parole caseloads make it nearly impossible to provide offenders with the type of supervision and services that are needed to reduce their likelihood of returning to crime. Yet that is what the public expects from corrections—to be kept safe from offenders. Knowing that their credibility depends on meeting this expectation, correctional agencies are working hard to meet the challenge. Through community justice initiatives and the privatization of prisons, correctional agencies are expanding resources and developing valuable partnerships. Through sentencing guidelines and case classification systems, correctional agencies are better equipped to allocate resources in a fair and equitable way and are able to reserve scarce prison space for the most dangerous offenders. And for the first time ever, through ongoing research, corrections agencies have a blueprint for developing programs capable of reducing recidivism. These proactive initiatives suggest that, despite the political context within which correctional agencies operate, there is hope for a more rational approach to managing state corrections policy.

REFERENCES

1. Freeman, R.M. *Correctional Organization and Management: Public Policy Challenges, Behavior, and Structure*; Butterworth Heinemann: Boston, MA, 1999.

2. Cullen, F.T.; Gilbert, K.E. *Reaffirming Rehabilitation*; Anderson Publishing: Cincinnati, OH, 1982.
3. Rothman, D.J. *Conscience and Convenience: The Asylum and its Alternatives in Progressive America*; Little, Brown and Company: Boston, MA, 1980.
4. Martinson, R. What works? Questions and answers about prison reform. *Public Interest* **1974**, *35*, 23–54 (Spring).
5. Cullen, F.T.; Gendreau, P. The effectiveness of correctional rehabilitation: reconsidering the “nothing works” debate. In *The American Prison: Issues in Research and Policy*; Goodstein, L., MacKenzie, D.L., Eds.; Plenum: New York, 1989; 23–44.
6. Beccaria, C. *On Crimes and Punishments*; Bobbs-Merrill: Indianapolis, IN, 1963 [1764].
7. Geis, G. Pioneers in criminology, VII: Jeremy Bentham (1748–1832). *J. Crim. Law Criminol. Police Sci.* **1955**, *46*, 160.
8. Shichor, D. Following the penological pendulum: the survival of rehabilitation. *Fed. Probat.* **1992**, *52* (2), 19–25.
9. Wilson, J.Q. *Crime and Public Policy*; Institute of Contemporary Studies: San Francisco, CA, 1983.
10. Zimring, F.E.; Hawkins, G. *Incapacitation: Penal Confinement and the Restraint of Crime*; Oxford University Press: New York, 1995.
11. Schichor, D. Penal policies at the threshold of the twenty-first century. *Crim. Justice Rev.* **2000**, *25* (1), 1–30.
12. Bureau of Justice Statistics. National correctional population reaches new high grows by 126,400 during 2000 to total 6.5 million adults. <http://ojp.usdoj.gov/bjs/abstract/ppus00.htm> (accessed November 2001).
13. Clear, T.R. *Harm in American Penology: Offenders, Victims, and their Communities*; State University of New York Press: Albany, NY, 1994.
14. Zimring, F.E.; Hawkins, G. *The Scale of Imprisonment*; University of Chicago Press: Chicago, IL, 1991.
15. Parent, D.; Dunworth, T.; McDonald, D.; Rhodes, W. *Key Legislative Issues in Criminal Justice: Mandatory Sentencing*; Research in Action, National Institute of Justice: Washington, DC, 1997 (January); 1–6.
16. Tonry, M. *Sentencing Reform Impacts*; National Institute of Justice: Washington, DC, 1987; 1–113.
17. Wooldredge, J. Research note: a state-level analysis of sentencing policies and inmate crowding in state prisons. *Crime Delinq.* **1996**, *42* (3), 456–466.
18. Langan, P.A. America’s soaring prison population. *Science* **1991**, *251*, 1568–1573 (March).
19. Bureau of Justice Statistics. Nation’s state prison population falls in second half of 2000 first such decline since 1972. <http://ojp.usdoj.gov/bjs/abstract/p00.htm> (accessed November 2001).
20. Paulus, P.; McCain, G.; Cox, V. The effects of crowding in prisons and jails. In *Reactions to Crime: The Public, Courts, and Prisons*; Farrington, D., Gunn, J., Eds.; John Wiley and Sons Ltd.: New York, 1985; 113–134.
21. Gates, G.; McGuire, W. Prison violence: the contribution of crowding versus other determinants of prison assault rates. *J. Res. Crime Delinq.* **1985**, *22* (1), 41–65.
22. Gottfredson, S. Institutional responses to prison overcrowding. *Rev. Law Soc. Change* **1984**, *12*, 259–273.
23. Joyce, N.M. A view of the future: the effect of policy on prison population growth. *Crime Delinq.* **1992**, *38*, 357–368.
24. Fulton, B.; Stone, S.; Gendreau, P. *Restructuring Intensive Supervision Programs: Applying “What Works.”* American Probation and Parole Association: Lexington, KY, 1994.
25. Bureau of Justice Assistance. *Electronic Monitoring in Intensive Probation and Parole Programs*; U.S. Department of Justice: Washington, DC, 1989.
26. Cullen, F.T.; Wright, J.P.; Applegate, B.K. Control in the community: the limits of reform. In *Choosing Correctional Interventions That Work: Defining the Demand and Evaluating the Supply*; Harland, A.T., Ed.; Sage: Newbury Park, CA, 1996; 69–116.
27. Petersilia, J. A crime control rationale for reinvesting in community corrections. *Prison J.* **1995**, *75* (4), 479–496.
28. Camp, C.G.; Camp, G.M. *The Corrections Yearbook 1999: Adult Corrections*; Criminal Justice Institute, Inc.: Middletown, CT, 2000.
29. Doble, J. *Crime and Punishment: The Public’s View*; Public Agenda Foundation: New York, 1987.
30. Cullen, F.T.; Cullen, J.B.; Wozniak, J.F. Is rehabilitation dead? The myth of the punitive public. *J. Crim. Justice* **1988**, *16*, 303–317.
31. Tilow, N. New public opinion poll cites support for intermediate punishment programs. *Perspectives* **1992**, *16* (1), 44–46.
32. Applegate, B.K.; Cullen, F.T.; Fisher, B.S. Public support for correctional treatment: the continuing appeal of the rehabilitative ideal. *Prison J.* **1997**, *77* (3), 237–258.
33. Langan, P.A.; Cuniff, M.A. *Recidivism of Felons on Probation, 1986–1989*; Bureau of Justice Statistics: Washington, DC, 1992.
34. Bureau of Justice Statistics. *Forty-Two Percent of State Parole Discharges Were Successful*; <http://ojp.usdoj.gov/bjs/abstract/tps00.htm> (accessed November 2001).
35. Jolin, A.; Stipak, B. Drug treatment and electronically monitored home confinement: an evaluation of a community-based sentencing option. *Crime Delinq.* **1992**, *38* (2), 158–170.
36. Petersilia, J.; Turner, S. *Evaluating Intensive Supervision Probation/Parole: Results of a Nationwide Experiment*; Research in Brief, National Institute of Justice: Washington, DC, 1993 (May); 1–11.
37. Smith, L.; Akers, R. A comparison of recidivism of Florida’s community control program: a five-year survival analysis. *J. Res. Crime Delinq.* **1993**, *30* (3), 267–292.
38. Wagner, D. Reducing criminal risk: an evaluation of the high risk offender intensive supervision project. *Perspectives* **1989**, *13* (3), 22–27.
39. Gendreau, P.; Little, T. *A Meta-Analysis of the Effectiveness of Sanctions on Offender Recidivism*; Department of Psychology, University of New Brunswick: Saint John, 1993. Unpublished Manuscript.
40. Federal Bureau of Investigation. *Press Release—2001—Crime in the United States*; 2000. <http://www.fbi>.

- gov/pressrel/pressrel01/cius2000.htm (accessed November 2001).
41. Rennison, C.M. *Criminal Victimization 2000: Changes 1999–2000 with Trends 1993–2000*; National Crime Victimization Survey NCJ-187007, Bureau of Justice Statistics: Washington, DC, 2001; 1–16.
 42. Reynolds, M.O. *Crime and Punishment in America: 1997 Update*; NCPA Policy Report 209; National Center for Policy Analysis: Washington, DC, 1997.
 43. Rand, M.R.; Lynch, J.P.; Cantor, D. *Criminal Victimization, 1973–95*; National Crime Victimization Survey NCJ-163069, Bureau of Justice Statistics: Washington, DC, 1997; 1–8.
 44. Reiss, A.J.; Roth, J.A. *Understanding and Preventing Violence*; National Academy Press: Washington, DC, 1993.
 45. Visher, C. Incapacitation and crime control: does a “lock ‘em up” strategy reduce crime? *Justice Q.* **1987**, 4 (4), 513–543.
 46. Campaign for an Effective Crime Policy. *What Every Policymaker Should Know About Imprisonment and the Crime Rate*; Campaign for an Effective Crime Policy: Washington, DC, 1995; 2 pp.
 47. Barajas, E., Jr. Community justice: an emerging concept and practice. In *Community Justice: Concepts and Strategies*; American Probation and Parole Association, Ed.; American Probation and Parole Association: Lexington, KY, 1998; 11–26.
 48. Fulton, B.A. *Restoring Hope Through Community Partnerships: The Real Deal in Crime Control*; American Probation and Parole Association: Lexington, KY, 1996.
 49. Zehr, H.; Mika, H. *Restorative Justice Signposts (Bookmark)*; Mennonite Central Committee and MCC U.S.: Akron, PA, 1997.
 50. Pranis, K. Promising practices in community justice: restorative justice. In *Community Justice: Concepts and Strategies*; American Probation and Parole Association, Ed.; American Probation and Parole Association: Lexington, KY, 1998; 37–58.
 51. Kurki, L. *Incorporating Restorative and Community Justice into American Sentencing and Corrections*; Sentencing and Corrections, National Institute of Justice: Washington, DC, 1999; Vol. 3, 1–11.
 52. Pranis, K. *Community Involvement in Offender Sentencing*; Minnesota Department of Corrections: St. Paul, MN, 1995. Unpublished paper.
 53. Levant, S.; Cullen, F.T.; Fulton, B.; Wozniak, J.F. Reconsidering restorative justice: the corruption of benevolence revisited. *Crime Delinq.* **1999**, 45 (1), 3–27.
 54. Kinkade, P.T.; Leone, M.C. Issues and answers: prison administrators’ responses to controversies surrounding privatization. *Prison J.* **1992**, 72 (1), 57–76.
 55. Bates, E. Prisons for profit. In *The Dilemmas of Corrections: Contemporary Readings*, 4th Ed.; Haas, K.C., Alpert, G.P., Eds.; Waveland Press, Inc.: Prospect Heights, IL, 1999; 592–604.
 56. Vardalis, J.J.; Becker, F.W. Legislative opinions concerning the private operation of state prisons: the case of Florida. *Crim. Justice Policy Rev.* **2000**, 11 (2), 136–148.
 57. Durham, A.M. *Crisis and Reform: Current Issues in American Punishment*; Little, Brown: Boston, 1994.
 58. Shichor, D.; Sechrest, D.K. Quick fixes in corrections: reconsidering private and public for-profit facilities. *Prison J.* **1995**, 75 (4), 457–478.
 59. Archambeault, W.G.; Deis, D.R. *Cost Effectiveness Comparisons of Private Versus Public Prisons in Louisiana*; Louisiana State University, School of Social Work: Baton Rouge, LA, 1996.
 60. Logan, C.H. Well kept: comparing quality of confinement in private and public prisons. *J. Crim. Law Criminol.* **1992**, 83 (3), 577–613.
 61. Thomas, C. *Comparing the Cost and Performance of Public and Private Prisons in Arizona: An Overview of the Study and its Conclusions*; University of Florida, Center for Studies in Criminology and Law: Gainesville, FL, 1997.
 62. Camp, S.D.; Gates, G.G. Private adult prisons: what do we really know and why don’t we know more? In *Privatization in Criminal Justice: Past, Present, and Future*; Shichor, D., Gilbert, M.J., Eds.; Anderson Publishing: Cincinnati, OH, 2001; 283–298.
 63. Bureau of Justice Assistance. *Emerging Issues on Privatized Prisons*; Bureau of Justice Assistance: Washington, DC, 2001.
 64. Jones, P. Risk prediction in criminal justice. In *Choosing Correctional Options That Work: Defining the Demand and Evaluating the Supply*; Harland, A.T., Ed.; Sage: Thousand Oaks, CA, 1996; 33–69.
 65. Reitz, K.R. The status of sentencing guideline reforms in the U.S. *Overcrowded Times* **1999**, 10 (6), 9–10.
 66. Lubitz, R.L.; Ross, T.W. *Sentencing Guidelines: Reflections on the Future*; Sentencing and Corrections, National Institute of Justice: Washington, DC, 2001; Vol. 10.
 67. Tonry, M. *Sentencing Matters*; Oxford University Press: New York, 1996.
 68. Frase, R.S. State sentencing guidelines: still going strong. *Judicature* **1995**, 78 (4), 173–179.
 69. Wright, R.F. *Managing Prison Growth in North Carolina through Structured Sentencing*; Program Focus Series, National Institute of Justice: Washington, DC, 1998; 1–16.
 70. Van Voorhis, P. An overview of offender classification systems. In *Correctional Counseling and Rehabilitation*, 4th Ed.; Van Voorhis, P., Braswell, M., Lester, D., Eds.; Anderson Publishing: Cincinnati, OH, 2000; 81–110.
 71. Clements, C. Offender classification: two decades of progress. *Crim. Justice Behav.* **1996**, 23 (1), 143–212.
 72. Lipsey, M. Juvenile delinquency treatment: a meta-analytic inquiry into the variability of effects. In *Meta-Analysis for Explanation*; Cook, T., Cooper, H., Cordray, H., Hartman, H., Hedges, L., Light, R., Louis, T., Mosteller, F., Eds.; B. Russell Sage: New York, 1992; 83–127.
 73. Andrews, D.; Zinger, I.; Hoge, R.; Bonta, J.; Gendreau, P.; Cullen, F. Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology* **1990**, 28, 369–404.
 74. Gendreau, P.; Andrews, D. Tertiary prevention: what the meta-analyses of the offender treatment literature

- tell us about “what works.” *Can. J. Criminal.* **1990**, 32 (1), 173–184.
75. Izzo, R.; Ross, R. Meta-analysis of rehabilitation programs for juvenile delinquents: a brief report. *Crim. Justice Behav.* **1990**, 17 (1), 134–142.
 76. Lipsey, M.; Wilson, D. Effective intervention for serious juvenile offenders: a synthesis of research. In *Serious and Violent Juvenile Offenders: Risk Factors and Successful Interventions*; Loeber, R., Farrington, D.P., Eds.; Sage: Thousand Oaks, CA, 1998; 313–345.
 77. Palmer, T. The youth authority’s community treatment project. *Fed. Probat.* **1974**, 38 (1), 3–14.
 78. Whitehead, J.; Lab, S. A meta-analysis of juvenile correctional treatment. *J. Res. Crime Delinq.* **1989**, 26 (3), 276–295.
 79. Clements, C. Delinquency prevention and treatment: a community-centered perspective. *Crim. Justice Behav.* **1988**, 15 (3), 286–305.
 80. Palmer, T. Programmatic and nonprogrammatic aspects of successful intervention. In *Choosing Correctional Options That Work*; Harland, A.T., Ed.; Sage: Thousand Oaks, CA, 1996; 131–182.
 81. Gendreau, P.; Ross, R. Revivication of rehabilitation: evidence from the 1980s. *Justice Q.* **1987**, 4 (3), 349–407.
 82. Gendreau, P. The principles of effective intervention with offenders. In *Choosing Correctional Options That Work*; Harland, A.T., Ed.; Sage: Thousand Oaks, CA, 1996; 117–130.
 83. Warren, M. Application of interpersonal maturity theory to offender populations. In *Personality Theory, Moral Development, and Criminal Behavior*; Laufer, W., Day, J., Eds.; Lexington Books: Lexington, MA, 1983; 23–49.
 84. Fabiano, E.; Porporino, F.; Robinson, D. Canada’s cognitive skills program corrects offenders’ faulty thinking. *Correct. Today* **1991**, 53 (5), 102–108.

Council/Commission Representation

Tari Renner

Department of Political Science, Illinois Wesleyan University, Bloomington, Illinois, U.S.A.

INTRODUCTION

How well do local government legislatures (city councils and county commissions) reflect the interests, opinions, and characteristics of their communities? This article provides an overview of the empirical research on the linkage between elected local officials and constituents. Given their particular importance, the focus is primarily on the impact of varying electoral structures used to choose elected representatives.

The legislative bodies of cities and counties have a central place in the processes of popular representation, democratic governance, and accountability in their communities. Unlike their counterparts at the state and federal level, local legislatures are unicameral and rarely have to contend with a powerful separately elected executive who has veto power over their actions.^[1] Legislative bodies in cities are most commonly referred to as councils, but a small number may use the titles of board, selectmen, or trustees. The most common name of county legislatures is commission, followed by council and board.

OVERVIEW

The procedures for electing representatives to city and county legislative bodies vary widely between the two types of local governments. The overwhelming majority of cities use non-partisan elections. The latest city form of government survey by the International City/County Management Association (ICMA) puts the figure at 76%. Alternatively, the overwhelming majority of counties use partisan ballots (82% in the latest ICMA survey of county forms of government). There is also a distinct difference between the two types of local governments in the distribution of at-large, district, and so-called mixed systems. At-large systems have representatives elected from the entire jurisdiction. In contrast, district systems have legislators elected from different areas of the city or county. The mixed systems have some legislators elected communitywide (at-large) and some elected from individual districts. A clear majority of American cities use at-large elections, whereas few use districts. The latest figures from the ICMA indicate that approximately 60% of cities have at-large elections and less than 20%

have district elections (the remainder report having combination or mixed election systems). In contrast, nearly one-half of counties report having district elections and only one-fifth elect their legislatures at-large. Cities, therefore, are more likely than counties to have so-called reformed electoral systems (promoted by the structural reform movement in the late nineteenth and early twentieth centuries).

The federal Voting Rights Act prohibits the use of racially discriminatory election procedures by both state and local governments. In fact, jurisdictions with a history of discrimination, disproportionately in the South, must have any changes in their election rules and procedures “precleared” by the U.S. Justice Department or the federal district court in Washington, DC, *before* they can go into effect. Although the most overt forms of racial discrimination are largely a thing of the past, election systems can have subtle discriminatory effects that inhibit or “dilute” minority strength. At-large elections, for example, may permit a White majority to consistently elect *all* members of a council over Black minority opponents. When faced with such a case in *City of Mobile v. Bolden*, 446 U.S., 55 (1980), the Supreme Court stated that in order for an electoral system to be overturned under the Voting Rights Act, it must be established that there was some *intent* to discriminate by those who established the structure. The demonstration that an election procedure had discriminatory consequences was not deemed by the court to be sufficient to invalidate the system. In response to this case, the 1982 amendments to Section 2 of the Voting Rights Act make it clear that laws that *result* in diluting minority representation are just as impermissible as those accompanied by discriminatory *intent*. As amended, Section 2 states that an electoral system is invalid if, based on a totality of circumstances, it is apparent that minorities will have fewer opportunities than others to participate in the electoral process and elect representatives of their choosing. The factors to be considered in making this determination include the degree of historical discrimination in the community; the degree of racially polarized block voting; racial appeals in political campaign rhetoric; the proportion of minorities elected to public office (although the Supreme Court in *Thornburgh v. Gingles*, 106 S.Ct. 2752, 1986, made it clear that minorities do not have

a right to have a fixed percentage of elected positions in any legislative body); and the extent to which there is responsiveness to minorities on the part of public officials in a community. These legal changes are largely responsible for the decrease in the use of at-large elections by local governments in the late 1980s.^[2]

The body of academic literature examining the interactive effects of local government election systems on the linkage between minority population percentages and minority representation on elected councils is quite clear. Beyond a certain population threshold (usually 15% or 20%), district and mixed systems tend to facilitate minority representation, whereas at-large systems tend to inhibit minority representation in both cities^[3] and counties.^[4] These systematic patterns persist despite anecdotal evidence that minorities can win in a progressive political environment with at-large elections.^[5] There is also evidence that racial minorities living in communities with district elections have higher levels of political efficacy than those living in communities where councils are elected at-large.^[6]

The extent to which elected councils and commissions are demographically representative of their constituencies (sharing similar distributions of race, ethnicity, gender, income, and education levels) is only one component of the broader concept of representation. Hanna Piktin^[7] referred to this as descriptive representation. In addition, she identified another form—substantive representation—which refers to the extent to which the policies supported by elected representatives are in the interests of their constituencies. Eulau and Karps^[8] referred to a similar type of representation as policy responsiveness. To illustrate the basic distinction between the descriptive and substantive concepts, a Hispanic commissioner who represents a majority Hispanic district might provide good descriptive representation but poor substantive representation if the commissioner voted against policies or programs that were in the best interest of their constituents.

A central question linking the two concepts is “what is the substantive effect of more equitable descriptive representation?” Some academic literature suggests that Black representatives can modestly change local hiring practices and spending priorities, whereas others suggest the change in the Black community is negligible.^[9,10] Research by Zolton Hajnal,^[11] however, focuses on the White community. The project concludes that Black representation in local government tends to decrease racial tension and increase White support of Black elected leadership, although White Republicans were “largely immune” from these effects.^[12] “Once an African American is elected, . . . whites obtain important information about the effect black representation on their lives. They can now base their assessments on an incumbent’s record rather than

on stereotypes, exaggerated fears, or the incendiary predictions of white opponents.”^[13]

Welch and Bledsoe^[6] explored two additional aspects of representation in their analysis of a national survey of city council members. These include representational focus and service representation. The former refers to whom council members or commissioners believe they should represent (a geographic area of the community or neighborhood, a racial or ethnic constituency, a partisan or ideological constituency, the city as a whole, a single issue group, a business or labor constituency, etc.). Service representation, however, refers to how commissioners or council members view their roles in providing constituency service.

Welch and Bledsoe’s data on representational focus indicated that council members who are elected from districts are more likely to report that a particular geographic area or neighborhood is more important to their representation than council members who are elected at-large. The latter are, however, the most likely to report that they focus on the city as a whole or a business constituency. In service representation, the data indicate that at-large council members devote less time to constituency service or ombudsman activities than district members and are less likely to report that their constituents expect personal service. This relationship persisted even after the size of the constituency served was controlled.

In their analysis of policy or substantive representation, Welch and Bledsoe found that there were no statistically significant differences in policy attitudes between council members elected under different structures. However, councils with district elections were engaged in more conflict than those elected at-large. In addition, the nature of the conflict differed because district councils were the most likely to report that their disputes were geographically based. Predictably, the data indicated that councils elected with party labels on the ballot had more partisan conflict than those elected in non-partisan systems. However, it appeared that partisan conflict on the council made both the partisan and non-partisan issues more salient to the public.

In descriptive representation, Welch and Bledsoe’s data indicated at-large council members tend to have higher incomes and education levels than those elected by district. This provides additional evidence to suggest that district elections facilitate council demographic diversity. Non-partisan elections were found to result in more upper-income council members and Republicans. The partisan bias of non-partisan systems, in favor of Republicans, has been documented in earlier academic literature.^[14,15] In Welch and Bledsoe’s data, however, non-partisan elections only advantaged Republicans when they were also elected at-large. District systems, in other words, eliminated the partisan bias of non-partisan elections.

Few, if any, city councils or county commissions can claim to be fully representative of their constituencies. Those in the least affluent stratum of society, likely to be disproportionately Democratic in political identification, are underrepresented in every type of electoral system. However, the literature demonstrates quite clearly that "some kinds of structures block representation of these groups more than do others."^[16] Research on cumulative voting in American local elections by Bowler, Brockington, and Donovan^[17] concluded that all plurality/majoritarian election systems tend to depress voter participation. They found that cumulative voting encourages voter mobilization and turnout.

Since the 1990s, it appears that city councils and county commissions are becoming somewhat more demographically representative of their constituencies. Susan MacManus,^[18] for example, found that from 1988 to 1996, the percent of city council members who were black increased from 4.1% to 5.5%. She also found that the percentage of women increased from 15% to 21% over the same period. Women were found to be most prevalent in the largest cities with racially and ethnically diverse, and more affluent and highly educated, populations. On county commissions, the trend toward gender diversification was even stronger. In the 1988 national survey by the ICMA, only 9% of county legislators were women, compared with 29% in the organization's 1997 questionnaire. Comparable data on the racial change in county legislatures are not available because the ICMA dropped the race question from its most recent survey.

In the early 1990s, a term limits reform movement was successful in adopting their agenda through popular referendum. These reforms appear to be an effective strategy to ensure that the tenure of elected officials is limited and, therefore, that turnover is maximized. The core assumption seems to be that policy-making bodies need constant membership change to remain responsive because over time incumbents become increasingly insulated from their constituent's interests. There is, however, little empirical evidence to support this assumption in local legislatures. Data from ICMA indicate that few city council or county commission members are term limited (less than 10%), yet, there appears to be a great deal of turnover in these positions. Nearly one-half of all city council members were reported to be serving their first term.^[19]

CONCLUSION

The linkage between constituents and their elected local representatives is a complex and multifaceted phenomenon. The existing academic literature has focused primarily on so-called descriptive representation,

the correlation between elected officials' demographic characteristics and those of their respective constituencies. The conditioning or interactive effects of election systems on this relationship have also received considerable scholarly attention. The available longitudinal data indicate that city councils and county commissions have experienced trends toward greater demographic representativeness over the last decade as minorities and women have increased their presence in these bodies. Comparatively little research exists, however, on the degree of substantive representation in local legislatures. There is no systematic evidence, for instance, on the congruence between constituency ideological characteristics and the roll call voting of representatives, which is common in the literature at the national level. The existence of thousands of jurisdictions and councils with various forms of record keeping means that such data would be extremely difficult, if not impossible, to collect. Future research could fill this gap by exploring the other dimensions of representation through elite surveys similar to those used by Welch and Bledsoe.^[6] However, all aspects of local legislative representation need continuing scholarly attention. Comparatively high membership turnover on these councils suggests that all facets of the representation nexus will be dynamic over time.

REFERENCES

1. Renner, T.; DeSantis, V. Municipal forms of government: issues and trends. In *The 1998 Municipal Yearbook*; International City/County Management Association: Washington, 1998.
2. Renner, T. Municipal election processes: the impact upon minority representation. In *The 1998 Municipal Yearbook*; International City/County Management Association: Washington, 1988.
3. Welch, S. The impact of at-large elections on the representation of Blacks and Hispanics. *J. Polit.* **November 1990**, 52, 1050–1076.
4. DeSantis, V.; Renner, T. Minority and gender representation in American county legislatures: the effect of election systems. In *United States Electoral Systems: Their Impact on Women and Minorities*; Rule, W., Zimmerman, J., Eds.; Greenwood Press: New York, 1992.
5. MacManus, S.; Bullock, C., III. Minorities and women do win at-large. *Natl. Civic Rev.* **May/June 1988**, 77, 231–244.
6. Welch, S.; Bledsoe, T. *Urban Reform and Its Consequences: A Study in Representation*; University of Chicago Press: Chicago, 1988.
7. Pitkin, H. *The Concept of Representation*; University of California Press: Berkeley, 1967.
8. Eulau, H.; Karps, P. The puzzle of representation: specifying components of responsiveness. *Legis. Stud. Q.* **August 1977**, 2, 233–254.

9. Mladenka, K. Blacks and Hispanics in urban politics. *Am. Polit. Sci. Rev.* **March 1989**, 83, 165–191.
10. Reed, A. The black urban regime: structural origins and constraints. In *Power, Community and the City*; Smith, M., Ed.; Transaction: New Brunswick, NJ, 1988.
11. Hajnal, Z. White residents, black incumbents, and a declining racial divide. *Am. Polit. Sci. Rev.* **September 2001**, 95, 603–617.
12. Hajnal, Z. White residents, black incumbents, and a declining racial divide. *Am. Polit. Sci. Rev.* **September 2001**, 95, 603.
13. Hajnal, Z. White residents, black incumbents, and a declining racial divide. *Am. Polit. Sci. Rev.* **September 2001**, 95, 604.
14. Hawley, W. *Nonpartisan Elections and the Case for Party Politics*; Wiley: New York, 1973.
15. Lee, E. *The Politics of Nonpartisanship*; University of California Press: Berkeley, 1960.
16. Welch, S.; Bledsoe, T. *Urban Reform and Its Consequences: A Study in Representation*; University of Chicago Press: Chicago, 1988; 53 pp.
17. Bowler, S.; Brockington, D.; Donovan, T. Elections systems and voter turnout: experiments in the United States. *J. Polit.* **August 2001**, 63, 902–915.
18. MacManus, S. The resurgent city councils. In *American State and Local Politics: Directions for the 21st Century*; Weber, R., Brace, P., Eds.; Chatham House: New York, 1999.
19. DeSantis, V.; Renner, T. Term limits and turnover among local officials. In *The 1994 Municipal Yearbook*; International/City County Management Association: Washington, 1994.

BIBLIOGRAPHY

1. Karnig, A.; Walter, O. Decline in municipal voter turnout. *Am. Polit. Q.* **October 1983**, 11, 491–506.
2. Menzel, D., Ed.; *The American County: Frontiers of Knowledge*; The University of Alabama Press: Tuscaloosa, AL, 1996.

County Government

J. Edwin Benton

*Department of Government and International Affairs, University of South Florida,
Tampa, Florida, U.S.A.*

INTRODUCTION

County governments are perhaps the oldest and most familiar units of local government in the United States. Currently, there are 3034 county governments or their equivalents^a in 48 of the 50 states. (Rhode Island and Connecticut abolished county governments in 1842 and 1958, respectively.^b) The number, size, and population of counties vary considerably. While the number of county governments per state ranges from three in Delaware to 254 in Texas, the geographic area of counties can be as small as 46 mi² (Hudson County, New Jersey) or as large as 89,783 mi² (North Slope Borough, Alaska).^[1] The typical county covers an area of 600–700 mi².^[1] In 2002, the average population served by county governments was 83,075, but this ranges from 67 in Loving County, Texas, to over 9.5 million in Los Angeles County, California.^[1] Around 15.6% of all county governments in 2002 ($N = 473$) serve populations of 100,000 or larger, with 201 of these counties in 40 states containing 250,000 or more inhabitants.^[1] But, 71.8% of all American counties ($N = 2178$) serve populations of less than 50,000 persons, with 700 of them containing fewer than 10,000 persons.^[1] Finally, some counties (e.g., Hernando and Orange in Florida) have experienced a tremendous growth in their population (over 100% in Hernando since 1990), while the population in counties in some states, such as Iowa and West Virginia, has remained unchanged or even declined.

To better understand what county governments are and what they do, the remainder of this entry is divided into three parts: 1) a brief history of the development of counties; 2) a description of county government structure; and 3) an overview of the roles counties play as service providers.

HISTORICAL DEVELOPMENT OF COUNTY GOVERNMENTS

County governments were originally established to serve the needs of their state. From the outset, county

governments were viewed as instruments by which their state governments could ensure the delivery of essentially state programs at the local level. The administration of public welfare and health-care programs; provision of roads, bridges, and highways; operation of court systems; maintenance of law and order; supervision of elections; tax assessment and collection; and furnishing agricultural information are examples of state services or services of statewide concern for which counties are responsible. Because of the role that county governments were to play, they frequently have been referred to as political or administrative subdivisions of their state government.

In spite of this general view of the purpose and functioning of county governments, counties developed differently in different parts of the country. To a large extent, the powers that county governments exercise and the functions they perform today—in short, the form of county government that was created—can be traced back to the needs of the settlers of the southern, northern, and middle colonies.^[2–4]

Those states that adopted the model of the Virginia colony (primarily southern states) developed a strong system of county governments. Southern colonies required a form of government that could effectively respond to the needs of a society consisting mainly of plantations, farms, and small agricultural communities. In fact, the wide geographical dispersion of the population in most southern states made the English strong county model a suitable form of local government. Therefore, county government became the primary unit of local government, through which the state guaranteed the coordination and provision of many state services (e.g., welfare, health care, roads, courts, and public safety) at the local level.

By contrast, counties in the northern part of the nation (New England) historically were vested with fewer powers and responsibilities and consequently played a much less important role in the lives of residents of these colonies (later to become states). For the most part, states in this region traditionally relied much more frequently on municipal governments and townships, and even special districts, than on county governments to provide state services and programs to citizens at the local level. This arrangement was a result of the fact that most New England states followed the example set by the Massachusetts Bay

^aIn Alaska, they are referred to as boroughs, while they are called parishes in Louisiana.

^bSince 1997, Massachusetts has abolished nine of its 14 country governments.

Colony, where towns, cities, and townships were established before counties. Therefore, many of the functions and responsibilities that were assigned to counties in the southern colonies were given, instead, to these other units of local government. Because county governments never attained the level of importance in this region of the country as they did elsewhere, Rhode Island and Connecticut eventually abolished their county governments, and more recently, counties appear to be headed for extinction in Massachusetts.

The middle colonies of Pennsylvania, New Jersey, and New York established local governments different from those found in the southern and northern colonies. Cities and towns in New York and New Jersey, like in the New England region, were the primary units of local government. Townships also were established and played an important role. However, county governments were not only established in New York and New Jersey, but they were given substantially more responsibilities. County government in Pennsylvania developed even differently from that in New Jersey and New York. County governments in that state were intentionally established as the primary unit of local government and, hence, developed as important administrative units before cities and towns. Townships were also created in Pennsylvania, to help counties in the provision of services on a level larger than the city or town but smaller than the county.

As the nation expanded westward, the other 37 states that were to eventually join the Union looked to the original 13 states for guidance when creating a system of local government. The model of the southern states, whereby counties became the central focus of local governance as well as the principal agents of service delivery, was adopted by many of the western states. The example set by the states of the Middle Atlantic region was most influential in the scheme of local governance that was created in many states in the Midwest. Here, county governments developed as important political entities, although not to the extent found in southern states. No future state, however, adopted the model set by New England states, where county governments were insignificant units of local government.

STRUCTURE OF COUNTY GOVERNMENT: TRADITIONAL VS. MODERNIZED

To understand the manner in which county governments are structured, one must recognize that counties are legally “creatures of the state.” Emphasizing this fact, the U.S. Supreme Court in the case of *Maryland vs. Baltimore and Ohio Railroad* (1845) held that “the several counties are nothing more than certain portions

of the territory into which the state is divided for a more convenient exercise of the power of government.” Later, in *Merriam vs. Moody’s Executors* (1868), the Iowa Supreme Court ruled that local governments, including counties, could exercise only: 1) those powers expressly delegated to them by the state legislature via enabling legislation; 2) those powers necessary and incident to execution of the express powers; and 3) those powers absolutely necessary to discharge the express powers. These two court cases (the latter known as Dillion’s rule) reinforced the principle that county governments were essentially administrative districts of the state. Therefore, it should come as no surprise that the states historically have dictated the structure of county governments. The form of county government that resulted frequently is referred to as a traditional county commission structure.

Although traditional county commission governments may differ noticeably in their organization, they have three distinctive features in common. First, they have a governing (legislative) body usually referred to as the “county commission” (sometimes, however, it can be called the “county board,” “board of supervisors,” or even “board of judges”), which is comprised of anywhere from three to 50 elected members. Typically, there are three or five county commissioners elected for overlapping 4-yr terms. Second, although the governing board may supervise some functions, it generally shares responsibility for the administration of many county functions with a number of separately elected officials with county-wide jurisdiction (usually referred to as “row officers”), such as the sheriff, county attorney, auditor, recorder, coroner, assessor, judge, treasurer, tax collector, and clerk of the circuit court. For instance, the county commission and the elected sheriff share responsibility for the county jail, with the sheriff overseeing day-to-day operations but the commission deciding on its construction, repair, and financing. The commission usually determines the property tax rate to be imposed on property owners (subject to limitations or tax caps or ceilings in the state constitution or imposed by the state legislature). But, the tax assessor determines the value of each piece of property in the county against which the rate is to be applied. And, in some counties, a separate tax collector actually sends out the tax bills and assumes responsibility for collecting revenue, while another separately elected county official (treasurer) may maintain the county’s financial accounts and write checks. Hence, responsibility for the administration of county government functions is fragmented and dispersed and generally beyond the control of one county official or county office.

A third common feature is the existence of a number of special boards or commissions that have authority over various functions, whose members may be elected

or appointed by the county commissioners or may even include the county commissioners in an *ex officio* capacity. Generally, these include such things as welfare, library, zoning, code enforcement, and pension boards; hospital, housing, and sanitation authorities; and water, sewer, and fire districts.

Reformers historically viewed the traditional structure of county government as lacking in efficiency and political accountability. In addition, reformers have lamented the fact that very little local home rule or self-determination exists for county governments, because they are creatures of the state. This means that the local governing body is limited in its ability to respond to public expectations and the service needs of their constituency. While reformers have advocated that state legislatures grant counties greater discretion in local matters as well as give them the authority to adopt more modern governmental structures, an overwhelming majority of all counties still operate under the traditional county commission structure.

In recent years, however, counties have been granted greater discretionary authority or home rule and the opportunity to modernize their structure of government.^[5] Since the early 1900s, 36 states have relinquished substantial control over the internal affairs of counties through the grant of home rule, while another four states have given additional authority in functional and fiscal areas to their counties.^[6] Moreover, some of the 12 states without home rule have granted authority through special legislation.^[7]

The most common type of home rule is charter government, which is available to counties in 24 states.^[7] Charter home rule permits counties to devise and adopt their own charter and basically results in greater autonomy than other types of home rule, particularly in functional and fiscal areas.^[8] While around 120 of the 1307 counties eligible to adopt a charter have done so, 2920 counties (95%) remain general law counties.^[8] As a general rule, urban counties and counties with more affluent, more educated, and reform-minded citizens are more likely to adopt a charter form of government.

While counties have been slow to take advantage of the opportunity to adopt home rule charters, they have been more willing to move toward greater centralization of executive and policy-making authority. To date, over 800 counties have adopted the commission-administrator form of government, and over 400 counties have opted for an elected executive form of government.^[8,9] Moreover, there has been a steady movement by counties to centralize their government by abolishing or merging the elective positions of constitutional officers under the governing board. For instance, a survey of state county associations in 1992 revealed that 19 of 41 responding states reported some kind of change in the status of elected constitutional

officers.^[10] Many states even permit this type of structural change, regardless of home rule or charters.^[5]

These changes have coincided with the willingness of the states to permit counties greater latitude in raising revenue. Specifically, an increasing number of states have granted counties wider authority to implement sales and gas taxes and to levy a variety of fees and charges for services. At the same time, states have substantially increased the amount of state aid to counties and have demonstrated a greater willingness to share a larger proportion of their revenue receipts with counties.^[5] To a large degree, it seems that financial pressures resulting from reductions in federal aid have been responsible for the willingness of states, in recent years, to increase state aid and shared revenues and grant expanded revenue-raising authority to counties.^[5]

COUNTY SERVICES

County governments, regardless of their structure of government, provide three basic types of services—traditional, municipal type, and regional. Traditional services (many of which are mandated) are typically some of the oldest county services and result from the fact that counties are administrative subdivisions of the state and are provided to all county residents. As such, counties are responsible for services that include public welfare, health and hospitals, highways, police, judicial and legal matters, corrections, elections, tax assessment and collection, and public buildings. Municipal-type services (sometimes referred to as “newer” county services) are usually provided to residents of the unincorporated area who expect the county to assure the delivery of a number of services that historically have been provided only by municipalities. These include fire protection, utilities (water, sewer, and garbage collection), libraries, and protective inspections. Like traditional services, regional services are provided to the entire county (that is, to both incorporated and unincorporated area residents). These are services (also frequently referred to as “newer” county services) that deal with service issues that know no political boundaries (e.g., airports, housing and urban development, sewer and solid waste disposal, parks and recreation, natural resource conservation, and transit) and, hence, require a county-wide response.

To gain a better understanding as to the role that county governments play as providers of these three types of services, it is instructive to examine county financial effort. This can be done best by comparing county expenditures over time and, where appropriate, with other units of local governments

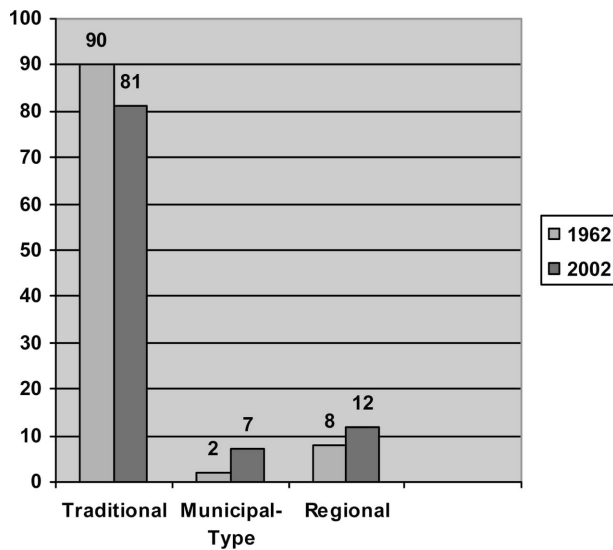


Fig. 1 County service proportions, 1962 and 2002. (From U.S. Bureau of the Census. 1962 Census of Government. *Compendium of Government Finances*; Government Printing Office: Washington, DC, 1964; Table 48; U.S. Bureau of the Census. 2002 Census of Governments. *Compendium of Government Finances*; Government Printing Office: Washington, DC, 2005; Table 48.)

(i.e., municipalities, townships, and special districts). The most reliable expenditure data exist for the 1962–2002 time period.

Because county governments were created to serve as administrative arms of the state, it is not surprising that counties historically have devoted the largest proportion of their fiscal resources to the provision of traditional services. As Fig. 1 indicates, counties in 1962 allocated \$9 out of every \$10 for the provision

of traditional services. As of 2002, counties still spend over four times as much to deliver traditional services as they do for municipal-type and regional services. In short, counties continue to be heavily involved in the provision of state services and programs at the local level. Nevertheless, it appears that this role diminished somewhat between 1962 and 2002, while the role as providers of municipal-type and regional services increased.

A comparison of the service role of counties to other local governments provides another perspective. For instance, it can be seen that county governments are the dominant providers of traditional services. In both 1962 and 2002, counties accounted for the largest proportion of all local government funds allocated to the provision of traditional services (see Fig. 2). In fact, the county proportion increased from 47 to 51 between 1962 and 2002.

With regard to municipal-type services, counties tend to play a relatively minor role in their delivery when compared to municipalities and special districts (see Fig. 3). Here again, this finding comes as no surprise, because most municipal-type services (e.g., fire, utilities, protective inspections, and the like) are ones that municipalities typically have provided to their residents. Nonetheless, two important observations should be noted. First, the county government proportion increased from 3% to 8% between 1962 and 2002. Second, the special district proportion increased substantially over this same time period (that is, from 14% to 34%). This is significant, because special districts are frequently created to provide services to some or all county residents when counties are unable or unwilling to provide them. Therefore, it is reasonable to conclude that, to a certain extent, special districts

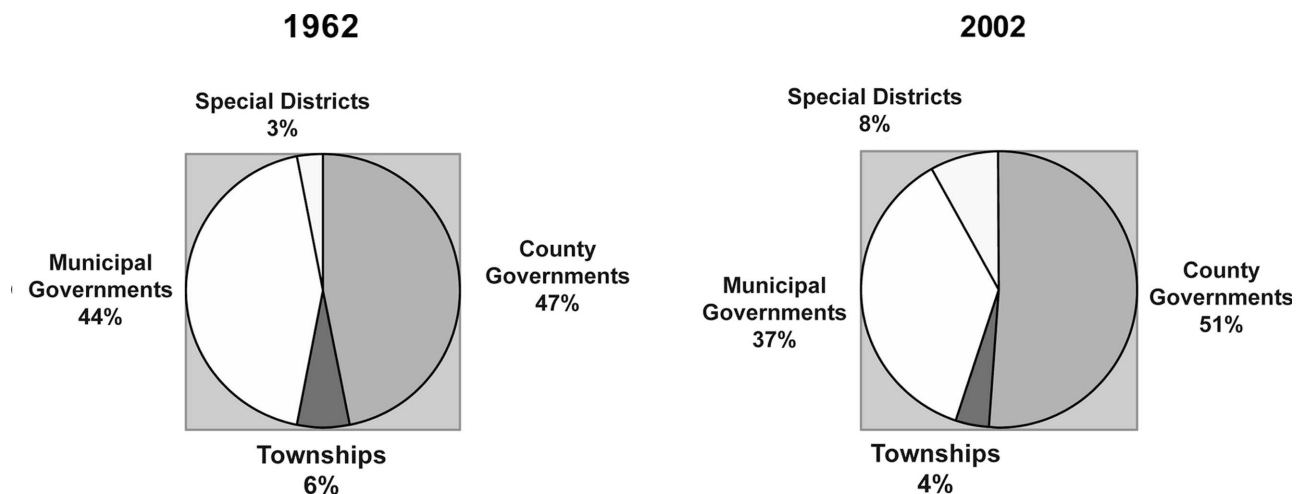


Fig. 2 Traditional service proportions, 1962 and 2002. (From U.S. Bureau of the Census. 1962 Census of Government. *Compendium of Government Finances*; Government Printing Office: Washington, DC, 1964; Table 48; U.S. Bureau of the Census. 2002 Census of Governments. *Compendium of Government Finances*; Government Printing Office: Washington, DC, 2005; Table 48.)

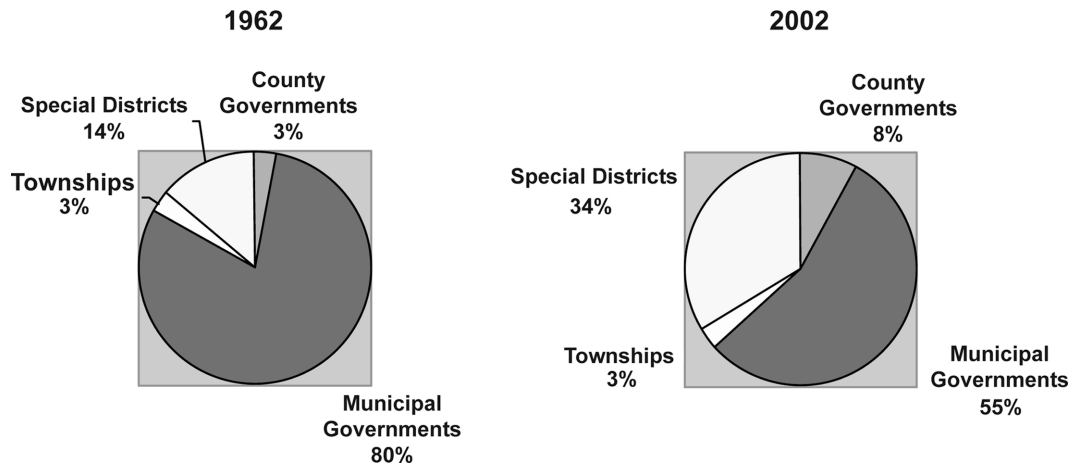


Fig. 3 Municipal-type service proportions, 1962 and 2002. (From U.S. Bureau of the Census. 1962 Census of Government. *Compendium of Government Finances*; Government Printing Office: Washington, DC, 1964; Table 48 U.S. Bureau of the Census. 2002 Census of Governments. *Compendium of Government Finances*; Government Printing Office: Washington, DC, 2005; Table 48.)

assist counties in their role as providers of municipal-type services.

The greatest surprise comes from the role that county governments play in the provision of regional services. As the data in Fig. 4 indicate, the county government proportion of all local government spending for regional services increased noticeably from 11% in 1962 to 18% in 2002, an increase of 7%. Over the same period of time, the municipal government share decreased by 8%, while the township and special district proportion basically remained unchanged. In short, it appears that in recent years, county governments have been called upon to play a much more important role in the delivery of regional-type services, at the same time that municipalities are finding it more

difficult or are even unwilling to provide some of these services, than they have in the past. Examples would be in the area of airports, sewer and solid waste disposal, and housing.

The role that county governments play in the provision of traditional, municipal-type, and regional services, however, varies considerably across counties. On the one hand, the service role of counties vis-à-vis municipalities, townships, and special districts is greatest among counties in the south and west regions of the country,^[2] in rapid-growth counties (particularly if most of the growth is occurring in the unincorporated part of the county),^[2,11] in metropolitan counties,^[2,11] and in counties that have adopted a charter form of government and have an appointed administrator or

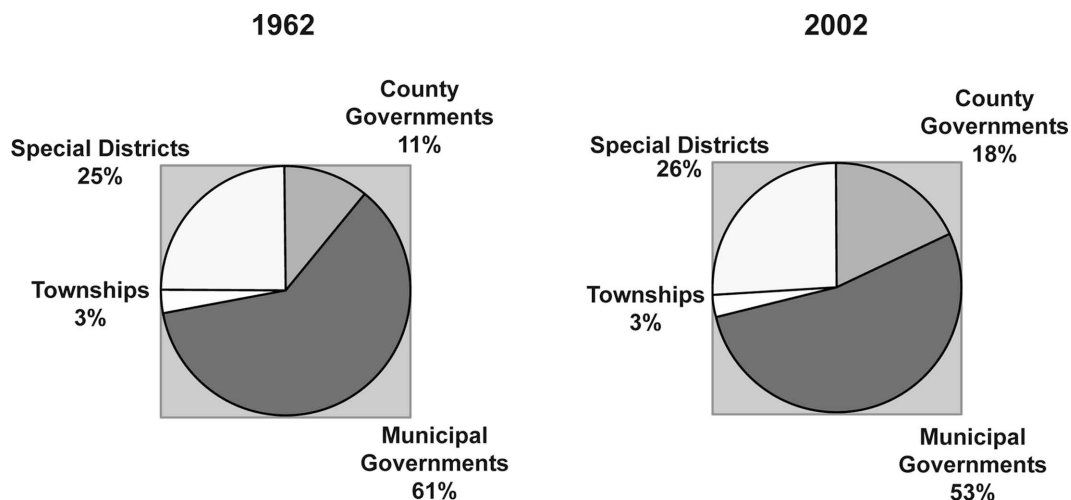


Fig. 4 Regional service proportions, 1962 and 2002. (From U.S. Bureau of the Census. 1962 Census of Government. *Compendium of Government Finances*; Government Printing Office: Washington, DC, 1964; Table 48 U.S. Bureau of the Census. 2002 Census of Governments. *Compendium of Government Finances*; Government Printing Office: Washington, DC, 2005; Table 48.)

elected executive (that is, a reformed structure).^[2,12,13] Moreover, counties tend to play a more important role in the provision of municipal-type and regional services, where their states have assigned more functional responsibility or provided more fiscal assistance.^[2] On the other hand, the service role of counties is much less important in the New England and Middle Atlantic region,^[2] in non-metropolitan counties,^[2,11] in negative- and no- or slow-growth counties,^[2,11] and in states with the traditional commission (i.e., unreformed) form of government.^[2,12,13] In addition, a much smaller county role in the delivery of traditional and regional services is the norm for counties, in which the state assigns greater functional responsibility to other local governments or distributes smaller amounts of state financial aid.^[2] In short, the role of counties as providers of the three types of services can vary significantly from region to region, from state to state, and even from county to county in the same state.

CONCLUSION

In addition to being the oldest form of local government in the United States, counties are perhaps the least understood. This is probably the result of the multiple expectations that we have for these units of government and the different service roles they are required to play. Although originally created to serve as administrative subdivisions of the state, they are frequently expected to provide municipal-type services to the vast numbers of unincorporated areas' residents and provide an increasing menu of services with regional implications to all county residents. In fact, county governments probably touch the lives of individuals more frequently than any other unit of government in the United States. Sometimes, counties are seen as assisting the state in the delivery of state programs and services, and at times, they are seen as trying to perform like municipalities. And, at still other times, they appear as direct providers or as coordinators of the activities of a number of local governments, in efforts to deal with issues that need an area- or

county-wide solution. In spite of the constantly changing image of counties, they have always been and are destined to continue as the cornerstone of local government in most states.

REFERENCES

1. U.S. Bureau of the Census. In *Government Organization, 2002 Census of Governments*; U.S. Government Printing Office: Washington, DC, 2005; Vol. 1, No. 1, vi.
2. Benton, J.E. *Counties as Service Delivery Agents: Changing Roles and Expectations*; Praeger: New York, 2002.
3. Duncombe, H.S. *Modern County Government*; National Association of Counties: Washington, DC, 1977.
4. Martin, LL. American county government: A historical perspective. In *County Governments in an Era of Change*; Berman, D.R., Ed.; Greenwood Press: New York, 1993; 1–13.
5. Berman, D.R.; Salant, T.J. The changing role of counties in the intergovernmental system. In *The American County: Frontiers of Knowledge*; Menzel, D.C., Ed.; University of Alabama Press: University, AL, 1996; 19–33.
6. Salant, T.J. *County Home Rule: Perspectives for Decision Making*; University of Arizona: Tucson, AZ, 1988.
7. Salant, T.J. County governments: an overview. *Intergovern. Perspect.* **1991**, 17 (1), 5–9.
8. Blake, J.; Salant, T.J.; Boroshok, A.L. *County Government Structure: A State By State Report*; National Association of Counties: Washington, DC, 1989.
9. International City/County Management Association. In *The Municipal Year Book 2005*; International City/County Management Association: Washington, DC, 2005.
10. Salant, T.J. *Survey of State County Associations Officials*; 1992; unpublished manuscript.
11. Benton, J.E. County government service provision: The impact of population size and population growth. *Urban Aff. Rev.* **2008**, 44, forthcoming.
12. Benton, J.E. The impact of structural reform on county government service provision. *Soc. Sci. Q.* **2003**, 84 (4), 858–874.
13. Benton, J.E. County service delivery: does structure matter? *Public Admin. Rev.* **2002**, 62 (4), 423–431.

Court Administrators as Managers and Executives

Steven W. Hays

Department of Government and International Studies, University of South Carolina, Columbia, South Carolina, U.S.A.

INTRODUCTION

Professional court administrators emerged relatively recently within the field of Public Administration. Court systems, which had always been managed primarily by judges and elected court clerks, had been slow to modernize until the introduction of these specially trained managers. Their roles are diverse, but their essential responsibility is to improve the operation of courts by freeing judges from non-judicial duties and providing them with competent managerial infrastructures. Known by a variety of titles—such as court executive assistant, court executive officer, and judicial executive—court administrators are now common fixtures in many sizeable judicial systems in federal, state, and local government. However, they are not a universally accepted innovation in some locales. Serving at the pleasure of the presiding judge of their court systems, their duties also vary as dramatically as one might imagine within such a diverse mosaic as the American justice system.^[1]

A BRIEF HISTORY OF THE PROFESSION

The development of the court administrator profession has been sporadic. Its origins can be traced to several catalysts. Most notable were the court reform movement promoted by Dean Roscoe Pound of the Harvard Law School, “good government” efforts that fostered professionalism among city managers and other civil servants, and the “crisis in the courts” that occurred during the 1960s and 1970s.^[2] Court delay, backlog, and notoriously inefficient operating practices led such groups as the American Bar Association (ABA), the American Judicature Society (AJS), and various collections of judges (including the Federal Judicial Council led by Chief Justice Warren Burger) to demand improvements in the administrative capacity of courts. Spurred by federal funding—most notably the Law Enforcement Assistance Program (LEAP)—professional court managers were hired at a feverish pace during the 1960s and 1970s. A complementary trend in this regard was the creation of specialized training programs designed to prepare individuals for the unique challenges of managing judicial

bureaucracies. The Institute for Court Management in Denver, Colorado, was the first program focused exclusively on honing the skills of court managers. This joint effort of the ABA and AJS was soon joined by several academic degree programs, most notably, those at the University of Southern California and at American University in Washington, D.C. By the turn of the twenty-first century, the court administrator profession had achieved a respectable degree of status inside and outside of courthouses across the nation. Every state now has a centralized Court Administrator’s Office (usually as an appendage of the Supreme Court), and most urban trial courts and judicial circuits employ at least one professional court administrator and an appropriate support staff.

A MANAGEMENT PROFILE

Not only are judicial systems extremely diverse, but they are also some of the most tradition-bound institutions of government. These facts are reflected in the heterogeneity that prevails among workers who carry the title of court administrator. Most surveys reveal that court managers at the federal and state levels are attorneys, whereas lower courts more commonly employ individuals with more eclectic backgrounds. There is a strong tendency for judges to appoint people to these positions who have prior experience in quasi-legal organizations such as parole offices, law enforcement agencies, and court clerks’ offices.^[3] Moreover, many of today’s “court administrators” are court clerks; the traditional role of court clerk, which usually involved housekeeping functions and jury management, has been expanded to include a wider array of responsibilities. Although many of these clerks are elected and therefore not known for their preentry professional qualifications, considerable numbers now enroll in training institutes and other programs intended to refine their managerial skills. About one quarter of all court administrators are actually the jurisdictions’ court clerks (either elected or appointed), while most of the remaining group come from diverse occupational backgrounds. Notably, more than 70% of all court managers possess a college degree or better, suggesting that the field is becoming more professional

over time.^[1] However, there are still many court administrators (about 20%) whose formal education ended with a high school diploma. The ABA strongly endorses an elevation in the educational and professional standards for court administrators,^[4] and progress is clearly being made (albeit slowly in many locations).

Organizational Considerations

Court administrators' professional lives are significantly affected by the judicial context in which they work. Courts are hierarchical organizations, just like any other public institution. This hierarchy will usually be mirrored in the court administrators' duties. Those employed by the highest court in a state (usually called the Supreme Court) typically serve at the pleasure of the Chief Justice. As "at will" employees, their job descriptions and job security will fluctuate depending upon the whims of their boss. Some are therefore integrated into high-level policy-making decisions concerning such topics as judicial assignments and budget priorities, while others may be relegated to relatively insignificant tasks such as data collection. If the Supreme Court exercises administrative responsibility over the lower courts in the state (a typical arrangement), then the state-level court administrator may have some supervisory authority over counterparts in lower courts (circuit, county, or specialized bodies such as Family Courts). Stated more directly, centrally mandated procedures and other requirements might be monitored and enforced through the state's highest court administrator. That individual usually has a support staff of professional and quasi-professional employees, including clerks, statisticians, information management experts, budgeting specialists, and the like.

As one descends the judicial hierarchy, the court administrator's role continues to be controlled (in part or completely) by the chief administrative judge of the court in which the administrator is employed. The simple fact is that court administrators always have "a boss," and that boss is a judge or a group of judges acting collectively. In the latter case, a judicial council may take responsibility for management of the judicial system, and thereby appoints the court administrator much like city councils appoint city managers. And, logically, the size and jurisdiction of the court system will directly influence the administrator's job responsibilities. Trial courts will often make jury management a central task of the court administrator, while housekeeping functions (courthouse maintenance and security) will be part of the job only when those obligations are not already assigned to someone else (e.g., an elected court clerk, a sheriff, or other local official). In sum, it is almost impossible to

generalize about the specific duties and responsibilities of court administrators. There is no accepted template or approved job description; each position differs, and each is subject to dramatic change whenever there is turnover among the judges directly responsible for court management assignments.

JOB DUTIES AND RESPONSIBILITIES

As noted, the essential function of court administrators is to relieve judges of as many of their administrative burdens as possible so that they can focus on their chief obligation: hearing and deciding cases. Another implicit expectation is that the presence of a full-time manager within the court will increase the likelihood that administrative problems will be confronted head-on, and that modernization and innovation are more likely to occur. Judges are not generally regarded as being managerial gurus, so the introduction of professional managers is expected by many experts to be an essential step in improving court operations.

Consistent with these expectations, surveys of court administrators indicate that most of their efforts are focused on providing judges with a support infrastructure that handles most of the administrative minutiae, and upgrading management practices within the courthouse. The most common duties of court administrators are in the areas of personnel and financial management. They hire, fire, and supervise most of the non-judicial staff (court reporters, deputy clerks) and are responsible for preparing and monitoring the judicial budget. While these tasks may not appear to be challenging at first glance, the reality is much different. Summoning, selecting, feeding, and paying jurors can be a huge undertaking, even in a fairly small jurisdiction. Record keeping, meanwhile, is a highly specialized and labor-intensive responsibility that requires constant attention. The number of subordinate employees necessary to manage a typical courthouse usually exceeds 75, all of whom must be hired, trained, supervised, paid, evaluated, and disciplined. Riding herd on such a labor force is a major undertaking, and one that judges (and sometimes court clerks) are more than happy to offload.

Categories of Duties

David Saari and his colleagues^[1] organize the duties of court administrators around three management functions: "Operations," "Fiscal Services," and "Special Support Services." Operations include case management, calendaring, courtroom support, research, and program development. Of these, research and program development are activities that rarely, if ever, occurred

before the arrival of court administrators. Overworked judges simply did not have the time to devote to research and planning functions, so professional managers have clearly been a major asset in this regard. Under the fiscal services category, the court administrators not only assume most of the tedious burdens associated with drafting and monitoring budgets, but they also might bring new skills to the table. In particular, professionally trained managers are better able to generate new revenue sources through grant writing efforts, and they can also provide audit and accounting services that were beyond the capacity of most judges.

The "special support" category has perhaps been the area in which court managers have made their biggest contribution to judicial administration. In addition to functions that have already been mentioned, such as jury management, court administrators increasingly offer assistance in emerging "problem areas" within courts. One prime example is in the provision of language interpreters during trials. The recent influx of immigrants spurred by the collapse of the Soviet Union and related international developments has led to an incredible upsurge in trials involving non-English speakers. Sacramento County's (CA) court system, as one simple example, must provide interpreters for 91 different nationalities (including Russian, Chinese, Vietnamese, Greek, and Bulgarian). Other more recent areas of responsibility include courtroom security (this was almost a non-issue prior to the 1970s, when some judges were physically attacked by defendants) and pretrial services (such as intervention programs that permit first-time defendants to expunge their records through community service).

Of special note is the contributions that court administrators are making in the areas of information management and the introduction of new technologies. Prior to the modernization movement, courts were notorious for retaining handwritten docket books and otherwise ignoring the technological revolution that was occurring as a result of the computer age. Computerizing legal records, jury management functions, payroll, judicial calendars, and other courthouse operations has been a major preoccupation of many first- and second-generation court administrators. By upgrading the internal capacity of courts to manage their caseloads efficiently, some commentators have credited court administrators with much of the credit for reducing court delay and eliminating backlogs that once prevailed throughout the United States.^[5]

JUDGES AS THE CRITICAL FACTOR IN JOB EFFECTIVENESS

For obvious reasons, much of the court management literature focuses on the relationship between judges and their court administrators. Ultimately, the courthouse is a highly constrained environment for a professional manager. Judges are unquestionably the key figures within courts, and their level of receptivity to court administrators spells success or failure. Unless they respect the role that a professional manager might play in enhancing judicial operations, the court administrator's job can be terribly lonely and frustrating. Conversely, visionary and progressive judges can fundamentally improve their organizations' functioning by making maximum use of the managerial skills that many of today's professionals possess. But, because many judges neither understand nor trust the concept of court administrators, much additional groundwork needs to be laid before the profession achieves its true potential. A model that is espoused by many proponents is that judges and administrators govern the courthouse as a collaborative "team," yet this approach has seldom (if ever) been realized. For this reason, most court administrators continue to function as part of the judicial system's support infrastructure, but few exercise the discretion or policy-making influence associated with comparable positions in Public Administration (e.g., city managers or agency directors). This reality does not diminish the contributions that court administrators have already made, and continue to make, in modernizing judicial operations.

REFERENCES

1. Saari, D.; Planet, M.; Reinkensmeyer, M. The modern court managers: who they are and what they do in the United States. In *Handbook of Court Administration and Management*, 1st Ed.; Hays, S., Graham, C., Eds.; Marcel Dekker, Inc.: New York, 1993; 237-261.
2. Frank, J. *Courts on Trial: Myth and Reality in American Justice*; Atheneum: New York, 1971.
3. Berkson, L.; Hays, S. Injecting court administrators into an old system. *Justice Syst. J.* **1976**, 2 (1), 55-76.
4. American Bar Association. *Standards Relating to Court Organization*; ABA: Chicago, 1998.
5. Solomon, M. Fundamental issues in caseflow management. In *Handbook of Court Administration and Management*, 1st Ed.; Hays, S., Graham, C., Eds.; Marcel Dekker, Inc.: New York, 1993; 369-381.

Court Budgeting: Judicial Branch Independence and Accountability

John K. Hudzik

*Global Engagement and Strategic Projects, Michigan State University,
East Lansing, Michigan, U.S.A.*

INTRODUCTION

Courts are large, complex organizations.^a With growth in workloads and programs, judicial branch costs have increased steadily (e.g., total state and local judicial and legal expenditures increased almost fourfold between 1984 and 2000).^[1] The emergence of performance standards [e.g., the American Bar Association (ABA) trial court standards] and legislative performance expectations lead to increased scrutiny of the resource planning and management practices of courts.^[2] Pressure throughout the public sector to cut costs requires greater attention to efficiency in the judicial branch through improved practices. Perceived declines in public confidence and satisfaction with the courts require that more attention is given to customer services in resource allocation decisions.

The growth in trial court costs places stress on local government's ability to pay. The movement toward state financing of trial courts and away from traditional forms of local financing is one consequence. Generally, larger judicial branch budgets, especially when aggregated into a single appropriation under state financing, draw greater political and public attention to how much the judicial branch costs and hence to how well it is managed. Critiques of judicial branch organization and performance have resulted in legislated reforms in the interest of greater efficiency and effectiveness. In several states, the adoption of state financing was tied to court reorganization and consolidation, performance standards, and greater controls over judicial branch allocation and expenditure decisions. Even where local financing remained the norm, increased costs brought increased supervision of expenditures by local commissioners and executives. In many jurisdictions, courts are treated as another agency of the executive branch, in direct competition with them for scarce resources, and subject to the same kinds of budget planning requirements

and controls. These changes are particularly salient for U.S. trial courts.

BUDGETS AND JUDICIAL BRANCH INDEPENDENCE

The separation of powers among coequal and independent branches of government has its blessings and challenges. Unproductive budget relationships between the judicial branch and the legislative and executive branches not only lead to insufficient funding for the courts but also attempts at micro-management by the legislative and executive branches on how courts and the judicial branch allocate and reallocate their funds. As the programs and costs of courts expand, their budgets and practices become more noticeable to those on the outside. Poorly managed courts and those which are inattentive to performance expectations invite attention from the outside and increase the likelihood of intervention. The power of the purse, lodged principally in the legislative branch, and the power of administration, lodged mainly in the executive branch, can be used to restrict the judicial branch's allocation of funds necessary to meet its objectives of impartiality, fairness, equity, and due process. Preserving judicial branch independence requires sufficient funding and freedom to allocate funds, but these two conditions are dependent on the three branches collaborating and being jointly responsible. Responsibility requires attention, on the one hand, to good stewardship of funds and attention to performance by the judicial branch; and, on the other hand, it requires good faith practice on the part of the other branches in support of judicial branch prerogatives to determine how best it can meet its constitutional responsibilities.

HISTORICAL CONTEXT FOR JUDICIAL BRANCH BUDGETING AND FINANCIAL MANAGEMENT

Court reform, court unification, and unified court budgeting provide a context for understanding

^aPortions of this article explore similar issues, although from a somewhat different perspective, to those in an article by the author in the 2002 ABA's volume on the administration of justice.

contemporary practice and debate over judicial branch budgeting and financial management. These terms originated in the court reform movement of the first half of the twentieth century; however, their impact has affected the reform efforts since the 1980s as well. Early reforms mainly sought to reduce variance in the number and types of trial courts and in local rules and procedures, and also sought to impose minimum standards of conduct and practice. The principal outcome was to create “one court of justice” in each state.

Besides unification and simplification, the three other components of early reforms included centralized rule making, centralized management, and centralized budgeting.^[3] The Supreme Court imposed rules to be adhered to statewide, and offices of state court administrators were established to increase centralized monitoring and control. In theory, centralized budgeting offered the ultimate means for unification because state-level preparation of the judicial branch budget would permit control over resource allocations at all levels and greater accountability to statewide standards.^[4]

Centralized budgeting requires state financing to be meaningful, and many see unification as being incomplete without state financing. Although all states moved to establish one court of justice, principally through rule making and some features of central monitoring, few adopted state financing. Using 1979 data, one study reported that only 8 states’ trial court systems were entirely or nearly entirely funded by the state; another 14 states had mixed state/local funding, with the state supplying more than half of that funding; and 28 states had mainly locally funded trial courts.^[5] Since 1985, more states have moved toward state financing, most notably California in the mid-1990s, and Florida by 2004. However, a mixture of state and local funding remains the norm in most states, and somewhat less than one-half the states’ trial court systems mostly remain locally funded. This shows that states are slowly taking on a greater share of court and judicial branch costs. For example, in 1984, states funded on average about 40% of total state and local costs for judicial and legal operations; by 2000 the proportion had grown to about 46% and with other changes since such as that in Florida, the proportion may now be approaching 50%. Greater state assumption is not always under the guise of adoption of state funding. Many times the creep toward greater state funding is programmatically focused such as was the case in Georgia when the state assumed in the early 2000s the responsibility for indigent defense costs.

State financing does not require highly centralized allocation and expenditure control from the state capitol. It is possible (and some would say preferable) to have relatively decentralized management control systems while using state financing. This was,

for example, part of the plan in the mid-1990s as California moved from mainly local to state support of trial courts. Florida may have ended up in 2004 with more of a balance between centralized control and decentralized management to the trial court level.

By the 1970s, court reform acquired a new emphasis that was less oriented toward reducing variance because, for the most part, one court of justice had been achieved in the states. The new goals were increased efficiency and effectiveness through continuing consolidation and centralization. In 1967, the President’s Commission on Law Enforcement and the Administration of Justice awakened reform interest throughout the criminal justice system and touted rational planning, consolidation, and coordination as means for dealing with the inequities and inefficiencies spawned by a political system that devolved too much to local control.^[6] The justice system, it was argued, was fragmented and in the absence of coordination it was inefficient and ineffective.

The press, for greater efficiency and effectiveness throughout government, was reinforced by the tax-cutting fever of the 1980s and 1990s, and by economic downturns in the early 1980s, early 1990s, and again 2002 through 2004. Improved court organization and management practices were directed toward reducing costs and improving services. The movement toward state financing was seen not only as a means to force additional reforms, but also as a way to shift the cost burden from local jurisdictions, which were increasingly unable to pay (for new staff, new programs, and for “black hole” areas of expenditure, such as indigent defense). Reducing budget inequities across trial courts and circuits that arise from largely locally financed systems also took on far greater importance from 1980 onwards. State financing and state-level allocation decisions in the interest of reducing funding inequities has become one of the strongest arguments for state funding.

EXPENSIVE AND DIVERSE JUDICIAL BRANCH OPERATIONS

Judicial branch costs have become a significant public expense, even though they constitute from 1% to 4% of total state expenditures (depending on what is counted and the proportion of total funding that comes from the state). State and local trial court expenditures nationwide are in the range of \$15 billion to \$18 billion annually; state appellate court expenditures (intermediate and supreme) add to the total. There are over 26,000 trial and appellate judicial positions nationwide and over one-quarter of a million other key court personnel.^[7] Large limited-jurisdiction urban trial courts spend well over \$100 million annually and, even among

the smaller courts, expenditures of several million dollars per year is not unusual.

Growth in cost has been driven partly by inflationary factors, but there are other important causes. The types of programs and services offered by trial courts have broadened considerably. Multiple formats of alternative dispute resolution, pretrial diversion programs, multiple types of pre- and post-trial services, help centers for unrepresented litigants, victim-witness assistance programs, a range of post-adjudication programming, and more customer-oriented services are common in courts today (and were not even 20 yr ago). Specialty courts are also much more common (e.g., mental health courts, drug courts, problem-solving courts, benches specializing in complex litigation). Growth in complex litigation such as toxic torts or class action product suits has increased the unit cost of trials on the civil side. Since the 1960s, annual caseloads have grown dramatically, and since 1988, caseloads have grown annually at an average of 5% per year. There are now nearly 1 million felony convictions per year in state courts.^[8]

INTERBRANCH RELATIONS AND CHALLENGES TO JUDICIAL BRANCH INDEPENDENCE

Increased costs have led to interbranch funding disputes. At both local and state levels, legislative and executive branches desire to manage judicial branch line items, and program allocations have strengthened, and in some cases threaten judicial branch independence. In addition, because most states have a mixed model of state and local funding for trials courts (including most of those states that are largely either state funded or locally funded), achieving coherency in the funding plan is challenged by inconsistencies between local and state priorities. Unitary budgeting is thereby challenged. Shifting or unclear agreements about which level of government pays for what (e.g., for court security, indigent defense, certain kinds of local programs, or shared use of local infrastructure, such as the County Clerks Office) can lead to “playing off” funding behaviors between levels of government.

Courts are sometimes caught between levels of government that argue over which gets to use court-generated fee and fine revenue. This revenue, generated by limited jurisdiction courts, typically produces surplus over cost, whereas the costs of general jurisdiction courts are rarely if ever covered. In the aggregate, trial courts produce insufficient revenue to meet costs; thus, pressure mounts on them to improve collection rates, tempting legislatures and executives to increase fee and fine rates. Increasing fee and fine rates, and asking courts to diversify their revenue streams (or spread costs to others) raises concerns about whether courts

can be entrepreneurial and do justice at the same time. Certainly, raising fees can limit access; and raising fines doesn’t mean they will necessarily be collected, perhaps leading to other sanctions as a consequence.

Stormy interbranch relationships over funding led to a spate of court-ordered funding actions during the 1970s, 1980s, and 1990s. Local funding authorities were angered. Bad press, challenges to public confidence in the courts, and concerns that judges were becoming too closely involved in local politics resulted. In some states, funding disputes between trial courts and local authorities reached sufficient heat and frequency that the Supreme Court and state court administrative offices imposed mediation and cooling off periods, and required Supreme Court or state court administrator’s approval before a local court order for funding could be enforced.

It was hoped that state financing would remove funding disputes from the local political thicket, and it usually did. However, in some states, it merely moved disputes to the state level. In some states, unpopular court decisions (e.g., the Supreme Courts striking down legislation as unconstitutional or otherwise flawed) led to retribution by legislative or executive branches in the form of “short” appropriations, or refusal to fund certain programs or new judicial positions. Severe budget shortfalls between 2002 and 2004 led many courts to adopt unpopular practices such as closing the court one or more days per week. These problems surfaced in both local and state-funded systems. State financing also did not necessarily solve the problem of unfunded mandates from the state legislature or from ballot initiatives (e.g., few, if any new funds followed to offset the impact of “three-strikes” legislation, which increased the number and cost of trials).

Interbranch funding disputes attracted sufficient national concern during the 1980s and early 1990s that the State Justice Institute funded a national conference in 1995 on funding the courts. The conference was cosponsored by the National Conference of Chief Justices, the ABA, and the National Conference of State Legislatures. During a conference plenary session, the problem was defined as a basic dissonance between the underlying assumptions of branches. The judicial branch asserted that it had the constitutional responsibility to define funding levels adequate to discharging its responsibility, and the legislative branch asserted its constitutional power of the purse. As a result, separate and coequal branches of government seemed inherently adversarial.

An alternative set of behaviors and assumptions was offered as a means to restore interbranch harmony. The legislative and executive branches, it was suggested, should 1) appropriate a lump sum to the judicial branch; 2) allow the judicial branch to set

priorities in allocating the lump to branch-set priorities; 3) permit the judicial branch authority to reallocate funds during the year; 4) permit the carryover of unspent funds across fiscal years as an incentive for implementing efficiencies; and 5) consult the judicial branch about effects on workload and attendant costs from proposed legislation. The quid pro quo was that the judicial branch should 1) become more managerially proactive in controlling costs; 2) improve or change practices in the interest of greater efficiency and effectiveness (e.g., by enforcing a firm continuance policy); 3) promulgate and be held accountable to performance standards; and 4) review priorities, discard non-performing programs, and reallocate resources to more important needs. Although not directly addressing whether the size of the “lump” might be sufficient in the first place, the quid pro quo sought a balance between power of the purse and judicial independence and sought to re-establish a functioning relationship between coequal branches.

Regardless of whether courts are locally or state funded, severe budget shortfalls impact them negatively as was the case from roughly 2002 to 2004. Budget shortfalls and reductions were serious enough that most states in a study conducted by the National Center for State Courts reported serious judicial branch budget reduction measures (e.g., of responding states 86% imposed hiring freezes, over half laid off temporary employees, nearly a quarter laid off permanent employees, a third delayed scheduled salary increases, all scaled back purchases, and 55% increased fee and fine rates). Courts are not isolated from budget reductions in time of budget or economic downturn. However, the lessons from the 2002–2004 period also support the view that compared to other agencies of government, courts tend to be held more harmless than most others. Courts are not immune, but do appear to have a somewhat favored status in lean times.

Interbranch relations remain problematic. Some state legislatures have become more engrossed in reviewing judicial branch line items, and some executives have begun to treat the judicial branch as another agency of executive branch in terms of budget submission and review. In a little over one-third of the states, the judicial branches’ budgets are submitted directly to the executive branch and its review lenses; in a little less than one-half the states, they are submitted directly to the legislature (and to both simultaneously in the remainder). In states where the judicial branch retains a fair degree of budgetary discretion and avoids the worst forms of interbranch warfare, politically astute and administratively skilled state court administrators and chief justices have been essential to forging positive and regular working relationships with key legislative and executive leaders. Similar skills of presiding judges and trial court

administrators are equally essential in locally funded systems. Perceived managerial competence^b and leadership has also risen in importance; courts perceived to be well run and that meet performance expectations appear to have a better environment for negotiating budgets and are given greater latitude to manage their affairs.

PROFESSIONALIZING THE MANAGEMENT OF COURT RESOURCES

In the early 1970s, the National Advisory Commission on Criminal Justice Standards and Goals, followed by the ABA, underscored the importance of professional administrators to improving the performance of the criminal justice system, including the courts. With the blessing and active involvement of Chief Justice Warren Burger, the Institute for Court Management (ICM) was formed in the early 1970s, with the mission of creating a professional cadre of court administrators for U.S. trial courts. Phase I of ICM’s core curriculum consisted of five topical areas, including budgeting. All graduates of ICM had to successfully complete the budgeting segment (along with the others). This requirement was the conscious recognition that competent budgeting and financial management skills were increasingly essential in court administration.

In 1998, the National Association for Court Management published the results of its 3-yr national study to identify the “core competencies” of court managers. Ten competencies were identified and “Resource Allocation, Acquisition, Budget, and Finance” was one among them. The study and subsequently published curricular guidelines concluded the following:

The allocation, acquisition, and management of the court’s budget affect every court operation, and arguably, determine how well, even whether, courts achieve their mission in the American political system. Allocating, acquiring, and managing financial resources is a core court management competency carried about by court managers, both judicial and administrative, and other court staff.... Effective court performance requires that court leaders—the executive team—have the ability: to set priorities and to manage competing demands on existing court resources in ways that build credibility, both internally and externally; to link resource requests to fundamental court purposes; to state court needs and objectives clearly and compellingly; and to protect judicial independence and

^bFor twenty years, the author has informally surveyed presiding judge and court administrator attendees at executive management programs. They have been asked to identify the factors most important to favorable relations with funding authorities. In the 12–15 factors usually identified, managerial reputation of the leadership team is typically ranked within the top two or three factors.

essential court functions while constructively negotiating with elected and appointed executive leaders and their staff.^[9]

This quotation is striking for many reasons, but for two in particular. First, there is implicit recognition of just how important and complicated budgeting and financial management have become in the modern court system. Second, there is reinforcement of the points made during the 1995 interbranch conference about the need for constructive behaviors between the judicial and the other branches of government.

The complexity of interbranch negotiations is the new reality. It is reflected in presiding judge rules in many states that assign presiding judges formal responsibility for the sound financial management of their courts. Although the addition of professional court administrators has helped, the sometimes complicated relationship between presiding judges and their court administrators confounds attempts at sound financial management and good interbranch relationships. Most judges, including presiding judges, have not had much management experience or much if any training in management practice. In addition, it is not unusual for presiding judges to change by election or rotation every 2 yr or so. Frequent changes in leadership make consistency in budget decision making, priority setting, stability in multiyear allocation plans, and the building of stable interbranch relationships problematic. The frequent turnover in presiding judges underscores the need for continuous training opportunities in budgeting and financial management.^c

The fact that professional budgeting practice in courts impacts judicial branch performance, accountability, and independence was underscored when the ABA's *Judges' Journal* issued a special edition in the summer of 2004 devoted entirely to the topic of, "Justice in Jeopardy: The State Court Funding Crisis."^[11] This was the first such publication devoted entirely to court budgeting. In part the special edition was in reaction to the serious budget shortfalls of 2002–2004, but also it reflected a "coming of age" of the centrality of budget management to judicial branch independence and accountability. Seventeen authors (academics included, but mainly judges, court administrators, and judicial officers) were united while making their individual contributions to underscoring

the importance of improved judicial branch administration of justice through more effective budget practice.

SHIFTS IN THE MIX OF LOCAL/STATE FUNDING

The movement toward state financing of trial courts has been underway for a number of years. By the late 1990s, nearly one-half of the states had wholly or substantially state-financed trial court systems and one-third of the states had substantially locally funded systems. The remainder were mixed with heavy components of both state and local support. Few states have full state financing (e.g., local jurisdictions almost always pay for facilities, and most locally funded systems still have state support for judges' salaries). However, the movement overall is toward greater portions of trial court budgets being supported through state appropriations.

The rhetoric for and against state financing involves numerous claims. The arguments in favor include greater equity in funding across trial courts, greater efficiency through scale, improved management quality, and more funds overall because of access to state coffers. Opponents assail proponents' claims as unsupported and predict that there will be loss of local control, more red tape, greater volatility in appropriations, simplistic formula funding, and a stifling of innovation.

Few if any of the claims on either side have been systematically researched, although there is strong anecdotal evidence that courts from the poorest jurisdictions are helped by state financing, that state financing makes it easier to temporarily shift judicial resources in a state, and that certain economies of scale are possible for programming that crosses trial court or jurisdictional boundaries (e.g., data systems).

OTHER REVENUE

As general fund appropriations fail to keep pace with perceived needs, the search for alternative sources of revenue intensifies. Fees and fines are substantial sources of court-generated revenue for both state and local jurisdictions (and for some states trial courts directly). However, fees and fines as a direct source of budget support for court operation are inherently problematic. If courts are expected to fund their own costs through fees and fines (or perhaps even produce surplus), it is not hard to imagine how justice and entrepreneurial motives could clash. Fees set too high limit access to the judicial system, and fines levied against the poor and indigent suffer from low collection rates if no consideration is given to the ability to pay. Fine revenue is a function of both the amount of the fine and its collectability. In addition, numerous states have already raised fee schedules over the last

^cFor an introduction to trial court budgeting, see Ref.^[10]. Also, the relationship between sound financial management skills and other managerial and leadership skills was underscored by the NACM study mentioned in Ref.^[9]. In that study, sound financial management was linked to relating resources to fundamental court purposes, to clear communication of purposes, and to an understanding of budget planning formats, problem solving skills, performance measurement, audit and request justification.

several years (especially during the 2002–2004 period), calling into question the “market’s” ability to absorb greater increases.

Numerous states, when moving from local to state financing, experienced difficulties accurately predicting how much of the cost of state assumption would be offset by redirecting fee and fine revenue from the local to the state level. Part of the problem was local government resistance to losing the fee and fine revenue, even though the local jurisdiction would no longer have to fund trial courts. Compromises were sometimes reached, which left a portion of the fee and fine revenue at the local level. Another problem is that the costs of state assumption have been underestimated significantly in several states. This, coupled to shortfalls in fee and fine revenue yield continuing controversy between the judicial and other branches of government.

Grants, principally from federal sources, such as the Office for Juvenile Justice and Delinquency Prevention or from various other programs within the U.S. Department of Justice, and from similar kinds of state programs are used by courts to fund innovation, special programs, or services and to provide enhancements not supported by general fund appropriations. However, grants rarely if ever provide ongoing funding, and pressure is put on court administrators to find replacement dollars for successful programs, either from other external sources or through internal reallocation. Federal reimbursement programs such as those for IV-D child support enforcement can return revenue for meeting federal objectives. Federal anticrime legislation adds potentially renewable revenue for targeted purposes (e.g., drug courts). On the plus side, grants provide the “risk capital” for trying new things, and also a funding bridge-period to finding permanent dollars, including through eventual reallocation of existing funds.

Other ideas for supplementing revenue include the use of volunteers, who if properly trained and assigned, supplement the personnel budget. Other ideas periodically surface such as “cut-of-the-action” schemes for the civil side of the docket where winners pay the courts a percentage of their “take” if they refuse mediation or others forms of dispute resolution, which are usually less costly than full-scale trials. In the ABA special edition of *The Judges’ Journal* referenced earlier, some authors identified other creative sources of funding such as, in exchange for depositing funds in a particular bank, the bank provides the court free or discounted financial services.

EMERGING ENVIRONMENTS FOR ACQUIRING RESOURCES

The reality of court funding is that a new relationship is evolving between the judicial branches and the

other branches, one that requires a quid pro quo, mutual responsibilities, and ongoing dialog across branch leadership on issues of court costs and financing. Regardless of promising alternative revenue sources, general fund appropriations, whether from state or local levels, will remain by far the largest source of money for funding judicial branch operations.

Courts will have to compete with other agencies of government for scarce resources. The *acquisition* of resources and the *allocation/management* of existing resources will increasingly be tied. Improved practice with the view of increasing both performance and efficiency can potentially yield resources for internal reallocation, as well as improve court management reputation with funding authorities.

Priority setting in resource requests helps to focus attention on the most important of needs, and a strategic approach to priority setting can help courts to focus requests rather than to diffuse or constantly change them. Frequent shifts in funding priorities send a message of unclear leadership. Multiyear budget planning will be required for many areas of new expenditure, such as for technology and other forms of organizational change, because the costs of these can only be met over several years. This would further suggest that courts increasingly will need to “stabilize” their strategic expenditure priorities, thinking through not only multiyear commitments to achieving long-range objectives, but also whether proposals involving new costs (e.g., new personnel) unintentionally commit the court to additional multiyear obligations. New multiyear obligations reduce allocation flexibility by tying up ever larger amounts of money. In addition, in times of fiscal constraint, funding authorities, although generally inclined not to incur new obligations of any kind, are particularly resistant to new multiyear rather than one-time obligations. One-time or 1- or 2-yr commitments are more palatable than permanent new allocations.

Interest in having line agencies self-fund or at least partially fund new initiatives is growing as a means for leveraging central resources (and vice versa). Self-funding or partnership funding requires courts to institutionalize organizational change as a means of freeing resources (e.g., by discarding low value-added programs or figuring out how to do the job differently and less expensively, such as, substituting court recorders for court reporters, or looking for partnerships with other public agencies or the private sector (e.g., in collections).

Forced efficiency reductions to court budgets by funding authorities is another reality with which court managers may have to contend. “Efficiency” reductions are a way for funding authorities to create policy reserves for reallocation, as well as a means to squeeze fat out of line agencies. In the process, the ongoing

budget is deliberately underfunded as a means to force organizational change (typically 1% to 3% in a given year). Agencies can usually absorb these over 1 or 2 yr, but over longer periods they are nearly bled to death unless fundamental change takes place either in the service mix or in how tasks are performed, or in how successful they are in attracting new funds from the policy reserve. Shorting budgets in the interest of creating policy reserves or in forcing efficiencies may or may not be openly announced, but are real nonetheless.

Change required to improve judicial branch efficiency and effectiveness will require a reshaping of programs and services that in some instances will require forming partnerships and effective cross-walks between the judicial and other branches of government, and between courts and other agencies of the criminal justice system. Joint commissions, interagency working groups, and standing interbranch staff analysis groups will be needed not only to review proposed legislation, but also to cooperate in meeting the costs of the many new kinds of programs that have become a more or less standard part of court programming.

CONCLUSION

Judicial branch acquisition of sufficient resources and the authority to allocate them are fundamental to meeting branch missions and purposes. The efficient and effective use of such resources is a responsibility and a core element of accountability. Responsibility and accountability are complicated by the nature of work in courts, by rising and more visible judicial-branch costs, by diverse expectations for branch performance, and by increased competition for scarce public resources. Court budgeting is not simply an issue of dollars and what dollars buy, but also of work performance and valued outcomes from the use of such resources. In the end, the judicial branch purposes of justice, equity, access and timeliness are served by a conjoining of resource acquisition and effective resource management.

REFERENCES

1. Pastore, A.L.; Maguire, K.; Eds. *Sourcebook of Criminal Justice Statistics 2000*; 3–4; http://www.albany.edu/sourcebook/Sourcebook_of_Criminal_Justice_Statistics/1995/pdf/t12.pdf (accessed December 2001).
2. American Bar Association. *Commission On Standards Of Judicial Administration Standards Relating to Court Organization*; American Bar Association: Chicago, 1974.
3. Berkson, L.C. The emerging ideal of court unification. *Judicature* **1977**, *60*, 372–382; Lawson, H.O.; Gletne, B.J. *Fiscal Administration in State-Funded Courts*; National Center for State Courts: Williamsburg, VA, 1981.
4. Skoler, D.L. Financing the criminal justice system: the national standards revolution. *Judicature* **1976**, *60*, 32–38.
5. Hudzik, J.K. Rethinking the consequences of state financing. *Justice Syst. J.* **1985**, *10* (2), 135–158.
6. The President's Commission On Law Enforcement. *The Challenge of Crime in a Free Society*; Commission Report: Washington, DC, 1967.
7. Rottman, D.B.; Flango, C.R.; Cantrell, M.T.; Hansen, R.; Lafountain, N. *State Court Organization 1998*; NCJ 178932; U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics: Washington, DC, 2000; 7–18; Also see, Pastore, A.L.; Maguire, K.; Eds. *Sourcebook of Criminal Justice Statistics 2000*; 26–30; http://www.albany.edu/sourcebook/Sourcebook_of_Criminal_Justice_Statistics/1995/pdf/t120.pdf (accessed December 2001).
8. Pastore, A.L.; Maguire, K.; Eds. *Sourcebook of Criminal Justice Statistics 2000*; 453; http://www.albany.edu/sourcebook/Sourcebook_of_Criminal_Justice_Statistics/1995/pdf/t540.pdf (accessed December 2001).
9. NACM Professional Development Advisory Committee. Core competency curriculum guidelines: history, overview, and future uses. *Court Manager* **1998**, *13* (1), 6–18.
10. Tobin, R.W. *Trial Court Budgeting*; Court Management Library Series; National Center for State Courts: Williamsburg, VA, 1996.
11. Hudzik, J.K. Justice in Jeopardy: the State Court funding crisis. *Judges' J.* **2004**, *43* (Summer).

Court System Strategic Planning

Deborah A. Botch

New York State Unified Court System, Albany, New York, U.S.A.

INTRODUCTION

The courts of the United States have experienced a quiet revolution over the past two decades. Once strictly neutral forums for case adjudication, courts have evolved into tribunals with a dual function. They remain the primary forum for resolving civil and criminal matters; but today courts also proactively seek ways to solve human and community problems. This fundamental change has brought with it new service demands and organizational challenges that have created a need for more systematic and integrative long-term planning. To confront this increasingly complex and dynamic environment, court systems have turned to long-established strategic planning tools and methods. This article explores concepts, models, and common practices adopted by federal, state, and local court systems in developing and implementing strategic plans.

PUBLIC SECTOR STRATEGIC PLANNING

By definition, strategic planning is a formalized and systematic process used by any organization, enterprise, or community to identify and achieve future goals through collective action. Long employed by large private sector corporations as a means of improving competitive advantage, strategic planning did not come into widespread public sector use until the 1990s.^[1] Faced with ever-increasing social and technological challenges, coupled with vocal demands for improved performance and accountability, government and non-profit organizations are now keeping pace with the private sector in instituting strategic planning processes as a matter of course. Today, strategic planning plays an important role in the management of many governments, government agencies, and non-profit organizations.

The public sector has adopted strategic planning as part of its organizational management structure for a number of reasons. According to Mintzberg,^[2] strategic planning processes serve multiple purposes. They are effective devices for self-assessment, improved communication, and consensus building. Bryson^[1] points also to strategic planning's utilization of participative

decision making, which is central to solving public problems in a democracy, as a key reason why the approach has been widely adopted and persists.

Legislative mandate provides another impetus to public sector adoption of strategic planning. Some governments have enacted laws and regulations that call for the submission of regularly updated strategic plans to executive and legislative authorities. A case in point is the Government Performance and Review Act (1993) that directs federal agencies to develop and submit strategic plans to Congress and the Office of Management and Budget (OMB).^[3] Strategic planning and routine submission of plans and performance reports to Congress and OMB are an ongoing part of the management processes of federal agencies.

The availability and proven track record of a number of strategic planning models have also contributed to the rapid expansion of public sector strategic planning. Bryson^[1] has developed a widely recognized and frequently referenced model for public sector strategic planning in the United States. Bryson's^[1] process consists of the following stages: initiate and agree on a strategic planning process; identify mandates; clarify mission and values; assess the organization's external and internal environments; identify strategic issues; formulate strategies to manage issues; review and adopt the strategic plan; establish a vision of success; develop an effective implementation process; and, reassess strategies and the strategic planning process.

COURT SYSTEM STRATEGIC PLANNING

Like its executive branch and local government counterparts, the judicial branch operates in an increasingly dynamic and uncertain environment. Increased case loads, public demands for greater accountability, and the introduction of grassroots problem-solving court programs are changing the environments in which courts operate. To respond to these changes, many courts have instituted strategic planning processes over the last decade and they continue to do so. The National Center for State Courts (NCSC) provides an extensive bibliography of court-related strategic planning materials as well as a comprehensive state-by-state reference list of court strategic plan documents.

The NCSC also supplies access to many electronic documents and links to sites containing court strategic plans and process information at their web site, <http://www.ncsconline.org/>.^[4]

Strategic Planning in Federal Courts

The federal judiciary has long operated under a formalized process of strategic long-term planning that reflects its unique Constitutional obligations. Administration of the federal courts, which includes long-range strategic planning among its various functions, is carried out under the policies and direction of the Judicial Conference of the United States. The Judicial Conference, made up of Senior Circuit and District Court Judges, directly oversees the Administrative Office of the Courts and authorizes and establishes standing and advisory committees that deal with specific areas of administration.^[5] In 1990, the Judicial Conference began a (<http://www.uscourts.gov/lrp/index>) long-range planning process with its establishment of the Committee on Long-Range Planning. The Committee's work reflected a highly decentralized and consultative process consistent with the federal judiciary's organizational structure and legally defined operations.

In meeting its charge, the Committee reviewed key historical plans and reports and coordinated its work with other committees of the Judicial Conference. These committees collaboratively prepared a number of subject-specific long-term plans, which included input from individual judges, supporting documentation from several judiciary sponsored research projects, and extensive public review and comment. Four years later, in 1994, the final product was a comprehensive draft long-range strategic plan containing dozens of long-term goals and implementation strategies.

In December 1995, the Judicial Conference of the United States adopted a final version of the *Long-Range Plan for the Federal Courts*, which contained a Vision and Mission Statement and 93 goals and 76 implementation strategies. A complete copy of the *Long-Range Plan for the Federal Courts* is available at the federal courts web site, <http://www.uscourt.us.gov/lrp>.^[6]

Strategic Planning in State and Local Court Systems

At the state and local level, while many different approaches have been applied to strategic planning development and implementation, most court systems have relied on only a few strategic planning methodologies and models to guide and coordinate the process. According to Martin and Wagenknecht-Ivey,^[7] these methods and models address organizational improvement issues and reflect an external and community

focus. They invite broad-based public participation and include stakeholders from outside the court to assure sustained community involvement. According to Rottman et al.,^[8] community collaboration in the court planning process inculcates greater understanding of the role of the courts and fosters public trust and confidence.

While the composition of planning committees varies, two distinct approaches are generally used in structuring court strategic planning committees. According to Martin and Wagenknecht-Ivey,^[7] the first approach is broad-based, made up of judges, court personnel, local government officials, lawyers, citizens, and community groups. In the second approach, the committee is entirely composed of judges, court staff, and court system administrators.

In the following section, two models that illustrate the key attributes, processes, and methods common to court strategic planning will be discussed:

1. Court Community Planning Model (Judicial Council of California).
2. Nine-Step Court Strategic Planning Process (Center for Public Policy Studies).

COURT STRATEGIC PLANNING MODELS

State of California Court Community Planning Program

The State of California Court Community Planning Program is an excellent example of the application of the community-focused model of court planning. In keeping with the model, the California program has been designed to promote community outreach and development of court–community partnerships.^[9]

The Judicial Council of California, the governance board for the State's judicial branch, is responsible for improving the administration of justice in California. The Council has adopted an official strategic planning process and management cycle that is integrated with other administrative processes, including annual judiciary budget preparation. In 1998, under the direction of the Judicial Council, the California courts launched its community-focused strategic planning process. The process fosters inclusion of community stakeholders as full partners in the planning process and relies on a five-stage community-outreach model developed in consultation with national experts.^[10] Much of the planning in this model is delegated to courts at the county level and is accomplished by teams of county court judges, court staff, local government officials, bar representatives, and community groups.

Several features of the five-stage model are common to other public sector strategic planning models. For

instance, there are similarities with strategic planning steps recommended in Bryson's^[1] framework, such as clarifying vision and mission, scanning the environment to determine important issues and trends, and development of goals and measures of success. But what sets the California model apart is its emphasis on broad-based public participation and the "enablement" of public involvement.

To enable community participation, the California model recommends creation of planning teams that include broad-based representation both from within and outside the courts. In the enablement phase, training is provided to all planning team members in the basics of strategic planning and use of the five-stage model as a means of building planning capacity and group cooperation and cohesion.

The California Courts Community Planning Model has generated a variety of local ideas for improving court operations and making courts more responsive to the unique needs of the State's socially and economically divergent communities. This planning process has also spawned a variety of community justice partnerships. For example, court- and school-based education programs have been instituted in a number of small rural jurisdictions to enhance student understanding of the legal system. Elsewhere, a citizen advisory board has been established to build court and community trust and cooperation in the Los Angeles Superior Court.^[8] The success of the collaborative

model used in California can also be seen in the rapid response and participation in the planning effort. As of 2002, all 58 counties of the State had submitted court community strategic plans to the State's Judicial Council.^[10]

Nine-Step Court Strategic Planning Process

By far, the most popular model used for court system strategic planning is that developed by John A. Martin and Brenda J. Wagenknecht-Ivey as consultants to the Center for Public Policy Studies (CPPS). According to Martin and Wagenknecht-Ivey, the *Nine-Step Court Strategic Planning Process* offers a standard systematic approach that can be readily adapted for use by any court system, court, or court subdivision. Although input from the wider community is a feature in the model, it is not a requirement. The nine-step model focuses on organizational change, learning, and improvement. As of the year 2000, more than 50 state and local courts and other justice agencies have reported adoption of this model for strategic planning (Fig. 1).^[7]

According to Martin and Wagenknecht-Ivey,^[7] because strategic planning requires significant resources and time commitments, courts need to be sure that conditions are right for introducing strategic planning. The authors recommend that, prior to undertaking

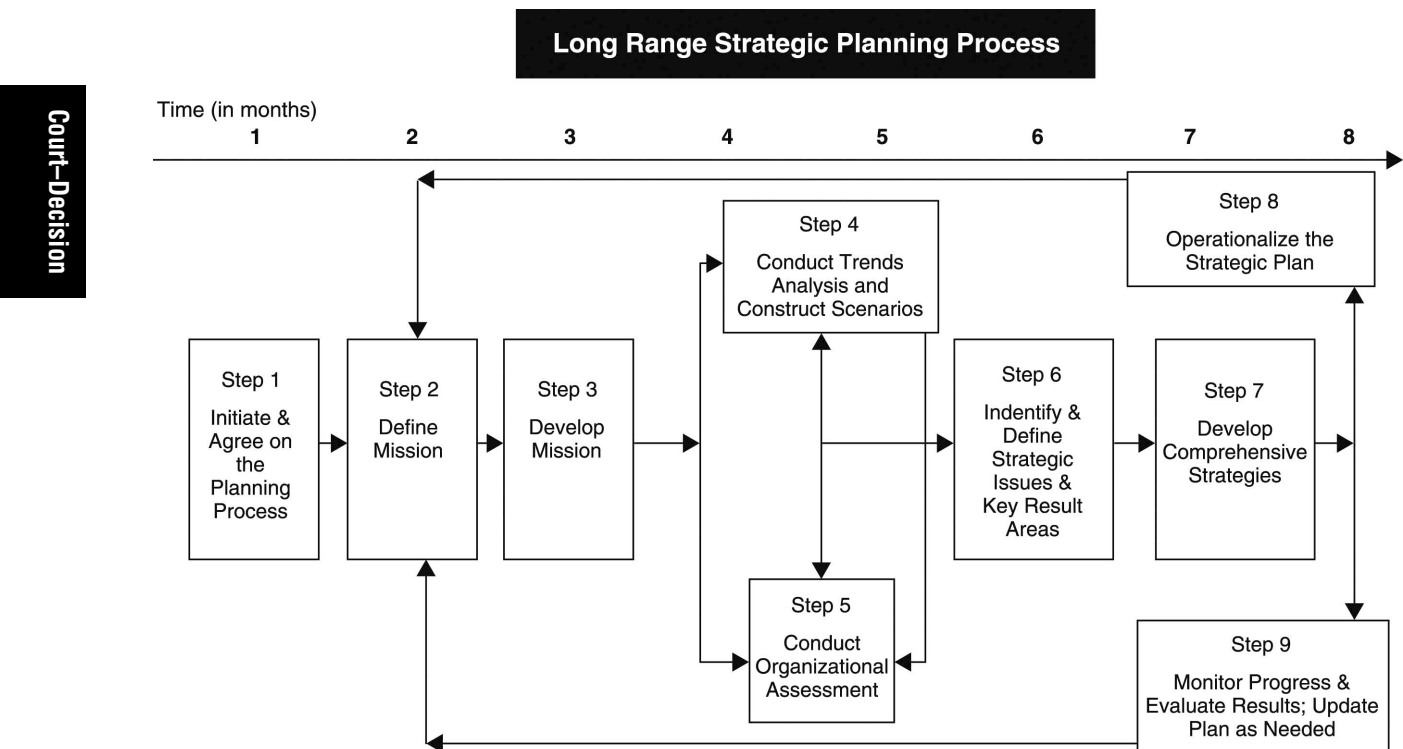


Fig. 1 A nine-step strategic planning process for courts. (From Ref.^[7])

strategic planning, court leaders must be able to focus on long-term priorities, be ready to embrace organizational change, and be willing to involve a wide-range of participants in planning. Wagenknecht-Ivey has developed a tool to score and evaluate an organization's conditions of readiness.^[7] If "readiness" criteria are met, court organizations are given the green light to move ahead with a strategic planning efforts.

Following the readiness assessment, the nine-step model begins with a commitment from court leaders and managers and an agreement from the planning group about the scope and purpose of the effort. Steps 2 and 3 focus on development of vision and mission statements. Step 4 consists of a trend analysis and scenario constructions. Scenario construction, according to the authors, can help planners picture how a variety of trends may impact on the courts and how courts can shape a positive future. They are creative tools that involve speculation and imagination, but are not predictions. Use of scenarios facilitate systematic future-oriented and "what if" thinking to help organizations better understand the potential forces impacting on desired future conditions.^[11]

Step 5 in this model is an organizational assessment that, according to the authors, allows the court to see how its structures, policies, attitudes, and resources align with "expected" and "desired" future conditions. In step 6, strategic issues and key-result areas are identified. Strategic issues are the critical policies or challenges to the court organization that will impact on future effectiveness. Step 7 involves development of comprehensive strategies to respond to each strategic issue area. The authors consider this step to be the heart of the model and recommend that broad goals, objectives, and end targets be established.

Step 8 moves plan formulation into operational and action planning stages and involves project selection, setting time frames, and evaluating resource availability and impacts.^[7] The model assumes that strategic planning will be an ongoing and continuous process. As a result, step 9 urges courts to set up methods for monitoring and evaluating results and developing plan updates.

Mentoring guidelines have been prepared as a practical tool to guide court strategic planners in using the nine-step model.^[7] The guidelines provide a case study of the Florida judicial branch long-range strategic planning effort. This case study illustrates the application of the nine-step model in an actual court system environment and shows how the model facilitates a court strategic planning process. The value of the nine-step strategic planning approach is confirmed not only by the Florida courts example, but by the number of other strategic plans it has helped foster. For example, see the plans of several statewide court systems—California, Michigan, and Wisconsin, and

at the trial court level—Los Angeles County Superior Court, 36th District Court in Detroit, Michigan; and, Yakima County Superior Court in Washington State.

Other Tools That Aid Court Strategic Planning

Along with models that are used to guide, coordinate, and quicken the pace of strategic planning, several other practical tools are available to facilitate the strategic planning process in state and local court systems. One such tool is a set of performance standards that have been adopted and used by hundreds of trial and municipal courts across the country to improve the quality and accountability of court programs and services.^[12] The Trial Court Performance Standards (TCPS),^[13] developed by the Commission on Trial Court Performance Standards, provides 22 guiding principles covering five public interest areas:

1. Access to justice.
2. Expedition and timeliness.
3. Equality, fairness, and integrity.
4. Independence and accountability.
5. Public trust and confidence.

To support vision and mission statement, development state and local courts often rely on these national performance standards as a key resource.

A second tool available to state and local court systems is one that can help shorten the time frame and reduce the cost, complexity, and work required of strategic planning. This tool is referred to as the "environmental scan" and is a concept presented in detail in the *Environmental Scan for State Courts 2002*, developed by the National Center for State Courts and Futures. com.^[14] As explained in the overview of the *Scan* report: "Environmental scanning attempts to identify events, trends, and developments or drivers, shaping the future . . ." Trends, for instance, can include those in social, scientific, technological, economic, political/governmental, and professional categories.^[14] In its broadest sense, environmental scanning involves an effort to take a court beyond its current ways of doing things (or encouraging it to rethink its "paradigms"). The scan report can be used by a court or court system to support strategic planning either as the sole resource for identifying forces and changes in the external environment or in combination with locally prepared environmental trends information and analyses.

A third tool that can be used as a foundation for strategic planning is the information gathered in court "futures" reports. As indicated by the National Center for State Courts web site, during the 1990s, 22 state court systems created futures commissions or committees. These study groups were generally initiated by

visionary state judicial leaders whose objectives were to assess economic, social, cultural, technological, and political trends and to consider how court systems should adapt their structures, rules, workforce, operations, and intergovernmental and community relationships to effectively respond to expected changes.

The Ohio Courts Future Commission undertook one such “futures” study. Conducted by a variety of legal experts, community and business leaders, and citizens, the Ohio study provides a good example of the breadth of issues and long-range horizons generally reflected in court futures studies. The Commission’s final report identified 10 key attributes of the Ohio courts that would be desirable by the year 2025, including, among other things, courts that are physically, economically, and functionally accessible to all citizens, a jury system that respects jurors and lets them take a more active role in court proceedings, selection of judges based on superior legal and personal qualifications (trained and continually reeducated on topics of relevance to the courts), and a workforce that is well trained and professional.^[15]

Strategic Planning Issues and Future Directions

Even with its benefits, such as clearer mission and goals, coherent decision making, and improved performance, strategic planning may not be suitable for some court organizations or in all circumstances. Bryson^[1] cites a number of difficulties and challenges that must be addressed to ensure success, including, foremost, the commitment of key leaders and stakeholders. Bryson further recommends that if implementation appears unlikely, strategic planning should not be considered. Martin and Wagenknecht-Ivey^[7] also recommend that courts assess their readiness for change and the commitment of leaders and stakeholders before embarking on a strategic planning effort. Critics of strategic planning are also concerned with the costs and time-consuming aspects of the process. Hamel,^[16] for example, argues that strategic planning is not only too costly but also does not provide for the fast-paced strategic development needed by today’s organizations.

Henry Mintzberg,^[2] in his classic work on the “rise and fall of strategic planning,” cites several other limitations of strategic planning including the conflict between the creativity necessary for innovation and the logical and systematic requirements of formal planning. For the public sector, he reiterates the challenge of ensuring that the desires of the people are adequately represented. Mintzberg^[2] does not entirely dismiss the process, but notes that while strategic planning may have benefits, at best, “[it is] a process with particular benefit in particular contexts” (p. 4). More recently, Raffoni^[17] argues that strategic planning often founders

at the execution stage because of a limited understanding by management of what it takes to successfully implement strategies. Raffoni also contends that few plans recognize the importance of translating broad-brush conceptual strategies into decisions about who will carry out the associated tasks and how much time and money will actually be expended.

In spite of these difficulties, the need for effective methods to deal with the dynamic forces of change remains a major challenge for court systems and other public sector institutions. Strategic planning and related techniques that facilitate communication, collaboration, goal setting, and accountability have helped many public sector organizations better prepare for the future. At the same time, there is recognition^[18] that more robust, focused, coordinated, and democratic methods will be needed if strategic planning is to remain a viable and beneficial tool for the next generation of public leaders, managers, and citizens.

CONCLUSION

Strategic planning concepts, tools, and models have helped many court systems prepare for the future and transform themselves into more responsive public organizations. As the pace of change quickens and public expectations for effective performance increases, systematic planning processes will become even more critical in ensuring the effectiveness of courts and court/community cooperation at all governmental levels. Recognizing this, several state court systems have made strategic planning an integral and ongoing organizational management process. Also, the National Center for State Courts continues to expand its resources and services to help court systems build and sustain their strategic planning capacity. While public sector conditions have dramatically changed in recent years, those courts that have a history of long-range, systematic, and broad-based strategic planning are likely to be able to adapt to changing conditions and meet new demands for service and effectiveness now and in the future.

REFERENCES

1. Bryson, J.M. *Strategic Planning for Public and Non-profit Organizations: A Guide to Strengthening and Sustaining Organizational Achievement*; Jossey-Bass Publishers: San Francisco, 1995.
2. Mintzberg, H. *The Rise and Fall of Strategic Planning*; The Free Press: New York, 1994.
3. GAO *Executive Guide—Effectively Implementing the Government Performance and Results Act (GAO/GGD-96-118)*; U.S. General Accounting Office, 1996.

4. National Center for State Courts. **2003**, <http://www.ncsconline.org/> (accessed 3/20/03).
5. Judicial Conference of the United States. <http://www.uscourts.gov/judconf> (accessed March 5, 2003).
6. Long Range Plan of the Federal Courts. **1995**, <http://www.uscourts.gov/lrp> (accessed March 5, 2003).
7. Martin, J.A.; Wagenknecht-Ivey, B.J. *Strategic Planning Mentoring Guidelines: Practical Tips for Court Leaders*; State Justice Institute: Alexandria, VA, 2000.
8. Rottman, D.; Efkenman, H.; Hansen, R.; Stump, S. *A Leadership Guide to Statewide Court and Community Collaboration*; 2002, <http://www.ncsconline.org/WCDS/Pubs/> (accessed February 19, 2003).
9. Judicial Council of California. *Community Focused Court Planning Initiative: Fact Sheet*. January, 2003. <http://www.courtinfo.ca.gov/reference> (accessed April 16, 2003).
10. California Courts. *Court Community Planning*; <http://www.courtinfo.ca.gov/programs/planning> (accessed March 6, 2003).
11. Martin, J.A.; Wagenknecht-Ivey, B.J. Courts 2010: critical trends shaping the courts in the next decade. *Court Manager* **2000**, 15 (1), 6–15.
12. Keilitz, I. *Performance Based Strategic Planning (PSPB) in the Courts: A Planning Process Anchored in the Trial Court Performance Standards*; Sherwood Associates: Williamsburg, VA, 1997.
13. Bureau of Justice Assistance. *Trial Court Performance Standards and Measurement System*. National Center for State Courts: Williamsburg, VA. <http://www.ncsconline.org/D-research/TCPS> (accessed April 16, 2003).
14. *National Center for State Courts & Futurist.com 2002 Report on Trends in the State Courts*; National Center for State Courts: Williamsburg, VA. <http://www.ncsconline.org/WCDSPubs/> (accessed March 21, 2003).
15. *A Changing Landscape: Ohio Court Futures Commission Report*; Ohio Courts Future Commission, May, 2000.
16. Hamel, G. Strategy as revolution. *Harvard Bus. Rev.* **1996**, 74 (4), 69–82.
17. Raffoni, M. Three keys to effective execution. *Harvard Manag. Update* **2003**, 8 (2), 1–4.
18. Rubenstein, H. Strategic planning tools for futurists. *Futures Res. Q.* **Fall 2000**, 16 (3), 5–17.

Court Technology and Information Systems

Ronald M. Stout Jr.

4th Judicial District, New York State Judiciary, Latham, New York, U.S.A.

INTRODUCTION

A full discussion of the concepts, applications, promises, and realities of court technology and information systems would require, and indeed deserve, an extensive analysis and discussion far beyond the confines of the space possible in the *Encyclopedia* (for example, see Ref.^[1]). Instead, this article will identify court functions that can be enhanced by the application of technology, the types of application software that are available in which to develop these functions, equipment and connectivity options depending upon the size and scope of the functions, and courts, design, implementation and maintenance considerations, and cautions to be exercised.

Developing technology and information systems for courts is a highly appropriate activity, because receiving, processing, and disseminating information is the method by which courts achieve their objective. From the initial filing of a case to the final postdisposition action, litigants, the bar, judges and their staffs, and the public, are entirely dependant upon timely and accurate information about the status and the substance of the case. Courts have always had information systems. Manually maintained docket books, ledgers, index cards, and file folders, combined with handwritten documents and notices are information systems. Some courts to this day continue to process information in this manner, while many more have hybrid systems that combine manual and automated components.

ENHANCEMENT OF FUNCTIONS BY TECHNOLOGY AND INFORMATION SYSTEMS

Virtually every function required by a court can benefit from the application of technology and information systems. Perhaps the first function that comes to mind when considering court information systems is case-flow management. Case-flow management, dealt with more extensively elsewhere in the *Encyclopedia*, “involves the entire set of actions that a court takes to monitor and control the progress of cases, from initiation through trial or other initial disposition to the completion of all postdisposition court work,

to make sure that justice is done promptly.”^[2] A comprehensive case-flow management system should include the ability to:

- Record all incoming documents or actions that initiate and continue a case, including those that are filed electronically.
- Assign a case to a judge.
- Schedule all events and record the outcome of those events.
- Monitor compliance with interim and final deadlines, and produce appropriate warning notices.
- Produce scheduling and non-compliance documents (such as warrants).
- Record all fees and fines levied, collected, or delinquent, and produce appropriate documents.
- Record all individuals associated with the case, including litigants, attorneys, probation officers, and court-appointed representatives or fiduciaries, and produce appointment and termination notices.
- Allow ad hoc inquiries to determine the status of one or more cases, the schedules for any day or judge, and the cases with which one or more participants are involved.
- Produce all required statistics, including those describing case processing, judicial workload, and court assignments of attorneys and fiduciaries.

If automating case-flow management can be considered applying technology to the processing of cases, a second function that can be assisted by automation is the substance of a case. Applying automation to this function assists judges by, for example, producing substantive orders, calculating structured settlements in civil cases or child support payments based upon income and expenses, producing transcripts of judgments and criminal commitment orders. In addition, automated systems can permit electronic filing of initial and subsequent papers, answers, motions, depositions, fees, and proposed orders. The line between this function and the case-flow management function is not definitive—after all, a case-flow management function should produce warrants and commitment orders. However, automated systems can offer judges the option of selecting a wide variety of standard provisions and provide calculation power for complex financial proposals. Even orders that require extensive

tailored provisions can benefit from the automated capture of the names of the parties, the issues, and the attorneys.

A third function is computer-assisted legal research, permitting judges and staff to conduct rapid and comprehensive legal research from their offices or from the bench. Materials available on-line include:

- Federal, state and local statutes.
- Appellate and trial court decisions and opinions.
- Treatises on specific areas of law.
- Citation checking in a draft opinion or decision.

A fourth function that can benefit from the application of technology is case conferencing and presentation. This includes:

- Telephone and videoconferencing to conduct scheduling, settlement, and pretrial conferences.
- Videoconferencing for trial testimony.
- Video display units for presentation of evidence during trial, with images of evidence displayed on screens of opposing attorneys, court clerks, the judge, and jurors.

Establishing a record of court proceedings is another function in which technology can improve efficiency and effectiveness. Software is available that:

- Instantaneously transcribes court reporters' notes, which greatly facilitates read-backs and the production of transcripts.
- Where permitted by law, electronically records court proceedings using either audio or video.

Courtroom utilization in large facilities can also benefit from technology. The demand for courtroom space in large urban settings is frequently very high. When a case in a trial settles, a trial is suspended for a significant period of time, or a case goes to a jury, the use of telephones, e-mail and even radios can notify judges and court staff of the availability of a courtroom.

Technology and information systems can also be applied to records management with great effect. Microfilm is an excellent records preservation tool but generally not a good management tool. Imaging systems that capture documents directly or in concert with microfilming systems make documents available to judges, staff, and the public, under the appropriate controls.

Technology can improve jury operations by:

- Merging mailing lists from different sources to establish the widest possible pool of potential jurors.
- Printing qualification and summoning notices.

- Tracking the response to those notices and producing corrective documents.
- Monitoring juror utilization once in the courthouse.
- Recording attendance.
- Producing documents necessary to authorize payment to jurors.

Communication within and without the court is another function ripe for the application of technology and information systems. Effective communication is essential for any organization, but particularly for the courts. The multiplicity of players in the courts makes establishing and maintaining effective communications especially difficult. This includes judges, staff, the private bar, executive branch agencies at the local, state, and federal levels, and litigants.^[3] Each organization and individual may have vastly different communications capabilities, ranging from integrated information and technology systems to homes without a telephone.

Communication covers a broad area, including:

- Informing each other about events affecting the status of a case, such as motions, agreements, scheduled appearances, proposed and final orders, and settlements.
- Advising attorneys about court rules, forms, schedule of terms of court, and changes thereto.
- Providing the public with a description of the court system and the purposes and general processes of the different components of the court, hours of court, directions to the court, types of proceedings that can be brought and the forms to be used, and a directory of related agencies.

Automating administrative functions has long been practiced by the public and private sectors. Administrative functions that can be automated include:

- Preparing budgets, and submitting them to the appropriate legislative body.
- Establishing expenditure accounts based on appropriations, tracking encumbrances, expenditures, reporting balances, and providing projections.
- Issuing purchase orders, and preparing vouchers for payment.
- Maintaining inventories of accountable equipment.
- Maintaining lists of authorized position titles and their qualifications.
- Maintaining lists of positions authorized to be hired, production of vacancy reports, and turnover rates.
- Tracking employee information, including data on hiring, salary, promotions, disciplinary actions, fringe benefits, training, and separation.
- Issuing paychecks.

- Tracking the availability of judges and non-judicial staff for assignment throughout the appropriate jurisdiction.

Many of these functions may be performed in whole or in part by local or state governmental agencies. Nevertheless, a careful review of the court's needs is advisable to determine if supplemental systems would be useful.

TYPES OF APPLICATION SOFTWARE

There are many types of software in which applications can be developed to perform the functions needed by the courts. In addition, the types of software have increasingly been bundled into "suites" of software that easily interact with each other. Increasingly, also, each type of software has elements of other types of software built into their design.

Database software is highly useful for developing systems to augment most court functions. A proper and comprehensive database application will identify all entities (e.g., people, places, events) and have the capability of storing all required information about those entities (e.g., names and addresses of people, court appearance dates, purposes of appearances and their outcomes, courtroom numbers, appropriations, voucher amounts, and personnel titles).

Database software programs have the capability of retrieving and presenting information in printed reports, on video display terminals, or electronically to other systems inside or outside of the courts. Data can be retrieved on a periodic or ad hoc basis. It can be tailored for any number of purposes, from providing a complete case history for a judge, to confirming a court date to a member of the public, to providing the latest court decision on a topic of interest.

Once the data is entered, it can be used for a virtually endless stream of documents, reports, and displays. Routines can be written to periodically scan records for compliance with processing dates and produce exception reports, compile statistics, and compute balances.

Word-processing programs can be used alone or in concert with database programs to produce documents for every court function. Templates can be written for standard documents, and automated messages can be built in to prompt for specific information. Alternatively, data can be automatically transferred from database programs to populate standard information, and specific information can be added to the document. Word-processing programs also have the capability of storing names, addresses, and other information in a database-like format, which can then

be merged with a standard letter to produce mass mailings.

Spreadsheet programs are used primarily for manipulating numbers through mathematical formulas. Templates can be designed that are extremely useful for calculating the consequences of different proposed settlements or payments required for support payments. They can also be used for projecting expenditures under different assumptions. It is also interesting to note that relational database systems store data in tables that are, in effect, spreadsheets. Under carefully controlled design criteria, a spreadsheet can be used as a source of data in a database program, and vice versa. Finally, word-processing programs usually have a table feature, which also resembles a spreadsheet. Spreadsheets can be imported into word-processing programs as tables. This can be very useful, if, for example, a structured settlement worked out in a spreadsheet needed to be incorporated into a settlement order written in a word-processing program.

Electronic mail is an increasingly widespread communications tool capable of very quickly reaching one or more individuals or institutions. Information can be contained in the body of the message or in attached documents. In addition, most database, word-processing, and spreadsheet programs have the capability of attaching documents created by the software directly to an e-mail message.

Scanning and document retrieval software has the ability to capture images of documents filed with the court, including signatures and seals, and to retrieve those documents for printing or display on terminals. It is becoming increasingly common for database programs to be able to electronically link to document images, so that, for example, all papers that have been filed regarding a case can be retrieved while reviewing case-processing information contained in the database. In addition, scanned documents can be attached to e-mail messages.

Facsimile transmission software is available that will permit sending documents or reports that have been produced by database, word-processing, or spreadsheet software to recipients who have fax machines or computer-based fax software. In addition, scanned documents can be distributed using facsimile transmission software.

Word recognition software permits an individual to enter information into software applications by using a microphone rather than a keyboard. This software is useful for judges and staff to dictate information into a word-processing document, and for the public in responding to automated information sites.

Touch-screen software can greatly simplify accessing information from automated systems. It is typically applied to programs that present structured responses to a set of structured questions.

EQUIPMENT AND CONNECTIVITY OPTIONS

There is an ever-increasing variety of equipment available to the courts that can operate independently or be connected in a wide array of options. Manufacturers are constantly developing equipment that has increased capacity and speed, as well as new items. It is difficult to keep current with developments in this area. Consequently, this section of the article will endeavor to be as generic as possible.

Telephones, copiers, and fax machines have been part of the office and the court for so many years that we tend to overlook them as components of court technology. They should not be overlooked as vital components of court automation. As mentioned previously, telephone conferencing can be a vital part of case processing. Fax machines can be used in every aspect of communication involving case documents, including initiating cases, filing motions and supporting papers, scheduling orders and notices, and making interim and final orders. The use of copiers is obvious, and their utility can be greatly increased by placing them throughout the court facility. For example, placing a copying machine in a courtroom can increase productivity by providing instant copies of orders, documents in evidence, and other case-processing information.

The most immediately recognizable configuration of automation equipment is a desktop computer with an attached printer. Although relatively simple, desktops have become sufficiently powerful to run every type of application software performing every type of function listed above. To have e-mail, a modem would be necessary to connect to the internet. To be able to use document imaging, a scanner would be required. There are multipurpose devices that incorporate printing, faxing, scanning, and copying in one machine.

Laptop computers have the same basic capacities as desktop computers. Their advantage is that they are portable, capable of, for example, being carried from a judge's chambers into a courtroom. Their disadvantage is that they are more expensive and less powerful than desktop computers for the same price.

The next step in automated systems is the local area network (LAN). A LAN connects two or more computers and allows users to share data, application software, and devices such as printers, scanners, and modems. In addition, a LAN can have an internal e-mail system installed, as well as a system of equipment and software to establish an intranet for court judges and staff to have access to information and applications within their own environment. A LAN can also have a gateway to allow users to access the Internet without having individual modems for each computer.

A wide area network (WAN) is a LAN with users separated by distances further than can be reached

by signals generated by LAN equipment. Routers or switches that convert and send signals over telephone lines are required. The features of a WAN are essentially those of a LAN, but extended beyond the reach of a LAN.

Kiosks are capable of displaying information about the courts in general, court schedules, forms to use and print out, and receiving information from the public, including court filings, pleas, and payments of fees and fines. The information provided and functions performed by these devices can be updated and modified from direct manipulation of the application software on the kiosks or "remotely" from court staff using their computers connected to a LAN or WAN. Status boards or running displays, similar to those found in airports, can also provide information to the public, and their messages can be modified from local devices or LAN- or WAN-based computers.

E-mail systems can also feed pagers, cell phones, and personal digital assistants via wireless Internet connections or through other communications systems. It is no longer necessary to be physically connected to receive information about courts, the status of a case, or a change in schedules.

DESIGNING, IMPLEMENTING, AND MAINTAINING SYSTEMS

The most challenging aspect of applying technology and information systems to the courts is determining the needs of the courts and establishing the appropriate mix of functions, software applications, and equipment and connectivity options. The most critical step in designing a system is to obtain a project sponsor for developing or improving an information system. A project sponsor should be placed high enough within the organization to be recognized as having the necessary authority to cause sufficient resources to be devoted to the effort to enable it to succeed. The project sponsor will change radically depending upon the size of the court system and the size of the project. It could range from a chief justice of a supreme court in a statewide system, to a magistrate in a rural town overseeing a two-person staff.

Once the project sponsor has been established, the court must carefully examine its resources to determine if it can reasonably design, implement, and maintain an information system. Larger court systems may well have sufficient in-house staff to undertake such an effort, while others may need consultant services. Even the largest organizations may contract with outside firms to provide guidance, support, and technical advice. Consultant services may be available from local and state governmental agencies, national organizations such as the National Center for State Courts,

and the private sector. For smaller courts, the state court administrator's office can frequently provide services directly or offer guidance in selecting consultants and vendors.

An important step in starting a court automation project is to become familiar with and adhere to basic principles of project management. There are a number of different methodologies and tools available, ranging from relatively simple to highly developed. The scope of the project will help to suggest the level of project management sophistication required to bring a project in on time and within budget.^[4]

Another critical step in designing information systems is to define the scope of the project. A careful definition of what functions are to be automated in which courts or offices, with a budget and timetable, and adherence to the scope definition, will help increase the probability of success. The scope should be committed to writing and signed by the project sponsor and key staff involved with the project. Any change to the scope of the project should be expressly addressed, complete with the additional time and expense required to include the change in the scope. Special attention should be given to insure that only those changes necessary to the success of the project are approved. Many requested changes can be dealt with at a later time, as a separate project.

In many instances, the court may find that it is not possible to incorporate all functions or courts in a current project. It is important that statements of scope that incorporate less than a total design include the possibility of expansion to other functions or courts in the future.

It is important that the judges and staff define the scope. They are the business experts and know what court functions would benefit the most from automated systems. While internal information technology analysts or consultants may offer advice as to the structure of the scope statement, or the time and cost implications of elements of the scope statement, the project sponsor, judges, and staff should have control of the final statement of scope.

Systems designs can range from a simple single desktop computer running a word-processing program to automate repetitive documents, to completely integrated data, voice, and imaging systems incorporating virtually all court functions with thousands of users spread over thousands of miles. Regardless of the size of the system, as with the preparation of the scope of the project, judges and court staff must be involved with the design of the system. This may require a significant amount of time, depending upon the scope of the project. It will also require a joint effort with in-house and consultant technology staff. Court staff and technology staff must take the time to insure that the needs of the court are completely understood, while

the cost and maintenance implications are clear and unambiguous. The final design, backed up by test runs of software and hardware, should be reviewed and approved by the project sponsor. Finally, it is imperative that all aspects of the design be well documented. The documentation provides the basis for corrections, modifications, and support in the future.

Implementing a new or modified system requires careful planning and patience. The system may require physical modifications, new equipment, and furniture. A certain amount of disruption of ongoing operations is inevitable, so careful coordination of court schedules, vacation plans, and backup will minimize problems. Arrangements must be made for training and intensive support after installation. Finally, last minute adjustments to the system may be required as a result of using the system in actual operations.

The assets required to maintain the system once it is installed should be included in the design phase. The need to modify the system for new requirements and correct problems will always occur. Providing for adequate maintenance is imperative, if the court's investment and operations are to be protected. Maintenance services can be provided by in-house staff or through contracts. The utility of solid systems documentation will become readily apparent at this time. However, the systems documentation must be maintained and updated by the group maintaining the system, or future efforts will be severely hampered.

CAUTIONS

One of the overriding concerns in any automated system is security. This is particularly important in court system records. Statutes and case law establish strict security requirements. Every effort should be made to incorporate appropriate security measures in the scope and design phase of a project. Judges and staff should ensure that the technical staff understand security requirements, and that they are satisfied with the security measures being described to them.

A second cautionary note is not to simply automate existing manual systems. The design phase of an automated system is an excellent opportunity to carefully review systems and procedures to insure that they are required by law, rule, good management practices, and are not redundant. Automating a confusing and unnecessarily complex system will only perpetuate past problems.

A third concern is the potential of e-mail to exacerbate the problem of ex parte communications. E-mail is widely used and can be an efficient method of communications with participants in a court case. Because it is so easily used, caution must be exercised by all parties to ensure that the contents of an e-mail message

from a litigant or an attorney does not cross the line into discussing a proceeding before the court without the full knowledge of all parties.

CONCLUSION

Technology and information systems have great applicability to the processes and procedures used by the courts. Virtually every function of the courts can benefit from carefully designed automated systems. Application software is widely available to develop automated systems in the courts. The capacity of equipment and interconnectivity continues to grow.

A well-known observer of court systems noted, “The great unknown for court leaders is the ultimate effect of telecommunications and computer technology on the courts As information systems integrate all types of information and the court starts to become a self-aware, self-correcting system, there will be fundamental changes in time-honored court practices.”^[5]

In order to take advantage of the promise of technology and information systems, courts must devote the resources necessary to design systems. In addition, judges and staff must devote a significant proportion of their time to ensure that the systems meet the needs of the courts.

Technology and information systems present a great opportunity and a great challenge to the courts. To keep abreast of ever-increasing workloads and to respond to the public’s demand that the courts become more efficient, courts will have to meet the challenge and take advantage of the opportunity.

REFERENCES

1. Webster, L.P. *Automating Court Systems*; National Center for State Courts: Williamsburg, VA, 1996.
2. Steelman, D.C.; Goerdt, J.A.; McMillan, J.E. *Caseflow Management: The Heart of Court Management in the New Millennium*; National Center for State Courts: Williamsburg, VA, 2000; xi pp.
3. Steelman, D.C.; Goerdt, J.A.; McMillan, J.E. *Caseflow Management: The Heart of Court Management in the New Millennium*; National Center for State Courts: Williamsburg, VA, 2000; 97–100.
4. Webster, L.P. *Automating Court Systems*; National Center for State Courts: Williamsburg, VA, 1996; 17–24. For a comprehensive overview of the subject, see *A Guide to the Project Management Body of Knowledge*; Project Management Institute: Sylva, North Carolina, 2000.
5. Tobin, R.W. *Creating the Judicial Branch: The Unfinished Reform*; National Center for State Courts: Williamsburg, VA, 1999; 205 pp.

Court Unification

James A. Gazell

*School of Public Administration and Urban Studies, San Diego State University,
San Diego, California, U.S.A.*

INTRODUCTION

Court unification is the degree to which federal and state judicial systems in America have established a single level of trial and appellate courts administered from the top by one institution or individual. It is a facet of centralization in government—an application to one of its branches: the courts. It proceeds along a continuum from autonomous individual judges and court administrators at one end to close hierarchical supervision of both at the other end. Synonymous terms include court centralization, court modernization, court bureaucratization, and court reform. It was a structural and administrative concomitant to the development of the nation's court systems.

PRE-EMINENCE

Court unification has been the pre-eminent characteristic of judicial administration since the early 1900s. The next comparable feature has been the emergence of judicial administration as a profession in the 1970s. Although court unification started at the federal level in 1891, its pre-eminence has been at the state level, beginning with New Jersey in 1948, as a result of a successful campaign, led by the Chief Justice of its highest court, Arthur T. Vanderbilt. Since then, most other states followed in greatly varying degrees.

ORIGIN

In 1906, legal scholar Roscoe Pound credited the origin of court unification to the experience of the United Kingdom in the 1870s, which took this action with respect to its national court system. It eliminated levels of trial and appellate courts that had jurisdiction that overlapped, work that was duplicative and wasteful of judicial personnel, and operations that spawned long delays in the resolution of cases. He agreed with the popular aphorism that justice delayed was justice denied. Pound wanted America's state court systems (many of which encountered the same problems) to undergo an overhaul and adapt the new British model to their own circumstances. He thought that these

systems would become more effective if they compressed their several levels of trial courts into one or two and established only one appellate court, standing alone or as a branch of a single trial court with comprehensive original jurisdiction. Presiding over such a structure would be one judge, empowered to make court-related rules and to assign other judges from one division of a court to another, equalizing the distribution of caseloads.

Furthermore, Pound's advocacy was one part of the Progressive Era in American history (1900–1914), which sought in part to strengthen all branches of government at the federal, state, and local levels. Pound had looked to the United Kingdom for a model suitable to fortify one branch of government: the judiciary. Progressives like him lived at a time when the business world had become increasingly centralized and bureaucratic, as large corporations, monopolies, oligopolies, and trusts emerged. Their size made them effective but also, to progressives, dangerous to the public interest because of restrictions on production, exploitation of labor, unsafe products, and excessive prices. However, they were too powerful to be broken up as populists wanted. But, government might be strengthened enough for their effective regulation in the public interest. Only government had the potential to become a genuine counterweight to the growing concentration of corporate power and its increasing domination of the nation.

Effective regulation also necessitated a degree of government centralization: the creation of honest and competent government machines to fight corporate machines. Government had to be run more like a business to be more effective, as a future president, Woodrow Wilson, once wrote. The famous sociologist Max Weber saw such machines as forms of bureaucracy, which he viewed as the most effective model of organization, the one best way of organizing. In a complementary vein, management writer Frederick Taylor argued that, within bureaucracy, there was one best way of doing and supervising work.

A degree of countervailing centralization began at the level of the federal government, with the enactment of antitrust laws, the formation of independent regulatory commissions, and the strengthening of administrative capacity of the executive branch through

expansion of a civil service system, budgetary control, and the development of the presidential staff. Government centralization slowly spread horizontally and vertically. It first extended to the legislative and judicial branches of the federal government, which sought to manage their own affairs. They sought to emulate the executive branch and to avoid becoming dependent on it. They hoped to function as a bona fide coordinate branch of the federal government, roughly equal in power. Otherwise, government by separation of powers and checks and balances would have remained mostly a concept in constitutional theory, not an operational reality. The federal legislative response largely took the form of strengthening its oversight of the federal executive branch, through imposing a degree of due process on its operations through administrative procedure statutes, expanding its structure of committees and their staffing and developing its ability to provide its own alternatives to executive budgets.

FEDERAL COURT UNIFICATION

The federal judicial response mostly took the form of persuading the other two branches to divest themselves of their traditional but increasingly onerous role of managing numerous aspects of federal judicial operations and to permit the federal judicial branch to handle its own finances, personnel, information, and rules to virtually the maximum extent permitted by the U.S. Constitution. Between 1891 and 1984, a series of notable structures came into existence. Some ran counter to court unification, but others fostered it. On balance, they brought a degree of centralization to the administration (but not the structure) of the federal judiciary.

Three structural developments diluted such unification:

1. In 1891, a second level of federal appellate courts—circuit courts of appeal—was established to ease the workload of the Supreme Court of the United States, generated by the population growth, urbanization, and industrialization of the nation. Furthermore, in 1925 the Supreme Court was granted extensive control over its docket, which in 1988 was made virtually complete.
2. In 1968, a second level of federal trial courts, the U.S. Magistrates, was also established. These courts were to decide minor criminal and civil cases and thereby enable the original level of trial courts, the federal district courts, to focus on more complex cases.
3. Various specialized appellate and trial courts were created to relieve circuit and district courts

of some of their technical and complex cases but at the cost of structural fragmentation. These included the United States Tax Court in 1942, the United S. Court of Military Appeals in 1951, the United States Court of International Trade in 1980, and the United States Court of Appeals for the Federal Circuit (for bankruptcy challenges), the United States Court of Federal Claims, and the United States Tax Court, all in 1982.

However, other developments more than offset such movements from unification. In 1922, the Conference of Senior Circuit Judges (which became the Judicial Conference of the United States in 1948) was established to permit the federal courts to make recommendations to Congress as to what the rules governing federal judicial operations should be. Congress transferred its judicial rule-making authority to this body in stages: in 1938, for civil procedure; in 1946, for criminal procedure; in 1968, for appellate rules; and in 1975, for rules of evidence. These transfers virtually ended legislative domination of the federal judiciary. But in 1984, Congress backtracked by asserting a degree of authority over sentencing, a traditional judicial function. It established the United States Sentencing Commission and entrusted it with formulating strict requirements for the courts. In 2005, the Supreme Court of the United States turned the sentencing requirements into advisory guidelines.

In 1939, movement toward federal court unification accelerated. The Administrative Office of the United States Courts was set up, ending the actual dependence of the federal courts on the executive branch. This office allowed the federal courts to gather statistics, to prepare a judicial-branch budget, and to hire their own staffs. Before this office came into existence, these matters were handled by a federal executive branch agency, the Department of Justice, also the chief litigant in the federal courts. Moreover, a council for each of the federal judicial circuits was established to manage its own workload and judges. Thus, the circuit council, not the Supreme Court of the United States, became the administrative model for running the federal judiciary and served to limit the degree of centralization that took place. In addition, the circuit conference was also created to enable judges within each circuit to meet periodically and keep abreast of legal and administrative developments.

Other unifying developments deserve mention. In 1967, the Federal Judicial Center was set up to become the research, development, and educational arm of the national court system. In 1971, the office of circuit executive was created. Each circuit chose a professional court administrator to run its operations and to allow judges to spend more time discharging

their primary duty: adjudication. In 1972, the office of administrative assistant to the Chief Justice of the United States was set up. It was the functional equivalent of the circuit executive. In 1974, the office of district court executive became a reality for the nation's federal trial courts.

STATE COURT UNIFICATION

After spreading horizontally at the federal level, the quest for government centralization moved vertically after World War II, seeping down to the executive, judicial, and legislative branches of state and local governments. One reason remained the same: the widely perceived need to establish counterweights to centralization in the business world. The state executive and legislative branches roughly emulated their federal counterparts. The state courts, like their federal counterparts, sought to increase their administrative capacity to perform functions entrusted to the other two branches, to end their dependence on them, and to become a genuine coordinate branch of state government. However, unlike the federal courts, the state courts favored a degree of centralization to escape a second broad source of dependence—reliance on local governments, which handled fiscal and personnel matters affecting the courts. In practice, state courts were often appendages not only of the executive and legislative branches of state governments but also of the corresponding branches of county and city governments. To end such dependencies, state courts established a set of structures roughly parallel to those in the federal judiciary. But, overall, the states went beyond the federal government in pursuit of court unification, because, as legal historian Lawrence Friedman suggested, their judiciaries were more atomistic to begin with and thus provided more room for such change.

States replaced their multiple trial courts of overlapping jurisdiction with one level or two levels of trial courts. In some states, there was to be only one place where all criminal and civil cases were to be filed—in short, a forum with comprehensive, original jurisdiction. In other states, there were to be two levels of trial courts modeled after the federal judiciary: a general trial court to handle major cases and courts of limited jurisdiction to handle minor matters. In addition, 39 states have followed the federal model by establishing an intermediate appellate level between the trial courts and the highest court.

All states established a judicial council, analogous to the Judicial Conference of the United States. It is the one body that makes policies applicable to an entire court system and reduces (but usually does not eliminate) the control of individual courts over

their own operations. This structure promulgates rules regarding various aspects of court administration: finances, human resources, assignment of judges, jury administration, education and training of judges and court staff members, and management information systems. The judicial council selects a state court administrator, comparable to the Director of the Administrative Office of the United States Courts (one official who supervises a staff who carries out the council's policies and who serves at its pleasure). This administrator sometimes exercises close supervision over the operations of individual courts but generally leaves them with discretion within broad boundaries, to handle their own business through appellate and trial court administrators. In a highly unified system, such officials are appointed by the state court administrator, subject to the approval of the judicial council. More characteristically, individual courts are permitted to make the choice within broad guidelines. Thus, judges of the individual courts or their presiding judge select their own court administrator as well as the deputies, assistants, division heads, clerks, reporters, secretaries, analysts, planners, and security officers—especially since the 9/11 tragedy and the independent rise of threats to the safety of judges.

The state court administrator oversees the performance of numerous judicial functions, none more important than the budgetary process. The work of this official often involves the use of a unitary judicial budget, one budget formulated for all the courts in a state, subject to legislative review, as opposed to the authority of each court to devise its own budget, subject to the same limitation. The unitary judicial budget is based on the executive budgeting model, which originated in the Progressive Era and served as a method for advancing the degree of administrative unity in federal, state, and local executive branches. Previously, individual executive branch departments, like individual courts, had prepared their own budgets and submitted them to the appropriate legislative bodies for disposition. Furthermore, unitary judicial budgeting has often gone hand in hand with still another unifying concept, state court financing. In this, all state courts submit all revenues raised through their own actions (fines, fees, and forfeitures) to one place (the state general fund) and receive their funding through appropriations from one source: the state legislature. State unitary judicial budgeting usually accompanies state court financing, but the converse is not always true. For instance, California has state court financing but not unitary judicial budgeting. The federal judiciary has both.

The state court administrator, like the Director of the Administrative Office of the United States Courts, oversees another aspect of court unification: a judicial civil service system. All court employees are under the

supervision of one system, which hires, retains, promotes, transfers, demotes, fires, and retires them. Thus, courts gain control over all the people who perform judicially related duties. Previously, some employees who worked in the courts were members of the executive branch of governments. In state trial courts, for example, deputy sheriffs provided security but were county executive branch personnel. They were comparable to deputy marshalls, who performed the same function in federal courts but worked for an executive branch agency. Other court employees worked within an overall state, county, or city civil service system, which also included legislative and executive branch employees.

The state court administrator follows the federal counterpart with regard to other characteristics of court unification. First, this official exercises the authority to assign judges and court staff members across and within jurisdictions. The administrator can more effectively utilize such employees by shifting them temporarily to those jurisdictions where they are most needed—for instance, from courts with light caseloads (usually rural trial courts) to those with heavy burdens (mostly urban trial courts). Second, this official oversees the pre- and in-service education of judges and other court employees, typically offered in one place, a subdivision of the state court administrator's office. Third, the same official often supervises an integrated judicial management information system like that of the federal court system. This form of computerization links courts at the same level, courts at different levels, and all courts to the state court administrator's office. Information can be readily disseminated throughout the system. If, for instance, a court where defendants are appearing on one charge can find out where they are also wanted by another court, such as for violating bail or failing to pay fines or fees. Moreover, states have taken this form of unification a step further by moving toward integrating information systems among courts and between courts and other state and local justice system agencies, such as police departments, prosecutors' offices, and probation and parole offices. The federal courts are proceeding in a comparable direction.

State court unification also encompasses two other features, neither of which exists at the federal level. One is the establishment of a judicial performance commission, which represents one place where all allegations of judicial misconduct can be heard. Its decisions also exert a unifying impact on the behavior of state judges, subjecting them to one set of expectations and one range of sanctions from reprimand to removal. By contrast, at the federal level, councils for each of the circuits may discipline judges by steps short of removal, which requires impeachment and conviction. The other is merit judicial selection. In pure

form, it involves the establishment of one body—a judicial selection (or qualifications) commission, which examines the backgrounds of all possible candidates for state judicial vacancies at all levels and compiles a list of what it regards as qualified individuals. All candidates must be certified by the commission. When vacancies occur, the governor is required to fill them by selecting someone only from the certified list. In one variation, the process is reversed. A governor is allowed to nominate candidates, contingent upon subsequent (rather than prior) commission approval before they can take office. The commission is not required to compile a list of qualified candidates. Instead, it handles nominees on a case-by-case basis, approving or rejecting them as it deems appropriate. In another variation, a judicial election is added and held within several years after a candidate takes office, regardless of whether there was prior or subsequent commission approval.

DOUBT

The spread of court unification at the federal and state levels has taken place despite much doubt. One source is rooted in the entrenched political sway of localism. County and city governments, for example, do not want to relinquish their budgetary oversight and concomitant financial control over trial courts in their jurisdictions. Hence, they resist unitary judicial budgeting and state court financing. They also do not want to give up control over court personnel. Thus, they fight efforts to replace elected court clerks with judicially appointed court clerks or court administrators. And, they do not want to entirely let go of their responsibility for court security. Furthermore, justices of the peace do not want their offices to be eliminated and absorbed into a unified state court system. Typically, these officials are unsalaried, lack a legal education and a license, and derive their compensation from finding defendants guilty of minor offenses and fining them. They are often so politically well organized that their position is secured in the state constitution and, therefore, cannot be abolished by statute.

Another source of doubt is more intellectual than political. Some acknowledge that, for decades, the trial courts in most states were too decentralized, too subservient to local culture and pressures, and too individualistic. They were virtually unsupervised by the highest state court and varied greatly in their performance. They needed at least a degree of court unification, although exactly what that entailed should depend on the peculiar milieu of each state or jurisdictions within it. But, to such skeptics, court unification cannot be treated as a one-size-fits-all type of remedy for judicial dysfunctions across and within states.

Dogmatic fervor allegedly accompanies court unification and ignores the particular historical legal, political, economic, social, and cultural factors. A high degree of court unification may fit some environments but not others. Advocates of court unification often pursue their goal of a structurally simplified and centrally administered system despite the lack of contextual knowledge. Consequently, there is no one best way of organizing and managing court systems, regardless of the specific local circumstances. Thus, it is therefore as doctrinaire to view court unification like a law of nature as it would be to espouse decentralization under all circumstances.

CONCLUSION

This article has provided a brief description and analysis of six aspects of court unification: definition, preeminence, origin, federal courts, state courts, and doubt. However, most writings on the subject discuss it as if it were fundamentally a matter of structure and administration. The desired outcome is an increase in the number of case dispositions, a quicker flow of filings through trial and appellate assembly lines of judicial systems. The result is usually considered from an empirical rather than a normative perspective. There is insufficient consideration of whether the eventual upshot will be a genuinely more just court system, greater public satisfaction, and more widespread respect for the nation's potpourri of court systems. These are presumed to accrue in proportion to an increase in court unification as automatically as day follows night. Whether a higher level of justice has accrued remains uncertain.

BIBLIOGRAPHY

1. Abraham, H.J. *The Judicial Process*; Oxford University Press: New York, NY, 1998.
2. Court Statistics Project. In *State Caseload Statistics, 2004*; National Center for State Courts: Williamsburg, VA, 2005.
3. Fish, P.G. *The Politics of Federal Judicial Administration*; Princeton University Press: Princeton, NJ, 1973.
4. Friedman, L.M. *A History of American Law*, 3rd Ed.; Touchstone Book: New York, 2005.

5. Friesen, E.C., Jr.; Gallas, E.C.; Gallas, N.M. *Managing the Courts*; The Bobbs-Merrill Company, Inc.: Indianapolis, IN, 1971.
6. Gallas, G. The conventional wisdom of state court administration: a critical assessment of an alternative approach. *Justice Syst. J.* **1976**, 2 (1), 35–55.
7. Gerth, H.H., Mills, C.W., Eds.; *From Max Weber: Essays in Sociology*; Oxford University Press: New York, NY, 1946.
8. Hays, S.W., Graham, C.B., Eds.; *Handbook of Court Administration and Management*; Marcel Dekker, Inc.: New York, NY, 1993.
9. Hofstadter, R. *The Age of Reform*; Vintage Books: New York, NY, 1955.
10. Neubauer, D.W. *America's Courts and the Criminal Justice System*, 8th Ed.; Wadsworth: Belmont, CA, 2005.
11. Pound, R. The causes of popular dissatisfaction with the administration of justice. *Rep. Am. Bar Assoc.* **1906**, 29 (Part 1), 395–417.
12. Rottman, D.B.; Strickland, S.M. *State Court Organization 2004*; United States Department of Justice (Office of Justice Programs, Bureau of Justice Statistics): Washington, DC, 2006.
13. Saari, D.J. *American Court Management*; Quorum Books: Westport, CT, 1982.
14. Steelman, D.C.; Goerdt, J.A.; McMillan, J. *Caseflow Management*; National Center for State Courts: Williamsburg, VA, 2000.
15. Taylor, F.W. *The Principles of Scientific Management*; W. W. Norton and Company, Inc.: New York, NY, 1967, originally published in 1911.
16. Tobin, R.W. *Creating the Judicial Branch: The Unfinished Reform*, 2nd Ed.; National Center for State Courts: Williamsburg, VA, 2004.
17. *United States v. Booker*, 543 U.S. 220, 2005.
18. Vanderbilt, A.T., Ed.; *Minimum Standards of Judicial Administration*; Law Center of New York University: New York, NY, 1949.
19. Wilson, W. The study of administration. *Polit. Sci. Q.* **1887**, 2 (2), 199–222.

Websites

Bureau of Justice Statistics (United States Department of Justice—Office of Justice Programs). <<http://www.ojp.usdoj.gov/bjs/>>.
 Federal Courts. <<http://www.uscourts.gov/>>.
 National Center for State Courts. <<http://www.ncsconline.org/>>.

Courts and Policy in the United States

Robert K. Christensen

*Political Science Department, University of North Carolina at Charlotte,
Charlotte, North Carolina, U.S.A.*

Jiang He

*Public Policy Program, University of North Carolina at Charlotte,
Charlotte, North Carolina, U.S.A.*

INTRODUCTION

The formation, implementation, and evaluation of public policy are the responsibility of many actors. By constitutional design, U.S. public policy is articulated and influenced across the respective powers of divisions of democratic government. The judiciary constitutes one of those divisions of power. While few would argue that the powers of the judiciary facilitate judges' unilateral control of public policy, a debate exists as to just how influential the judiciary's voice is if we conceptualize governance as a dialogue among democratic institutions such as legislatures, executives, and judiciaries.^[1,2] For example, some observe that among the three branches of government, the judiciary is the least dangerous branch, and "next to nothing" in its lack of the purse (or legislative power) and sword (or executive power).^[2,3] Others view courts as more influential, dominating executive and legislative actors in particular policy areas.^[4–6]

We discuss these competing perspectives in greater detail below. Our general approach in this piece is that the power of courts over public policy in the United States is context dependent—both temporally and in relation to the other institutions of government.^[7] In other words, we consider when, and under what circumstances, courts influence policy.

In exploring these dynamics, we will first introduce fundamentals of the judicial impact debate and relevant features of the U.S. court system. We then discuss judicial discretion, or how courts "choose" and influence public policy. Lastly, we will look at the interaction between courts and other political branches, to evaluate the relative policy potency of the judicial branch.

JUDICIAL IMPACT DEBATE

Observers have long examined the impact—both in theory and in practice—of the judiciary on public policy. Hamilton's writings draw upon Montesquieu's observations that "of the three powers above mentioned,

the judiciary is in some measure next to nothing" (in Book 11, Chapter 6, Vol. 8, No. 78 of Ref.^[3]). Hamilton's reasoning is relatively straightforward:

The Executive not only dispenses the honors, but holds the sword of the community. The legislature not only commands the purse, but prescribes the rules by which the duties and rights of every citizen are to be regulated. The judiciary, on the contrary, has no influence over either the sword or the purse (in No. 78 of Ref.^[8]).

Such a view of the judiciary underscores a judiciary whose relative social impact is innocuous. Other scholars observe that "it is sometimes said that courts implement policies made by other branches of government. But for many issues . . . the opposite is true: the courts make policy, and other political institutions respond to their lead" (on page 193 of Ref.^[9]). To further explore the territory between these contrasting perspectives, we offer a discussion of the structure of courts within the United States.

U.S. COURT SYSTEM AND PUBLIC POLICY

The U.S. court system is hierarchical. One system of hierarchy exists at the state level, where each state in the United States has a system of civil and criminal courts. Generally, cases are brought before trial courts, where juries and judges make findings based on the fact and law. Appellate courts have the ability to review trial courts (although usually only the application of law from lower courts' decisions). Each state also has its own court of last resort, most commonly called a supreme court, which serves to review appellate judicial decisions. Decisions of the state courts of last resort can also be appealed to the U.S. Supreme Court.

Parallel to and sometimes overlapping state court jurisdiction, federal courts also handle decisions with potential policy implications. Across the country, there are 94 U.S. judicial districts, which are organized into 13 regional circuits, with a U.S. court of appeals in each circuit. A federal appellate court usually hears

appeals from the district courts located within its circuit, as well as appeals from decisions of federal administrative agencies. These courts of appeals are subject to the review of the U.S. Supreme Court.

The structure of the judicial system in the United States raises at least two considerations regarding courts and public policy. On the one hand, the hierarchical nature of the court system limits unilateral policy making on the part of any single judge or court. On the other hand, while the finality of the federal Supreme Court's decisions draws much national publicity, a vast majority of judicial decision making is executed at the state level and goes without appellate (state or federal) review. Partially a function of sheer numbers, only a small fraction of cases are filed (recently around 7500), and far fewer are heard (recently less than 100 per year) with the U.S. Supreme Court, when compared to the millions of cases filed annually in state and federal courts.

JUDICIAL DISCRETION

In addition to understanding court structure, judicial impact on public policy is also shaped by the doctrine of justiciability. This means that judges do not possess a roving commission to pick and choose actual, or forward hypothetical, disputes. Judges are confined to actual disputes that are brought before them, within their jurisdiction.

These constraints notwithstanding, the notion that judges can act as policy proponents is widely argued among scholars.^[10] Many observe that court decisions actively advance policy goals and that individual justice can influence public policy by selecting among justiciable cases legitimately before them. This is particularly the case with the U.S. Supreme Court and state courts of last resort (in those states where there are intermediate appellate courts), where appellate review is not mandatory. Even in most appellate courts, where review is mandatory, judges have developed creative ways of managing their workload.^[9,11] In other words, it may seem that judges are passive in choosing public policy areas, but many judges are able to pick cases that they are particularly interested in and ignore the others. This ability gives judges varying levels of discretion in selecting and advancing public policy areas. Judges can put items important to them on the agenda and use court rulings as tools to influence public policy.

While the life tenure that many judges enjoy is meant to provide insulation from outside political expediencies when ruling on cases, justices do not operate independently. In practice, judges face pressures presented by the legislative and executive branches, interest groups, and general public opinion.^[9,12] In court, these pressures have been demonstrated in

the form of pressure to defer to government actor litigants^[13,14] or to be sensitive to briefs filed by parties^[15]—especially government parties—not directly involved (*amicus* briefs). From this perspective, public policy outcomes from court decisions represent a sensitivity to, if not a reflection of, majoritarian views. This is discussed in greater detail in the following section.

INTERACTION WITH OTHER POLICY MAKERS

There are many contacts between Congress and the Supreme Court in the U.S. political system. Many federal court cases deal with whether federal statutes created by Congress violate the Constitution.^[16] When a court decides that existing statutes are unconstitutional, Congress may rewrite the statute or enact a new statute to advance the same goal as the old statute within the limitations established by the Court's decision.^[16]

Even if the new/revised statute passes, judicial bodies often have the chance to interpret the law, based on their understanding of Constitution. For example, in successive exchanges with legislative and administrative bodies, courts have designated very specific methods to achieve racial integration within public schools^[6] or have, over time, elaborated increased liability for public employee misconduct.^[17] As some scholars have observed, the legislative branch's ability to check iterative, judicial decision making is limited, unless the high hurdle of amending the Constitution amendment is overcome.^[18] Thus, while Congress holds the power of the "purse," it may be difficult to discern whether courts or legislatures are dominant in particular policy areas. Both actors have their ability to shape policy processes and outcomes.

Some research focusing on the Supreme Court demonstrates that these interactions and abilities can be strategic. Justices can predict responses from other branches before ruling on a particular case. If judges anticipate being overturned by Congress, they may choose a less controversial and more incremental advancement of their policy preferences.^[16] Congress, in turn, has relied on the budgetary process to impact judicial policy making. If Congress is in conflict with the Supreme Court, Congress may reduce the budget for the judicial branch. These budgetary signals from Congress have occurred historically, and past judges—such as Chief Justice Warren—have criticized Congress for taking these steps.^[19]

The judiciary also interacts with executive and administrative actors in the policy process. One obvious area of this interaction occurs because the chief executive appoints and influences the nomination of federal judges. The president can select judges who

may share certain political views. When the policy goals of the chief executives and a majority of judges deciding a particular case are the same, conflict is minimized. However, when policy priorities are in conflict, courts can detract from executive policy influence. A recent example still in contention is the Court's removal of presidential line-item veto power in 1998, which President Bush asked, unsuccessfully, to be restored by Congress in 2006.^[20]

Observers also note that courts frequently exercise public policy influence over administrative agencies. Courts might demonstrate this influence in multiple ways, including reformations of administrative law, articulation/refinement of constitutional duties,^[21] and modifications to public official immunity.^[17] The public administration literature has been particularly helpful in illustrating the policy influence of courts in the administrative arena. For example, O'Leary documented that courts very often take first seat in policy control because of conflicting political and social contexts.^[4] She observed that the EPA was often more influenced by federal court decisions than by political directives. While she noted a range of judicial behavior, from passive to aggressive, O'Leary discovered that from both the agency and the individual civil servant level, compliance with court decisions had become priority (on pages 561–562 of Ref.^[22]). Spriggs explains that “it is unlikely that agencies can expect Congress or the president to protect their overturned policies from the Court” (on page 1143 of Ref.^[23]). Research by O'Leary and Wise focusing on school desegregation argues that in the partnership of judges, legislators, and administrators, judges are the most potent of the triumvirate.^[5,6]

CONCLUSION

As the final arbiter in the judicial branch, the U.S. Supreme Court claims definitive authority over constitutional interpretation—the yardstick against which all public policy is ultimately measured. However, even Justice Felix Frankfurter echoed early phrases of the judicial impact debate: “The Court's authority—possessed of neither the purse nor the sword—ultimately rests on sustained public confidence in its moral sanction.”^[24] Courts certainly face institutional constraints. The legislative branch controls the budget to implement court policy, the executive branch controls policy administrative functions, and the public, moral sanction. Justices cannot effectively pursue their own policy goals in a vacuum.^[25] As an illustration of these constraints, Rosenberg argued that Court decisions, such as *Roe v. Wade* and *Brown v. Board of Education*, had marginal impact on social change relative to the greater political and social forces shaping public policy.^[26]

These considerations notwithstanding, judges influence public policy in many ways. In their often-repeated interactions with other branches, they can be strategic in predicating their policy decisions on anticipated responses of the political institutions.^[18] Judges also play a central role in our society that some argue is increasingly adversarial. Kagan^[27] suggests that with rise of “adversarial legalism” policy making is increasingly articulated in our court system, giving the judiciary increasing power in defining policy (see also Ref.^[28]).

REFERENCES

1. Fisher, L. *Constitutional Dialogues: Interpretation as Political Process*; Princeton University Press: Princeton, NJ, 1988; Vol. x, 306.
2. Bickel, A.M. *The Least Dangerous Branch: The Supreme Court at the Bar of Politics*; Indianapolis: Bobbs-Merrill, 1962; Vol. viii, 303.
3. Montesquieu, C.D. *Spirit of Laws*, 1914 Ed.; G. Bell & Sons: London, 1748.
4. O'Leary, R. *Environmental Change: Federal Courts and the EPA*; Temple University Press: Philadelphia, 1993.
5. O'Leary, R.; Wise, C.R. Public managers, judges, and legislators: redefining the “new partnership”. *Public Admin. Rev.* **1991**, 51 (4), 316–327.
6. Wise, C.R.; O'leary, R. Breaking up is hard to do: the dissolution of judicial supervision of public services. *Public Admin. Rev.* **2003**, 63 (2), 177–191.
7. Miller, M.C. *The View of the Courts from the Hill: Governance as Dialogue*; PS Pol. Sci. Polit. **2007**, 40 (1): 179.
8. Manis, J., Ed.; *The Federalist Papers*, Electronic Classics Series; Pennsylvania State University, 2001.
9. Van Horn, C.E.; Baumer, D.C.; Gormley, W.T. *Politics and Public Policy*, 3rd Ed.; CQ Press: Washington, DC, 2001; Vol. xiii, 366.
10. Epstein, L.; Knight, J. *The Choices Justices Make*; CQ Press: Washington, DC, 1998; Vol. xviii, 200.
11. Carp, R.A.; Stidham, R. *The Federal Courts*, 3rd Ed.; CQ Press: Washington, DC, 1998; Vol. xiii, 236.
12. Miller, M.C.; Barnes, J. *Making Policy, Making Law: An Interbranch Perspective*; Georgetown University Press: Washington, DC, 2004; Vol. xi, 244.
13. Tanenhaus, J.; Schick, M.; Muraskin, M.; Rosen, D. The Supreme Court's certiorari jurisdiction: cue theory. In *Judicial Decision Making*; Schubert, G., Ed.; Free Press: London, 1963; 111–132.
14. Cohen, L.R.; Spitzer, M.L. The government litigant advantage: implications for the law. *Florida State Univ. Law Rev.* **2000**, 28, 391–425.
15. Puro, S. The United States as amicus curiae. In *Courts, Law and Judicial Processes*; Ulmer, S.S., Ed.; Free Press: New York, 1981.
16. Baum, L.; Hausegger, L. The supreme court and Congress—reconsidering the relationship. In *Making*

- Policy, Making Law: An Interbranch Perspective*; Miller, M., Barnes, J., Eds.; Georgetown University Press: Washington, DC, 2004; 107–122.
17. Rosenbloom, D.H.; Carroll, J.D.; Carroll, J.D. *Constitutional Competence for Public Managers: Cases and Commentary*; F.E. Peacock Publishers: Itasca, IL, 2000; Vol. xviii, 221.
 18. Epstein, L.; Knight, J.; Martin, A. Constitutional interpretation from a strategic perspective. In *Making Policy, Making Law: An Interbranch Perspective*; Miller, M., Barnes, J., Eds.; Georgetown University Press: Washington, DC, 2004; 170–188.
 19. Toma, E. Congressional influence and the Supreme Court: the budget as a signaling device. *J. Legal Stud.* **1991**, 20 (1), 131–146.
 20. *Clinton v. City of New York*, in 524 U.S. 417, 1998.
 21. Melnick, R. Courts and agencies. In *Making Policy, Making Law: An Interbranch Perspective*; Miller, M., Barnes, J., Eds.; Georgetown University Press: Washington, DC, 2004; 89–104.
 22. O’Leary, R. The impact of federal court decisions on the policies and administration of the U.S. Environmental Protection Agency. *Admin. Law Rev.* **1990**, 41, 549–574.
 23. Spriggs, J.F. The Supreme Court and federal administrative agencies: a resource-based theory and analysis of judicial impact. *Am. J. Pol. Sci.* **1996**, 40 (4), 1122–1151.
 24. *Baker v. Carr*, in 369 U.S. 186, 1962.
 25. McCloskey, R.G.; Levinson, S. *The American Supreme Court. The Chicago History of American Civilization*, 4th Ed.; University of Chicago Press: Chicago, 2005; Vol. xv, 331.
 26. Rosenberg, G.N. *The Hollow Hope: Can Courts Bring About Social Change?* University of Chicago: Chicago, 1991; Vol. xii, 425.
 27. Kagan, R.A. *Adversarial Legalism: The American Way of Law*; Harvard University Press: Cambridge, MA, 2001; Vol. xii, 339.
 28. McCann, M.W. How the Supreme Court matters in American politics: new institutional perspectives. In *The Supreme Court in American Politics: New Institutional Interpretations*; Gillman, H., Clayton, C.W., Eds.; University Press of Kansas: Lawrence, KS, 1999; 63–97.

Criminal Justice Public Policy

Barbara Sims

*School of Public Affairs, Pennsylvania State University at Harrisburg,
Middletown, Pennsylvania, U.S.A.*

INTRODUCTION

Perhaps one of the most complex of all public institutions is the American criminal justice system. With its myriad of agencies, each with a distinct and critical role, this open-ended system projects a picture of “doing justice” as though it were a well-oiled machine. People come in contact with the system through the police, and a decision is made at the street level whether that individual will stay in the system. If an arrest is made, the offender becomes part of a process that is often referred to as “assembly line justice.”^[1] Developing and implementing policies up and down the justice assembly line is no easy task. One jolt to the system can send rippling effects throughout all of these agencies, leaving criminal justice administrators in the unenviable position of having to make some rather “rushed” decisions. Even when administrators know that change is coming, as opposed to unplanned changes that occur as the result of events outside the purview of criminal justice agencies (demographic changes, for example, or drastic downturns in the economy), navigating the waters of how best to make those changes, and where, is not without its problems.

As is the case with any public agency, criminal justice agencies must constantly struggle on two major fronts. First, they must contend with a public that has become increasingly more concerned about crime and wants something done about it. Second, they must walk through political mine fields when attempting any approach to the crime problem. The result is that more often than not, crime policies are based on knee-jerk reactions to the more celebrated cases (O.J. Simpson, Willie Horton, etc.) as opposed to careful projections about the consequences of actions taken. It is those cases, after all, that are in the public eye, and as such, are closely paid attention to by the people elected to serve them.

As argued by Hancock and Sharp,^[2] many different groups have an interest in shaping crime policies, as it has become one of the most “popular for legislative, executive, and judicial manipulation and debate.” Sometime in the decade of the 1980s, the American criminal justice system moved from one of due process and concern for the constitutional rights of individuals to a crime control model seemingly intent

on incarcerating more individuals and for longer periods of time. The idea that “criminals” could be rehabilitated and turned into law-abiding citizens gave way to punishment as retribution, or “just desserts.” Americans began to see crime as the result of a lenient criminal justice system as opposed to wayward youth in need of assistance and supervision.

With “tough on crime” the new mantra of both major political parties, complete with phrases such as “war on crime” and “war on drugs,” the 1990s ushered in a time of unprecedented incarceration rates. The U.S. prison population rose from 196,429 in 1970 to 1.9 million in 1998, with the incarceration rate rising from 96 per 100,000 in 1970 to 452 per 100,000 in 1998. Compared with other countries, the United States is clearly out of step, with Japan at 42 out of 100,000 and Canada at 111 per 100,000.^[3] The new policies directed at waging the war on crime, and that resulted in this exponential growth in the incarceration rate, are the subject of this article. As will be illustrated, there are other negative consequences associated with these new crime policies in addition to skyrocketing corrections budgets that most state departments of corrections cannot afford. Further, a great deal of research has produced convincing evidence of certain policies that could work to curb crime in the United States. Findings from that research are also discussed, and an argument is made for a more sensible approach to the crime problem.

PUBLIC POLICY AND CRIMINAL JUSTICE

Dye^[4] defined public policy as anything that government chooses to do or not to do. Government leaders in the United States began using the “war” rhetoric when talking about crime policies beginning with President Lyndon Johnson in 1965, followed by Richard Nixon in 1969, and in 1989, President George Bush declared a war on drugs.^[5] As a result, policy makers have chosen to react to America’s crime problem, as Walker^[5] argues, by implementing certain “crime control fads,” such as “selective incapacitation,” or “three strikes and you’re out” legislation. These policies are in one day and seemingly out the next, and what government has not chosen to do is

to examine with some clarity the extent to which they do, or do not, work.^[5]

Problems Associated with Defining the Extent and Nature of America's Crime Problem

According to Welsh and Harris,^[6] the policy process begins by clearly defining the problem at hand. In the criminal justice policy arena, the problem is crime in U.S. communities. Determining the extent of the problem, however, is not without its problems. There are three ways that are used to measure crime in the United States, none of which is problem free. Referred to as official data, the Uniform Crime Reports (UCR) consists of collected and disseminated crime statistics as reported to the F.B.I. by about 16,000 local, county, state, and federal law enforcement agencies. The major problem with UCR data is the fact that it contains only reports of crimes known to the police and for which an arrest was made. Data from the National Crime Victimization Survey (NCVS), an annual survey of 60,000 U.S. households, seeks to uncover the so-called "dark figure" of crime by identifying those crimes not known to the police. NCVS data is criticized primarily for undercounting certain crimes (sexual and domestic abuse, for example), and because it does not contain crimes associated with businesses or commercial operations. Finally, self-reported studies ask people to disclose their crime and delinquency and may fill the gap between UCR and NCVS data. Those types of offenses often missed by both, public order crimes, for example, or drug use, may be more accurately measured through self-report studies. The major problem with self-report studies, however, is the fact that even guaranteed confidentiality and complete anonymity, individuals may not fully disclose their crimes to the interviewer.

Some criminologists have sought to get around the problem of reliability and validity when it comes to measuring crime by comparing the findings from different sources. The most successful of these undertakings have come from those comparing UCR data with NCVS data. Blumstein et al.^[7] for example, have found that, despite the different methodologies associated with both, similarities exist, especially for robbery and burglary. For policy makers, an awareness of the problems associated with any crime data is an important first step in the policy process.

Selective Incapacitation: Identifying and Locking up the Career Criminal

A good example of a modern-day criminal justice policy is that of selective incapacitation, a term used to describe the process of identifying and incarcerating

so-called "career criminals." Conservatives believe that if those individuals responsible for most of the crimes, those repeat offenders for which the threat of punishment does not appear to work, can be weeded out and targeted for long prison terms, crime rates will decline. The notion of the career criminal was first examined through cohort studies, the most well-known and respected of which is that by Wolfgang et al. in 1972.^[8] The authors tracked 9945 boys born in the city of Philadelphia in 1945 through their 18th birthday in 1963, using data from police reports, school records, and children and youth services. The overall finding was that the majority of all crimes (52%) associated with this birth cohort were committed by only 6% of these individuals. Approximately 627 of the 9945 individuals were identified by Wolfgang and his colleagues^[8] as chronic or career criminals.

When Wolfgang et al.'s^[8] results were published, crime rates had been rising. According to Walker,^[5] the time was ripe for any program that promised to get tough with chronic, repeat offenders. Selective incapacitation seemed like the perfect solution. The problem, however, is how best to predict future offending. In a thorough examination of the "prediction problem," Walker^[5] concluded that even if the criminal justice system could accurately identify the career criminal, the next hurdle for criminal justice policy makers would be to determine just how much crime could actually be prevented by incarcerating them. He argues that the three ways often used to predict criminal behavior might not be successful. Those three ways are: 1) the actuarial method, such as insurance companies use when predicting driving behaviors by certain groups, the young driver, for example; 2) examining prior criminal history; and 3) extensive clinical evaluations by the use of trained experts.^[5]

Attempts at predicting future violent behavior

When researchers have sought to develop instruments that can be used to predict future criminal behavior, they have often worked with a sample of incarcerated individuals. Wenk et al.^[9] examined closely the records of 4146 youths held by the California Youth Authority using an instrument that included information about prior criminal record coupled with the youths' histories of violent behavior and substance abuse. In this identified group of juvenile delinquents, 104 went on to become violent recidivists.^[9] Through a series of psychological evaluations and interviews, Wenk et al.^[9] discovered that only about half of the 104 recidivists were identified correctly. The remainder of the group were false negatives, that is to say, they were not predicted to become violent recidivists, but ultimately did. Further, 404 people who were predicted to become

violent recidivists did not. Had the instrument been used with the original sample and had the goal been to selectively incarcerate those predicted to become violent recidivists, these “false positives” would have been facing incarceration for violent crimes they would never commit.

The Wenk et al.^[9] study and others like it suggest that selective incapacitation is a policy that cannot succeed unless better prediction instruments can be developed to get around the problem of false negatives and false positives. In addition, prediction is further complicated by disagreement among researchers about just how much crime these so-called “career criminals” commit. As Walker^[5] points out, “If the averages are low (as some suggest they are), we will not get that much payoff in terms of crime reduction.”

“Gross incapacitation” versus selective incapacitation

Selective incapacitation, predicting future criminal behavior, will not work to reduce crime in the United States. The evidence suggests that the courtroom work group (attorneys, judges, jurors, etc.) is just as accurate as are prediction instruments. This means that the overall goal of this particular criminal justice policy is not feasible. Public policy is a fluid process, however, and what has happened in criminal justice is the rapid move toward “gross incapacitation,” “locking up a lot of people and sending them to prison for long terms.”^[5] The goal of carefully selecting individuals who are likely to cause the most harm to society has fallen away.

In 1987, a report titled “Making Confinement Decisions,” published by the U.S. Government Printing Office and authored by Edwin Zedlewski, purported to illustrate the public value of gross incapacitation. Using “costs to society” as his starting point, Zedlewski^[10] calculated that every dollar spent on incarceration actually saves taxpayers \$17. Other researchers have pointed out the problems associated with Zedlewski’s formula upon which his estimates were based (for an excellent review of the subject, see Ref.^[11]). Primarily, the critical problem is with his estimate of the number of crimes committed a year by criminals. Career criminals specialists cannot agree on what that figure might be, but Zedlewski chose the higher value of 187 crimes per year per criminal offender, as opposed to the more often used figure of about 18 crimes a year. Yet, policy makers relied heavily on Zedlewski’s^[11] calculation and continued down the path of gross incapacitation. One of the best examples of such a policy is mandatory sentencing laws, more commonly referred to as “three-strikes and you’re out.”

Three-Strikes Legislation

Mandatory sentences are extremely popular with the public and with elected officials. This popularity comes from, as noted above, the media play of some criminal justice cases. When 12-year-old Polly Klass was kidnapped and murdered by a twice-convicted felon named Richard Davis, out on parole at the time of the murder, the citizens of California were outraged. This outrage led to California’s three-strikes law, soon followed by other states. Today, 26 states have some sort of mandatory sentencing statute. The major premise of these laws is that for the third convicted felony, offenders will receive a life sentence in a state prison.

The goal of mandatory sentencing

Just as selective incapacitation is a policy aimed at reducing crime in the United States, so is mandatory sentencing. A major step in the policy process is to ensure that the targeted population is clearly defined and that steps to reach that population are equally clear. Determinations as to the feasibility of the proposed policy are crucial prior to implementing major changes. As mentioned above, the courtroom work group works not unlike other street-level bureaucrats to put up roadblocks to the implementation of newly enacted crime policies. This use of discretion has been described by Lipsky^[12] as a more important deciding factor as to how policies will be implemented than is the power of the executive or legislative branches of government, the sources of many crime policies.

Through mandatory sentencing, states are able to send a message to the public that something is being done about crime. This message meets the goal, in policy terms, of engaging in symbolic politics.^[13] Once again, a public in fear of crime has been reassured that government is cracking down on felony recidivists. Interest groups, those rallying around the family of young Polly Klass for example, are able to influence state legislators. Through widespread media footage of proposed mandatory sentences, the political climate becomes ripe for such legislation, as indicated by a series of public opinion polls.

Based on an excellent analysis of the policy process as it applies to “three-strikes” legislation, Welsh and Harris^[6] came to the conclusion that this law was doomed to fail because of poor planning. They posed the question, “What evidence was used to demonstrate a need for change in U.S. crime policies dealing with felony offenders?”^[6] As was pointed out above, although the perception in society was that crime was rising, actually, it had been declining. Thus, instead of working to assure the public that crime was not on the rise and to educate the public about the negative consequences that sometimes come with poor planning,

politicians were “too eager to cater to the perceived public will, rather than documented problems.”^[6]

Implementing three-strikes legislation

Welsh and Harris^[6] pointed out that prior to the implementation of any policy, all parties involved in that implementation should be fully apprised of their responsibility. The policy process requires brainstorming about possible obstacles to implementation and how those obstacles will be overcome. It became clear almost immediately that mandatory sentencing laws would be met with resistance or simply ignored. Why? The answer can be found again in the manner in which employees in bureaucratic organizations apply discretion in the everyday job of doing the public's business.

Welsh and Harris^[6] described the plea bargaining process, for example, and the manner in which prosecutors circumvent the law by reducing the charges. Mandatory sentencing laws give the prosecutor a tool to dangle before offenders in an effort to get them to plead guilty. Police officers may engage in more selective arrest policies, and at the federal level, it has been shown that the law has been used sparsely since its inception in 1994.^[6] Walker^[5] pointed out that Tennessee, New Mexico, and Colorado have not used their three-strikes law at all, and that Wisconsin only used its law once in the first two years after it was enacted.

Evaluating three-strikes laws

In the policy process, it is important to determine to what extent the goals and expected outcomes are being met. Evaluation research provides public agencies with this critical information. With three-strikes laws scarcely being used as they were intended, however, evaluating the impact of them is tenuous at best. The finding that criminal justice personnel often ignore the law or somehow get around it is the type of information that comes from what is termed “process evaluation.” Unlike impact assessments that provide outcome or impact measures of the new program, process evaluation is designed to determine whether any new program was implemented the way it was intended. Clearly, the usurping of the law by the courtroom work group means that these types of sentencing structures are not being implemented the way they were intended.

First, mandatory sentences add a tremendous burden to an already-overworked criminal justice system. With more at stake, defendants who are in a district where mandatory sentencing has been fully implemented are demanding jury trials. In California, third-strike cases represent about 24% of all trials, raising

court costs for local counties. Too, the increase in the number of defendants waiting in jail for their trial dates has resulted in astronomical increases in local corrections budgets. In Los Angeles, the new three-strikes law cost the county an additional \$169 million in the first two years of its implementation.^[5]

Another important piece of evidence from attempts to evaluate mandatory sentences is that they probably do little to bring down crime in the United States. On one hand, and as pointed out above, it is difficult to argue that such laws are tailored toward identifying an individual's propensity to commit future crime.^[5] On the other hand, Walker^[5] argued that under such a policy, older offenders are likely to end up in three-strikes statistics at about the time they would have naturally aged out of the crime-prone years. This means that there are more and more older people being sent to prison, and costs associated with housing the elderly inmate are quite substantial, adding even more to correctional budgets.

Welsh and Harris^[6] summarized the projected effects of California's three-strikes law by stating that fully implementing this policy will:

- Triple California's prison population over the next 25 years.
- Cost an average of \$5.5 billion more each year for the next 25 years.
- Corrections could consume 18% of the state's spending by the year 2002.

PROMISING APPROACHES FOR CRIMINAL JUSTICE POLICY MAKERS

The above discussion about two modern-day crime control policies was used as a backdrop for demonstrating that not unlike any other public institution, policy making in criminal justice is often driven more by symbolic politics than sound policy analysis. The nation's “war on crime” and “war on drugs” rhetoric has created “get tough” approaches targeted toward the so-called “career criminal” and major drug dealers. Much of the literature suggests, however, that these policies have failed, by not being implemented correctly or by not doing much to reduce crime rates. Instead, these policies have produced a plethora of negative backlashes against an already over-extended system. This leaves many forward-thinking and innovative criminal justice administrators thinking seriously about what *can* work to curb serious crime in local communities.

Programs such as Pennsylvania Governor Mark Schweiker's “Weed and Seed” programs, a state-sponsored offshoot of the federal government's program by the same name, are beginning to stress

approaches that are founded in research. Weed and Seed programming calls for the “weeding out” of hot spots of crimes in specially targeted areas by local law enforcement, in conjunction, in some cases, with the State Police. “Seeding” programs then seek to involve community-based agencies in a joint effort to provide these targeted areas with critical resources needed to correct problems found to be conduits for crime. Terminology such as “research-based” initiatives, “blueprints for change,” and “best practices” is emerging in the Weed and Seed policy arena, not only in Pennsylvania, but also in similar programs around the country.

These research-based initiatives are just that: programs that have been evaluated and found to have a positive impact on local communities and their ability to curb serious and violent criminal activity. Funding from state and federal sources is being tied to conditions placed on grant recipients to conform to specific guidelines for program implementation and future evaluation of the new programs coming on line. The balance of this chapter examines some of these programs and argues for a more “sensible” approach to the crime problem. This approach is grounded more in sound policy analysis and processes and less in symbolic politics.

Research-Based “Promising Approaches” to Curb Crime in U.S. Communities

In 1998, Lawrence Sherman and his colleagues^[14] at the University of Maryland released a report to the National Institute of Justice in which they summarized a comprehensive review of evaluation research in the area of crime prevention. In the report, entitled “Preventing Crime: What Works, What Doesn’t, What’s Promising,” Sherman et al.^[14] identified seven institutional settings in which these promising approaches take place: 1) communities; 2) families; 3) schools; 4) labor markets; 5) places; 6) police agencies; and 7) other agencies of criminal justice. The authors first distinguished between the terms “practice” and “program,” a distinction that is important for policy makers and practitioners. They suggest that a practice is any activity that is continuous in a particular setting, and it can only change if a new government policy restricts it or takes the place of the old practice. A program, on the other hand, is much more specific. It is a *focused effort* to affect change in a crime prevention setting. As such, unlike routine practices, actual program implementation and continuation requires a continuing source of resources. Given the amount of resources going toward crime prevention programs, it is not surprising that the Department of Justice would have a great interest in the findings from an examination of any evaluation of those programs.

The Maryland Report, as the Sherman et al.^[14] study has come to be called, takes a great deal of evaluation research and condenses it into more simple terms. At the end of each chapter, where various authors addressed those programs associated with the above-identified seven institutional settings, there is a concluding section that summarizes “what works” and “what doesn’t.” Beginning with the community setting, it appears that what will work to bring down crime is gang violence prevention focused on reducing gang cohesion and mentoring programs targeted toward 10- to 14 year-olds. What does not work at the community level are community mobilization efforts and gun buyback programs, both of which have received a great deal of attention and press coverage in the recent past.

Family programs identified in the Maryland Report as being promising are long-term frequent home visitation programs that have been shown to reduce child abuse, and family therapy for delinquent and predelinquent youth. In the school setting, programs aimed at improving communication about rules and behavior norms appear promising (anti-bullying programs, for example), as long as those rules are consistently reinforced. Also, comprehensive instructional programs that focus on such issues as stress management, problem solving in social settings, and self-control, and that are maintained over an extended period of time, appear promising. To bring down substance abuse by young people, behavior modification programs, in addition to the comprehensive instructional programs previously listed, are needed, according to Sherman et al.^[14]

Programs that do not work in the school setting to reduce substance abuse are peer-group counseling sessions and single efforts at offering more recreation activities. Further, Sherman et al.^[14] concluded that education programs based on fear arousal, moral appeal, and information dissemination will not work to bring down the use of drugs by young people.

When it comes to labor market programs aimed at reducing crime in U.S. communities, Sherman et al.^[14] identified such programs as short-term vocational training programs for older male parolees. Promising approaches include intensive training programs (Job Corps, for example) for at-risk youth, prison-based vocational programs for adults, and the development of Enterprise Zones. What does not work, according to Sherman et al.^[14] are subsidized work programs for at-risk youth and pretrial diversions for adult offenders (making employment training and finding a job a condition of case dismissal).

Targeting “places” and policing programs can work together to prevent crime in local communities but only under certain conditions.^[14] These two approaches are often termed “situational crime

prevention through problem-oriented policing.” One of the more promising approaches is one in which some form of crime analysis is conducted (mapping of types of serious crimes, for example) followed by a well-designed police effort in those targeted areas. In fact, of all the studies examined in the area of policing, Sherman et al.^[14] argued that none other is effective as a clearly, focused program. Simply putting more police on the streets is not going to do much to bring down serious crime. Saturated patrol may have some short-term effect, but that effect fades over time without a more focused, targeted approach.^[14]

Finally, and in the area of corrections, evaluation studies point to the conclusion that the following programs can work to reduce repeat offending: prison-based therapeutic community treatment of drug-addicted offenders; structured and focused treatment programs that use multiple components; programs that allow for a substantial amount of contact between the treatment provider and the offender; and programs that address the characteristics of the client.^[14] As alluded to above, programs that focus only on incapacitation, and that lock up less serious offenders for longer periods of time, are likely to be extremely expensive and to not do much to bring down serious crime.

SUMMARY AND CONCLUSIONS

This “tip of the iceberg” summary of the many studies examined in the Maryland Report suggests that there is ample evidence on which criminal justice policy makers and practitioners can rely when it comes to making critical decisions about program changes in the criminal justice system. The move toward incarcerating more individuals and for longer periods of time, while symbolically pleasing to many, may have failed as a crime prevention approach. Practices such as gross incapacitation through mandatory minimum sentences have been shown to overburden the criminal justice system while not doing much to reduce recidivism. More sensible approaches to the crime problem seem warranted, and research-based initiatives are appearing as blueprints for needed change. The question becomes, “To what extent are policy makers and practitioners willing to pay attention to such reports as those produced by Sherman and his colleagues?”

Petersilia^[15] has argued that research has affected how policy makers “*think* about issues, problems, and the range of viable solutions.” In her example, it is the *conceptual* use of research findings that may be more pronounced than the so-called *instrumental* use of research findings. In other words, although the effects of research on policy development may not always be measured and thus can be easily missed,

the effects are there, nonetheless, and present themselves in more subtle ways. The word gets out. It may not come through an examination of peer-reviewed journal articles by criminal justice practitioners, but there is now a substantial amount of reporting of program evaluations appearing in trade journals and being posted on government web sites to support an argument that research findings are now being disseminated beyond mere academic circles. As stated above, many sources of funding are tying that funding to programs that are research based. Making use of criminal justice research in the policy process has actually begun to take place.

Criminal justice research is better preparing practitioners and policy makers to analyze the problem at hand. The policy process is informed, through the research, about the history of the problem and the outcome of previous interventions. Using the information gained in this first step of the process, goals and objectives of any new program can be determined. The process can then move toward a discussion, by key stakeholders, about the feasibility of implementing the new program, including brainstorming about any obstacles that might be faced at that point in the process. During the actual designing of the new program, the research findings add even more to the mix through assisting policy makers and practitioners in determining the target population, how clients will be selected, and even what type of staff will be needed to fill the new positions created by the program. Finally, building in an evaluation component is critical to new program development and implementation. Impact evaluations of the new program can be used then to make any necessary changes to the new program and to guide others in development and implementation of similar programs.

As Welsh and Harris^[6] argued, “Rarely does planning go smoothly, and it takes willpower, a clear vision, and lots of communication to remain rooted in the planning process.” They also add that “planned changed” is better than ad hoc decision making, and that it “increases the likelihood of successful interventions.”^[6] With the safety of the community at stake, planned changes in the criminal justice policy arena, following closely the policy analysis process as well as research-based recommendations for what will work to accomplish that goal, are most certainly called for.

REFERENCES

1. Packer, H.L. *The Limits of the Criminal Sanction*; Stanford University Press: Stanford, CA, 1968.
2. Hancock, B.W.; Sharp, P.M. *Public Policy, Crime, and Criminal Justice*, 2nd Ed.; Prentice Hall: Upper Saddle River, NJ, 2000.

3. Bureau of justice statistics. *Sourcebook of Criminal Justice Statistics—1998*; Government Printing Office: Washington, DC, 1999.
4. Dye, T. *Understanding Public Policy*, 4th Ed.; Prentice Hall: Englewood Cliffs, NJ, 1981.
5. Walker, S. *Sense and Nonsense About Crime and Drugs: A Policy Guide*, 5th Ed.; Wadsworth Press: Belmont, CA, 2001.
6. Welsh, W.N.; Harris, P.W. *Criminal Justice Policy and Planning*; Anderson Publishing: Cincinnati, OH, 1999.
7. Blumstein, A.; Cohen, J.; Rosenfeld, R. Trend and deviation in crime rates: a comparison of UCR and NCS data for burglary and robbery. *Criminology* **1991**, *29*, 237–263.
8. Wolfgang, M.; Figlio, R.M.; Sellin, T. *Delinquency in a Birth Cohort*; University of Chicago Press: Chicago, IL, 1972.
9. Wenk, E.A.; Robison, J.O.; Smith, G.W. Can violence be predicted?. *Crime Delinq.* **1972**, *18*, 339–402.
10. Zedlewski, E.W. *Making Confinement Decisions*; Government Printing Office: Washington, DC, 1987.
11. Zimring, F.E.; Hawkins, G.E. The new mathematics of imprisonment. *Crime Delinq.* **1988**, *34*, 425–436.
12. Lipsky, M. *Street Level Bureaucracy: Dilemmas of the Individual in Public Service*; Russell Sage: New York, NY, 1980.
13. Stolz, B.A. *Criminal Justice Policy Making: Federal Roles and Processes*; Praeger: Westport, CN, 2002.
14. Sherman, L.W.; Gottfredson, D.; MacKenzie, D.; Eck, J.; Reuter, P.; Bushway, S. *Preventing Crime: What Works, What Doesn't, and What's Promising*; National Institute of Justice, Department Of Justice Programs: Washington, DC, 1998.
15. Petersilia, J. Policy relevance and the future of criminology. In *Public Policy, Crime, and Criminal Justice*, 2nd Ed.; Hancock, B.W., Sharp, P.M., Eds.; Prentice Hall: Upper Saddle River, NJ, 2000; 383–395.

Crisis Policy-Making and Management in Southeast Asia

Scott Fritzen

Public Policy Programme, National University of Singapore, Singapore

INTRODUCTION

Crises severely test the capacities of bureaucracies, politicians, and, at times, the overall resilience of political, economic, and social systems. Studying crisis management across country contexts enables the exploration of systemic capacities and policy-making patterns as these undergo (at times, extreme) stress. This review briefly introduces a framework for cross-country analysis and applies to Southeast Asia. It has four sections. First, it presents hypotheses for how different crisis types interact with country governance characteristics to influence decision making. Second, summary characteristics of crises influencing policy making and management are presented for a range of crises that have affected Southeast Asia in the past 10 years. Third, regional responses to the emergence and spread of the Severe Atypical Respiratory Syndrome (SARS) virus in 2003 are analyzed as one example of interconnections posited between governance characteristics and crisis management. The concluding section explores implications of the framework for those interested in improving the robustness of crisis management in the region.

CRISIS DECISION MAKING AND GOVERNANCE CHARACTERISTICS

Analytical frameworks for comparing crisis decision making across national boundaries are at an early stage of development, in contrast to the considerable literature that exists on individual types of crisis, such as terrorism or natural disasters (see the *Handbook of Crisis and Emergency Management*^[1] for comprehensive coverage). Two types of generalizations are likely to be important for future work in this area.

The first stems from the pioneering work of Grindle and Thomas,^[2] who analyzed ways in which problem characteristics and context help determine patterns of policy decision making. Distinguishing between “crisis-driven” and “bureaucratic politics-as-usual” reform contexts, they find that decision making in the latter case is likely to be addressed at senior levels of government and to be strongly influenced by the need to maintain regime stability. Together with Kingdon,^[3] they also stress the importance of the

leadership abilities of “policy entrepreneurs,” who are shown to make a profound difference on outcomes even in resource-poor environments.

Building on the Grindle and Thomas dichotomy, we can posit that “not all crisis are created equal.” Different types of crisis will “trigger” different decision-making modalities. Four characteristics seem particularly important:

1. Degree to which a problem threatens overall regime stability. Ability to threaten the tenure of senior leaders became in effect a defining characteristic of a crisis in Grindle and Thomas’ framework, but it is clear that different types of crisis threaten *political* stability to very different degrees; it is also clear that the impact of the same type of problem on political stability will vary between regimes. The greater the degree to which regimes and incumbents are threatened by a crisis, the more decision making will tend to be dominated by the political calculus of senior officials.
2. Degree to which a problem is potentially amenable to a “technical” solution. The effectiveness of decision making in crises involving considerable technical complexity (such as environmental protection) will in large part be determined by the bureaucratic capacities that exist in a given system.
3. Degree of cross-national spillover involved. Interstate crises raise the complexity of problem solving by forcing decisions to be negotiated with multiple stakeholders, both within and outside of a polity. The greater the spillover effect, the more outcomes will be shaped by the effectiveness of regional forms of cooperation and by differentials in power between states.
4. The continuity of issue as crisis (whether short-acting or chronic). All crises, by definition, thrust themselves unpredictably onto the policy agenda. Some, however, are recurring in nature—sometimes over decades—such that precedents and “organizational repertoires”^[4]—patterned national and bureaucratic responses to a given problem—are triggered with each recurrence of the crisis. In unprecedented crisis contexts, the role of individual leaders in shaping

responses may be highlighted to a greater extent than in the case of “chronic” crises.

A second broad generalization, to date insufficiently explored in the literature (but see Grindle^[5]), is that state capacities matter in crisis management. A strategic approach to crisis policy making and management^[6,7] focuses on three categories of capacities:

1. *Political and leadership capacities*, which, as the name implies, has two aspects depending on whether the capacities involved adhere primarily to an individual or to the system as a whole. A critical feature is the way political leaders (or, in rarer cases in Southeast Asia, policy entrepreneurs outside government) direct attention to a problem in a manner that shapes how the crisis is perceived.
2. *Bureaucratic and process capacity* involves the ability to manage a high-quality decision-making process during a crisis^[8] and to execute decisions taken with predictability and control. This is a profoundly institutional capacity resting on many foundations: a well-functioning bureaucracy, clear information systems and resources (both manpower and fiscal) to tackle implementation challenges that arise.
3. *Social capacity* in times of crisis relates to the willingness of social or economic groupings to accept overall government coordination and/or to participate in the constructive response to the crisis.

Looking across these categories, one can make what might be called the “weakest link” hypothesis. Any

given crisis may invoke multiple types of decision-making responses and rely for its resolution on multiple capacities. Yet some capacity constraints will be more binding for the “successful” management of crises than others. For politically sensitive crisis points, a governance system’s political capacities will, in particular, be tested, and crises involving high technical complexity will depend heavily on a system’s bureaucratic capacity. Crises with high spillover effects, in turn, will depend prominently on the strength of mechanisms for interstate cooperation—a capacity that in Southeast Asia has been growing steadily over the past two decades, but which is still stretched thin.

COMPARING CRISIS CHARACTERISTICS IN SOUTHEAST ASIA

Multiple kinds of crisis have affected Southeast Asian countries of the past 5 years or so, each with different characteristics based on the framework above (Table 1). These can be divided into three groupings that suggest different hypotheses for future research.

The first grouping includes chronic, largely political problems that are generally not amendable to any obvious “technical” solution, but are also unlikely to precipitate regime change on their own. Terrorism, internal violence, secession struggles, and human rights disputes all might fit into this category. In between inevitable crisis points, these problems have tended to simmer without resolution where they have affected low-capacity political systems (as in Indonesia or Cambodia). Their resolution will depend on the degree to which political and adaptive capacities develop in, as yet, weakly institutionalized democracies.

Table 1 Characteristics of crises buffeting Southeast Asia since 1997

	Technical solution	Spillover	Regime threat	Continuity
Health (SARS, bird flu)	High	High	Low	Low
Health (HIV–AIDS)	Moderate	Moderate–high	Low	High
Terrorism (e.g., response to Jemmah Islamiyah)	Low	High	Low	Moderate
Economic and financial crisis (1997–1999)	Moderate	Moderate	Moderate–high	Moderate
Regime challenge (e.g., “People’s Power II” or the fall of Soeharto)	Low	Low	High	Low
Interstate dispute (e.g., Cambodia–Thailand “Angkor Wat” incident in January 2003)	Moderate	High	Low	Low
Internal civil violence (e.g., Maluku unrest)	Low	Moderate	Moderate	Moderate
Secessionist struggles (e.g., Aceh, Mindanao)	Low	Moderate	Moderate	High
Human rights problems in regional countries (e.g., ASEAN response to Myanmar’s detention of Aung San Suu Kyi in May 2003)	Low	Moderate	Low	High
Environmental dispute (e.g., Indonesia–Singapore “smog” problem, 1997)	Moderate	High	Low	Moderate

A second category of problems includes technical problems with high “contagion,” moderately capable of threatening regime tenure. Handling of epidemic outbreaks and regional economic crises, such as (notably) the East Asian financial crisis beginning in 1997, fall into this category. Outcomes in these cases depend, to a large extent, on problem continuity and the technical capacity and interest of the affected governments to address the problems. Addressing major problems of high contagion in this manner clearly goes beyond the capacity of less-developed Southeast Asian countries, which must therefore rely, to a large extent, on regional management mechanisms. As shown by the SARS example below, the degree to which bureaucratic capacity and interstate norms can simultaneously develop becomes critical to the eventual resolution of crises in this category.

The third category involves direct challenges to government tenure, extending in the extreme to challenges to the entire governance system (as in Indonesia during the fall of Soeharto). These are, of course, political crises points, the resolution of which depends on the balance of state capacity, legitimacy, and institutionalized conflict-management mechanisms built up over the history of the polity concerned.

RESPONDING TO A CRISIS OF UNKNOWN PROPORTIONS: THE CASE OF SARS IN 2003

Different responses in Southeast Asia to the outbreak of a new and deadly disease—Severe Atypical Respiratory Syndrome (SARS)—serve as an illustration of the interaction effects of crisis and governance characteristics. SARS originated in southern China in November 2002, reached Hong Kong in February 2003, and peaked in May 2003, by which time it had spread eventually to over 25 countries. It was to cause just fewer than 800 deaths before its chain of transmission was declared broken in all countries in July 2003. Of the 10 worst affected countries, seven were in Asia, foremost among them mainland China, which accounted for almost 90% of worldwide cases.^[9] This health crisis gripped the public imagination in a manner much different from the far more devastating HIV virus, in part because it erupted in a short time, caused widespread uncertainty (including over transmission modalities and fatality rates) and massive economic costs to the countries worst affected. SARS posed a tremendous challenge above all to East and Southeast Asian countries and territories that bore the brunt of the disease—seven of the 10 worst affected countries were in Asia.^[10]

The SARS epidemic had a high technical content; the policy response to SARS was predicated on getting

better information on transmission routes and effective mechanisms of control. It was moderately threatening in political terms, with fears concentrated on the potential economic costs as well as the social fallout from being seen to handle the virus incompetently. Of course, SARS had high international spillover (and therefore generated high amounts of pressure from the international community for affected countries to institute transparent control mechanisms). Despite sharing these crisis characteristics, the way the SARS virus was handled differed substantially in China, Vietnam, and Singapore in ways that shed light on the intersection of governance qualities and crisis-management modalities.

In Vietnam, early detection of the virus and a swift response led to eventual containment. The government’s response drew praise from the World Health Organization (WHO):^[11] “Vietnam demonstrated to the world how a developing country, hit by an especially severe outbreak, can triumph over a disease when reporting is prompt and open . . .” Partly, the response was one of luck, as a rapid initial effort to contain the spread was successful. As was shown eventually in neighboring China, this success was a function of effective dissemination of information through Vietnam’s some 10,000 rural health centers, demonstrating that even in relatively poor countries, infrastructural and organizational capacities, once mobilized, may be impressive. The resolution of the SARS outbreak in Vietnam, and the slower but eventually effective response in China, was multifactorial, “involving improvements in management and triage in hospitals and communities of patients with suspected SARS and the dissemination of information to health care workers and the public.”^[12]

It is the initial Chinese response that calls for further analysis. Information regarding the true extent of the outbreak in China was suppressed for some weeks before the government launched a full-scale attack on the disease. The normally diplomatic (by necessity) WHO noted as much, saying, “SARS is now known to have begun in mid-November [2002] in Guangdong Province. Cases during the earliest phases of the SARS outbreak there were not openly reported, thus allowing a severe disease to become silently established in ways that made further international spread almost inevitable.”^[13] Harsher assessments were found in the international press, with *The Economist* magazine^[14] likening the Chinese performance to that of the Soviet Union during the Chernobyl disaster. It was as the international spotlight fully glaring on the Chinese that the stakes, in terms not least of national image, became great enough for the central leadership to initiate its rapid and comprehensive course correction that finally led to an effective response.

Singapore generally drew high praise for its response to the outbreak. As an international trading and tourist hub, Singapore clearly stood to lose a great deal to the virus. The government's early response to it was to pull no stops; virtually all containment measures imaginable, including use of military forces to enforce quarantine requirements,^[15] were taken with alacrity. The leadership was engaged in a high-profile manner in communicating information regarding the latest occurrences of the virus and its transmission mechanisms. A recent review of Singapore's experience, based on a survey of 1200 Singaporeans, found that respondents evaluated highly the authorities' management of the SARS crisis, with a special emphasis on their openness to communication. The authors also pointed to the importance of social capacity in responding to the virus, finding "a relatively high level of social discipline in the population."^[16]

Singapore, Vietnam, and China are all authoritarian states generally capable of mobilizing society behind high-priority national decisions society unencumbered by legal niceties and challenges. However, in this type of crisis, technical capacities and leadership—the ability of the authorities to grasp the situation at the grassroots and their willingness to act in a transparent manner on this information—mattered more than the macrogovernance characteristics of the regimes. As a technologically savvy city-state, transaction costs to access information about what was "really" happening over the course of the disease were reasonably low in Singapore, whereas the complexity of governing the world's most populous nation, coupled with classic distortions in center-periphery communications in China, led to its central government initial incapacity to take action transparently to thwart this emerging crisis.

IMPLICATIONS

Regime characteristics and systemwide capacities clearly matter in determining crisis-management outcomes. Three types of capacities can be examined, in particular, country contexts—political, technical, and social. Overlapping their boundaries are two additional "meta-capacities:" for systemwide learning and adaptation and for effective international cooperation and lesson sharing. Mechanisms for responding collectively to problems of high technical complexity are still at an early stage of development in Southeast Asia, although—as demonstrated in part by the SARS example—they are improving.

The characteristics of crises themselves also matter, and this has implications for those seeking to shore up regional capacity to address crises in Southeast Asia. Technical capacities will be particularly important

where the crisis context allows decision making to proceed in a relatively politically neutral fashion (unlike the early response to SARS in China). Reformers seeking better crisis management in such cases might hope to mobilize external parties affected by the issue—i.e., to raise the "spillover" salience of an issue—in order to prompt better cooperation and national responsiveness.

Where political motives are paramount in responses to crisis, answers will be more complex. Systems of conflict resolution are needed at the regional level that are both technically and politically viable. The influence of NGOs and others in keeping unpopular issues on the policy agenda will be essential, as will ongoing attempts to open policy making to greater external scrutiny and input. Although state capacities to handle crises in Southeast Asia are highly uneven, robust institutions to manage political transitions and accumulated social pressures are in fairly short supply in most Southeast Asian countries.

REFERENCES

1. Farazmand, A. *Handbook of Crisis and Emergency Management*; Marcel Dekker, Inc.: New York, 2001.
2. Grindle, M.; Thomas, J. *Public Choices and Policy Change: The Political Economy of Reform in Developing Countries*; Johns Hopkins University Press: Baltimore, MD, 1991.
3. Kingdon, J. *Agendas, Alternatives, and Public Policies*; Harper Collins College Publishers: New York, 1995.
4. Allison, G.; Zelikow, P. *Essence of Decision: Explaining the Cuban Missile Crisis*, 2nd Ed.; Longman: New York, 1999.
5. Grindle, M. *Challenging the State: Crisis and Innovation in Latin America and Africa*; Cambridge University Press: New York, NY, 1996.
6. Nice, D.; Grosse, A. Crisis policy making: some implications for program management.. In *Handbook of Crisis and Emergency Management*; Farazmand, A., Ed.; Marcel Dekker, Inc.: New York, 2003; Chapter 5.
7. Irving, K.J. Managing "Complex Emergencies:" U.N. administration and the resolution of civil wars.. In *Handbook of Crisis and Emergency Management*; Farazmand, A., Ed.; Marcel Dekker, Inc.: New York, 2003; Chapter 10.
8. Moore, M. *Creating Public Value: Strategic Management in Government*; Harvard University Press: Cambridge, MA, 1995.
9. Lam, W.K.; Zhong, N.S.; Tan, W.C. Overview on SARS in Asia and the World. *Respirology* **2003**, 8 (11).
10. World Health Organization. *Severe Acute Respiratory Syndrome (SARS): Status of the Outbreak and Lessons for the Immediate Future*; WHO: Geneva, May 20, 2003.

11. *Severe Acute Respiratory Syndrome (SARS): Status of the Outbreak and Lessons for the Immediate Future*; WHO: Geneva; 720 May, 2003.
12. Pang, X.H.; Zhu, Z.H.; Xu, F.J.; Guo, J.Y.; Gong, X.H.; Liu, D.L.; Liu, Z.J.; Chin, D.P.; Feikin, D.R. Evaluation of control measures implemented in the severe acute respiratory syndrome outbreak in Beijing, 2003. *J. Am. Med. Assoc.* **2003**, *290* (24), 1.
13. World Health Organization. *Severe Acute Respiratory Syndrome (SARS): Status of the Outbreak and Lessons for the Immediate Future*; WHO: Geneva, 20 May, 2003; 8 pp.
14. The economist. In “China’s Chernobyl?” London; 24 April, 2003. Available at <http://www.taiwandc.org/economist-2003-01.htm> (accessed January 10, 2004).
15. World Health Organization. *Severe Acute Respiratory Syndrome (SARS): Status of the Outbreak and Lessons for the Immediate Future*; WHO: Geneva, 20 May, 2003, 5 pp.
16. Quah, S.R.; Lee, H.-P. Crisis prevention and management during SARS outbreak, Singapore. In *Emerging Infectious Diseases [serial online]* [accessed December 20, 2003]; Feb 2004. Available from <http://www.cdc.gov/ncidod/EID/vol10no2/03-0418.htm>.

Customer Service

Evan M. Berman

*Public Administration Institute, Louisiana State University,
Baton Rouge, Louisiana, U.S.A.*

INTRODUCTION

Customer service is the activity of identifying and satisfying the needs of customers. In public organizations, the term “customer” is broadly used to include program clients, external customers of public organizations, internal customers (for example, employees who depend on the service of other employees in the same public organization), and citizens. The term “customer service” is sometimes also used to connote a standard, such as in the statement: “The customer service in this public agency is world class.” Customer service is currently viewed as essential to maintaining sound stakeholder relations.

RECENT DEVELOPMENT OF CUSTOMER SERVICE IN THE PUBLIC SECTOR

Although customer service is arguably as old as public administration, the importance of providing customer service by U.S. public organizations dramatically increased in the mid-1980s. This increased emphasis followed that of U.S. corporations in the late 1970s and early 1980s, during which period U.S. corporations lost market shares to Japanese and other foreign firms. These firms had innovated a new approach to product development based on the rapid utilization of customer feedback. The new approach, based on a management philosophy called “Total Quality Management” (TQM; see below) resulted in products with superior reliability and more user-oriented features. American firms quickly and successfully copied the new management philosophy that emphasized customer service in an effort to counter the advantage that foreign rivals enjoyed. Later, this new management philosophy was widely adopted by other organizations; TQM was broadly introduced into the federal government during the late 1980s, and local and state governments followed in the early 1990s.

In broader context, the period of the late 1970s and early 1980s is seen as a nadir in stakeholder trust toward American private and public institutions. Trust in the federal government reached a record low in the early 1980s. Whereas in the mid-1960s about 62% of respondents answered the question: “How often do

you trust the federal government to do what is right?” with “most of the time” or “some of the time,” by the early 1980s, this percentage had dropped to about 18%. It is probable that concerns in the late 1970s and early 1980s of high unemployment, high inflation, two energy crises, and deceit by elected officials (for example, Watergate) contributed to broad public distrust, as did poor service to program clients and citizens. State and local governments enjoyed a higher level of trust but were not immune from public apathy and cynicism. With 20/20 hindsight, the turn to customer service in the public sector can be viewed as an effort to address causes of citizen distrust that were narrowly related to program quality and service. The mantra “to run government as a business” (that is, borrowing management practices from the private sector) regained importance as a dominant ideology in the late 1980s. Although some critics regarded the customer service emphasis as a blind adoration of capitalist values and as inconsistent with the fiduciary and political responsibilities of government, most observers believed that an increased orientation toward customer service was long overdue in government, as it had been in business.

Customer service is an integral part of TQM. This management orientation emphasizes meeting customer needs through a comprehensive strategy that encompasses empowerment, process analysis and improvement, customer service, partnering and performance measurement, and benchmarking.^[1] Although initially organizations attempted to implement TQM as a coherent management strategy, today it is widely regarded as a management philosophy that encompasses a broad range of specific strategies, among which is customer service. As a management philosophy, the core values of TQM are to: 1) identify, meet, and exceed the needs of customers (and other stakeholders); 2) strive to produce services right the first time (that is, reduce errors that upset stakeholders and cause rework and increased costs); 3) use systematic analysis to evaluate and improve services delivery; and 4) consistently support workers in their efforts to improve quality and meet customer needs. TQM invites managers to think broadly about customer service and needs and to develop innovative approaches to improve service delivery. Many applications of

TQM begin with emphasizing meeting and exceeding customer expectations, in part by focusing on customer complaints. Hence, customer service is usually the first strategy in TQM efforts. Other strategies that affect customer satisfaction are discussed below.

Generally, the principles of establishing a customer service strategy involve addressing the following questions: 1) How do our long-term plans serve the needs of our customers?; 2) What specifically do customers expect from us?; 3) How can we exceed their expectations?; 4) Are we exceeding customer expectations?; 5) What can we do to further improve upon our satisfaction of customer expectations?; and 6) Do we have the capacity to meet our customer service goals? Regarding the latter, is the commitment of managers and employees as well as technology infrastructure, consistent with our customer service goals? If not, how can we improve?

THE IMPORTANCE OF ORGANIZATIONAL CULTURE

Widespread agreement exists that customer service outcomes reflect an organization's values, that is, the daily expectations and practices of organizational members as they deal with customers and plan their work. All organizations have expectations for how staff should deal with customers; they vary with regard to these expectations, including the extent that customer service is viewed as being important. At issue are not generally espoused values such as "we value our customers highly," but rather the specific guidelines that all employees and managers are given as they go about their daily work. The specifics matter, because general statements provide little guidance and assurance in the face of specific daily challenges. Some examples of specific customer service values that are found in organizations are the following:

- We always stay focused on resolving the customer's problem or need.
- We value changing what or how we do it to increase customer satisfaction.
- We want our customers to think good things about us.
- We use policies and procedures as guides to resolving requests or complaints, not as excuses for saying "no."
- We try to provide the entire service as soon as possible; at a minimum, we provide a concrete result within 24 hours.
- We always address customers with dignity, respect, and calm.
- Whatever you do, do not be the source of any customer complaint.

- We acknowledge and reward superior customer service in ways that matter to our employees and managers.

Agreement also exists that top managers play a crucial role in making customer service an important organizational value. Organizational values usually face competing challenges from other values, such as being responsive to elected officials, getting along with others in the organization (not rocking the boat), and respecting rules and policies. One role of top management is to clarify the importance of customer service, for example, "our customers vote for our elected officials, who are our bosses," in relationship to other competing values. When customer service is made an important priority, top management will also have to resolve, or provide a process for resolving, various issues that come up as staff attempt to achieve customer service, such as unclear or contradictory rules or agency practices. Top managers need to be consistent in these clarifications and in their support for others who deal with these clarifications.

Another role is ensuring that specific customer service values are ingrained throughout the organization. Broad-scale commitment is necessary, because customers come in contact with a broad number of employees. Also, employees often depend on other employees to satisfy the needs of "their" customers, thus, all employees must work in concert to achieve customer satisfaction. The job of top management is to ensure that organizational values relating to customer service are driven deep into every nook and cranny of the organization, affecting the actions and thought processes from the highest managers to the lowest-ranking employee. To this end, top management must ensure that the structure of rewards and expectations promotes its values with regard to customer service. Typically, middle managers and supervisors become accountable for ensuring that these values receive broad-scale support in their units. Training is usually part of instilling these values at lower levels (see further). Managers can also be asked to lead by example. Continuing top management support appears to be critical, even after work groups have embraced it. In short, top management plays an important role in shaping customer service actions of organizational members.

CHALLENGES TO ACHIEVING CUSTOMER SERVICE IN THE PUBLIC SECTOR

Organizations that begin on the path toward improved customer service often face a number of issues. An important top management task is to acknowledge these issues and to assist in clarifying them for others.

First, many authors have noted differences between private and public sector “customers.” Individuals often wear different hats as they interact with governments. Citizens expect something different from government than only easy access to its services, and these different roles are sometimes seemingly at odds: for example, as citizens, many residents value speed limit enforcement in their neighborhoods, but as motorists, they might not value enforcement. Public organizations must acknowledge these different roles and accommodate the differences or assist others such as elected officials in doing so.^[2] Also, the fact that individuals play different roles does not preclude improvement in their interactions with public agencies.

Second, in some public organizations, customers are involuntary, such as regulated businesses, taxpayers, and inmates. The paradox is presented that it may be impossible for regulators to satisfy the regulated, when their requests are denied, and actions are monitored. According to Bruce McClendon,^[3] surveys show that customer satisfaction has little to do with whether a request is approved; it is more strongly affected by whether employees are knowledgeable, respectful, and creative in finding solutions, and whether services are accessible, accurate, consistent, timely, and dependable. Explanations need to be given for how and why requests are granted or denied. The physical appearance of service areas is also a factor. It seems reasonable that inmates might be approached in a similar way; for example, they might be told why a request cannot be immediately granted.

Third, customer service is consistent with orientations toward empowerment, sometimes more than is customary in some organizations. It is obvious that “one size does not fit all,” thus employees will need to use a little ingenuity in dealing with customer requests. Empowerment means holding employees and managers accountable for results; rather than using rules and regulations to stipulate how service is to be provided, agencies may do better by providing a few stipulations as to what cannot be done (or which may require supervisory review), and a mix of general principles that aid in providing a high level of customer service. This approach gives employees considerable latitude in meeting customer needs, but it also implies the possibility of inappropriate actions that might be taken. How managers specify the latitude and responsibility of employees, and how they deal with customer service failures (for example, through punishment or as learning opportunity) greatly shapes employee orientations toward customer service efforts.

Fourth, customer service requires attention to resources and rewards. It is sometimes alleged that public agencies have inadequate resources for training and insufficient flexibility for monetary rewards that might shape behavior. Rather, the resources required

for training are quite minimal, though they are higher when investments are made in information technology as well. Reinforcement of employee and manager customer service behavior need not involve (many) monetary resources. A consensus exists that monetary rewards matter less in shaping behavior than immediate and consistent feedback; public acknowledgement (for example, in staff meetings) for tasks well done can be provided. In addition, formal approaches may be considered, such as incorporating customer service criteria in performance appraisal criteria, though these are seldom sufficient in the absence of informal approaches to instill and maintain strong customer service values.

CUSTOMER SERVICE TRAINING

Employees need to be told what actions are expected of them. Thus, training is usually part of an effort to increase customer service. Training usually begins with identifying the need for customer service, assessing current customer service outcomes and practices, specific organizational values and practices, applications to specific situations relevant to employees, and formulating an implementation or follow-up plan that discusses how change will be monitored, improved, and rewarded. Customer service training involves basic principles as well as advanced applications.

The basic principles of customer service are that it is about achieving real results for customers and the organization; it involves more than a slogan or merely putting on a friendly smile, although the latter may be part of the improvement. Customer service is attitude and expected behaviors that should always be present. The above values clarify specific attitudes and expectations that may be expected of employees. They involve matters of direct interaction (for example, how phone calls are transferred), as well as willingness to change the way that services are provided (for example, electronic acceptance of applications, whereas previously, only in-person or hard copy applications were accepted).

The basic principles include a four-step process of personal interactions with customers: open, interact, ask for an expression of satisfaction, and conclude. Examples in role playing are provided that help staff develop ways of dealing with customers that have a wide range of conditions, such those that are indecisive, irate, overly insistent, angry, violent, as well as those that are within a normal comfort zone in dealing with public agency personnel. These applications should be relevant to the actual jobs of employees. Basic training sometimes also includes a site visit or film of an organization that is known for excellence in customer service.

Advanced training involves the steps for creating customer service teams in departments, which are tasked with recommendations for improving customer satisfaction, and implementing specific proposals.^[4] Training involves a range of practical steps that such teams should take, as well as challenges that may stem from resistance or apathy. Advanced training also involves the development and use of a customer satisfaction survey. Although they are increasingly used, voluntary surveys are distinguished from systematic (or scientific) surveys. Voluntary customer surveys are completed at the initiative of the customer; although they are not generalizable to all customers, they do provide an early warning signal to managers about problems that may need to be addressed. Scientific surveys use a representative sample of customers; such surveys are generalizable, and are used for benchmarking customer service satisfaction.

MEASURING CUSTOMER SERVICE OUTCOMES

Organizations are increasingly measuring the level of their customer service:^[5]

1. Customer and citizen surveys are used to measure perceptions of customer service and program outcomes.
2. Objective measures, such as performance measures, are used track performance in areas relevant to customers.
3. Trained observers are used to observe conditions against customer service standards.
4. Other methods, such as tracking customer complaints or compliments or conducting focus groups, are also used.

Scientific surveys are increasingly used in public administration to obtain the perceptions of a representative sample of citizens or program clients. Such surveys typically include such questions as “Have you had contact with any (our office) employees during the last 12 months?,” and, if yes, “Thinking about your recent experiences with _____, please tell me whether you strongly agree, agree, disagree or strongly disagree with the following statements:”

- The employees were helpful.
- The employees treated me with respect.
- The employees were friendly.
- The service was provided without mistakes.
- The service was provided in a timely manner.
- The service experience exceeded my expectations.

Responses to such items help assess average customer satisfaction. In many cities and counties, it

may be found that more than half of residents have had some contact with an official from the jurisdiction, that most residents have only one or two interactions, and that the perceived level of customer service during those interactions significantly affects overall perceptions of the jurisdiction. It thus follows that customer service is of great importance. Of course, surveys are also used to assess perceptions about the overall functioning of programs and policies.

Systematic surveys are not usually designed to provide much insight into the reasons for high or low levels of customer satisfaction; they usually only provide a valid snapshot of the current performance. Focus groups and customer comment cards can shine light on the causes of customer satisfaction. Customers can identify specific program features that are irritants to them, hence, suggesting or implying alternative courses of action for improvement. Although these source of information are not generalizable, they stimulate insight and, hence, possible action.

Performance measurement is increasingly used and often involves collecting objective data about program activities and outputs. For example, an emergency response program might measure the number of calls acted on within the first 2 minutes of receipt, the average time it takes for a response team to appear at location, the percentage of response teams that show up exceeding a standard time, the number of dispatched teams that did not involve emergency, and so on. Many of these administrative data are related to customer satisfaction, and continuous improvement is likely to affect customer satisfaction.

Trained observers are not always widely used, but they, too, can help assess customer service outcomes. For example, parks programs often used trained observers to rate the appearance of their facilities in ways that are relevant to customer satisfaction. Role playing is also used to assess customer interactions, such as at service request counters, toll booths, and telephone response centers.

DESIGNING PROGRAMS AND SERVICES WITH THE CUSTOMER IN MIND: SOME EXAMPLES

In recent years, public organizations have found many ways to improve customer satisfaction. In most instances, this involved redesigning the ways in which services are delivered, as well as the manner in which customers and citizens are addressed. Even basic operations such as answering citizen requests are redesigned. In many agencies, a standard is now set that calls are answered on or before the fourth ring, and that a “warm” hand-off is used when transferring calls; no longer are callers transferred to voice mail

boxes. Standards are also set for resolving caller queries, such as immediately or within 24 hours for most calls. Fairfax County, VA has implemented software that tracks the status of citizen requests.

Agencies have also redesigned their services to better meet the needs of citizens, often through information technology. Sometimes, this has required rethinking of the business that they are in. For example, the U.S. Internal Revenue Service (IRS) has reshaped itself from an agency that only attempts to maximize revenue to one that now aims to ensure that citizens pay the correct amount of taxes; taxpayers can call the IRS for advice on ways to reduce their tax burden. The IRS has also improved the way it deals with tax refunds. To satisfy its customers, the IRS has an electronic system, whereby taxpayers call to learn whether their Income Tax Returns have been received, when they will be processed, and when a check has been issued for any refund that is claimed.

Transportation agencies have also increased their customer orientation. The Orlando-Orange County Expressway Authority operates over 80 miles of toll roads in Orange County (Florida). It was among the first of such agencies to use an electronic system for collecting tolls, thus reducing traffic backups at toll booths during rush hour and allowing users of the electronic to drive through toll booths without stopping. This saves users many minutes in their commute. It is now considering expanding the use of its electronic payment device for such alternative uses as parking. Other agencies are also inquiring about the feasibility

of providing users with real-time updates of traffic situations and alternative routing.

Many jurisdictions are now investing in their Internet web site to improve their delivery of services. On many sites, citizens are now able to download a wide range of forms that they might need, such as for regulatory inspections or enrollment. They are also able to electronically submit these forms, through e-mail or by on-line completion of them. Governments are also beginning to use their Internet sites for e-commerce, such as for paying parking fines or providing user fees for services. They are also able to order government products on-line, such as special reports and T-shirts and mugs. In short, agencies are increasingly rethinking how they can satisfy and exceed the expectations of citizens, through what they do and how they do it.

REFERENCES

1. Berman, E. *Productivity in Public and Nonprofit Organizations: Strategies and Techniques*; Sage, 1998.
2. Leon, L.; Denhardt, R. The political theory of reinvention. *Public Adm. Rev.* **2000**, *60* (2), 89–97.
3. McClendon, B. Customer service for regulators. *Public Manage.* **1996**, *78* (5), 19–22.
4. Otey, D. Citywide customer satisfaction in Peoria. *Public Manage.* **2001**, *83* (1), 15–18.
5. Epstein, P. Measuring the performance of public services. In *Public Productivity Handbook*; Holzer, M., Ed.; Marcel Dekker, 1992; 161–194.

Data Integrity

Theo Edwin Maloy

*Department of History and Political Science, West Texas A&M University,
Canyon, Texas, U.S.A.*

INTRODUCTION

Data integrity is the assurance that information can only be accessed or modified by those authorized to do so. Integrity measures include controlling the physical environment of networked terminals and servers, restricting access to data, and maintaining rigorous authentication practices. Data integrity can also be threatened by environmental hazards, such as heat, dust, and electrical surges.

Physical measures to protect data integrity include making servers accessible only to network administrators; keeping transmission media, including cables and connectors, covered and protected to ensure that they cannot be tapped; and protecting hardware and storage media from power surges, electrostatic discharges, and magnetism.

Network administration measures to ensure data integrity include maintaining current authorization levels for all users; documenting system administration procedures, parameters, and maintenance activities; and creating disaster recovery plans for power outages, server failure, and virus attacks.^[1]

magnetism. There should be an uninterruptible power supply attached to the server.

- Combustible materials should not be stored near network equipment, and fire and smoke detectors should be installed near network equipment. Carbon dioxide fire extinguishers should be located near network equipment.
- Power cords should be covered to ensure they are not a hazard. Cables and connectors should be installed in accordance with organization standards and the National Electrical Code and the National Fire Protection Code.
- Sensitive, confidential, or critical documentation, files, and software should be stored in a fireproof safe when not in use.
- Interconnection device access should be disabled after hours.

NETWORK OPERATING SYSTEMS

Access to network operating system software, files, and commands should be controlled and appropriately restricted:

- Each node attached to the network should be uniquely identified.
- The network should provide the ability to log and to audit network activity when a security threat has been identified.
- The network should provide the ability to produce reports on network utilization and on audited events.
- Access privileges granted to read, modify, create, and delete files should be limited to only what is necessary. Data should be available only to users authorized to access that data.
- Access privileges granted to execute only, read, update, create, and delete programs and applications should be limited to only what is necessary. Access to utility programs should be controlled, and their use should be monitored. Each authorized user should be uniquely identified.
- The network should authenticate by password each authorized user. Users should be able to change their passwords and should be required to change their passwords at a predefined interval. Passwords

PHYSICAL AND ENVIRONMENTAL SECURITY

Alternative processing methods for critical functions should be available in case local area network services are disrupted for extended periods:

- Servers should be located in a physically secure area accessible only to the network administrators.
- Where every workstation may be used as a server, all devices should be under supervision when in use and should be physically secured when not in use.
- Transmission media should be covered and protected to ensure unauthorized personnel or devices cannot tap them. Transmission media should be physically secured in conduits, trenches, or ducts to prevent electronic eavesdropping.
- Hardware and storage media should be protected from excessive dust, dirt, heat, and moisture. Hardware and storage media should be protected from power surges, electrostatic discharges, and

associated with special privileges, such as the administrator's login, should be changed more frequently. Passwords should not be displayed and should be passed on the network in a secure format. A specific number of unsuccessful login attempts should result in the suspension of the login ID.

- Users should be automatically logged out or the workstation should be locked after a period of non-use to prevent unauthorized access.
- Access to data or files classified as sensitive, confidential, or critical should be controlled appropriate to their classification. All sensitive data transmitted over the network should be encrypted if encryption is available.
- An audit trail of all station additions and deletions should be maintained.
- For authorized personnel, security tables and files should be easy to maintain. Additions and changes to the security tables and files should be effective immediately.
- The software should provide the ability to delegate administration of some functions and some users in such a way that the subadministrator can be assigned responsibility for those portions of the security tables and files that directly affect his or her organizational unit.
- Adequate network administration resources, including manuals and vendor support, should be available.

DATA AND SOFTWARE BACKUP

Regular backups of production software should be performed, with backup devices kept in a secured location, protected from environmental hazards and unauthorized use. Periodic backup copies of the entire system should be made, as well as selected backup copies of specific files as needed. Backup procedures should be documented, and external labels should be used to identify the contents of backups. Data sensitivity and security should be monitored during recovery from a backup and should be reinstated once recovery is complete. Alternative processing methods for critical functions should be available in case local area network services are disrupted for extended periods.

NETWORK ADMINISTRATION

Managers and supervisors within the departments using the local area network should understand the importance of security, and should communicate that importance to departmental personnel:

- Network administrators should be trained, and a backup administrator should be identified.

- Access privileges for transferred or terminated employees or employees on leave should be reviewed and adjusted.
- Network administration functions should be limited to specific users and to specific network nodes.
- A current inventory of workstation and file server hardware and software should be maintained.
- Users should be aware of the license and contract restrictions that affect their use of the local area network. Copyrighted software and documentation should not be copied, and concurrent access to single-user software should not be allowed on the local area network.
- Users should be informed about their responsibilities for equipment, software, data security, and copyrighted material.
- Requests to develop or change programs and files should be approved by authorized personnel.
- Adequate documentation of system administration procedures, parameters, and maintenance activities should be maintained.
- Contact information should be available for vendors or others who have agreed to provide network maintenance.
- A disaster recovery plan should be in place that lists symptoms, possible causes, preventative measures, and recovery steps for power outages, server failures, tape backup unit failures, application software failures, network operating system upgrade failures, cabling failures, client failures, and virus attacks.
- Procedures should be in place to ensure that all modifications to routing tables and translator tables are properly authorized, that such changes are made only by authorized personnel, that changes are tested prior to implementation, and that an adequate audit trail exists to document all modifications.
- Procedures for changing or updating existing programs and files should be established and maintained. Vendor-supplied software changes and updates should be tested, installed, and distributed where appropriate.
- Users should be informed of changes such as new versions of software changes and the effects of those changes, and users should be informed about any restrictions on using hardware and software for personal use.

COMMUNICATION CONTROLS

Access to interconnection devices including gateway servers, routers, and bridges, should be limited to authorized users. Monitoring systems should be in place to detect unauthorized access attempts. Interconnection device access should terminate

automatically if unauthorized use is detected and should be disabled after hours. Access to the communication devices' configuration files and tables should be restricted to authorized network administrators.^[2]

PUBLIC ADMINISTRATORS' SPECIAL CONSIDERATIONS

Governmental agencies and private organizations that use public agencies' data must assure their customers, including individuals, firms, and other governmental entities, that the confidentiality and privacy of information they electronically collect, maintain, use, or transmit is secure. Information security is especially important when information can be directly linked to an individual or firm. Confidentiality is threatened not only by the risk of improper access to electronically stored information but also by the risk of interception during electronic transmission of the information.

Many government records require access signatures, but there is no national standard for security or electronic signatures. Each governmental entity must define its own security requirements. To satisfy the legal characteristics of a written signature, an electronic signature must identify the signatory individual, assure the integrity of a document's content, and provide strong and substantial evidence that will make it difficult for the signer to claim that the electronic representation is not valid.

Public Administrators should establish data integrity standards that conform to these generally desirable principles:

1. Improve system efficiency and effectiveness by leading to cost reductions or improvement of service from electronic transactions.
2. Meet the needs of the user community, particularly individuals, firms, and other governmental agencies.
3. Be consistent and uniform with other private and public sector data standards, including data element definitions and codes and their privacy and security requirements.
4. Have low additional development and implementation costs relative to the benefits of using the standard.
5. Be supported by an American National Standards Institute-accredited standards-developing organization or other private or public organization that will ensure continuity and efficient updating of the standard over time.
6. Have timely development, testing, implementation, and updating procedures to achieve administrative simplification benefits faster.

7. Be technologically independent of the computer platforms and transmission protocols used in electronic health transactions, except when they are explicitly part of the standard.
8. Be precise and unambiguous but as simple as possible.
9. Keep data collection and paperwork burdens on users as low as is feasible.
10. Incorporate flexibility to adapt more easily to changes in the infrastructure (including new services, organizations, and provider types) and information technology.

A master data dictionary providing for common data definitions should be developed and maintained. Data element definitions should be precise, unambiguous, and consistently applied. The transaction-specific reports and general reports from the master data dictionary should be readily available to the public. At a minimum, information presented will include data element names, definitions, and appropriate references to the transactions where they are used.

Security Standards

A security standard is a set of requirements with implementation features that individuals, firms, and governmental agencies must include in their operations to assure that electronic information remains secure. The standard need not address the extent to which a particular entity should implement the specific features. Each affected entity should assess its own security needs and risks and devise, implement, and maintain appropriate security to address its requirements. These measures must be documented and kept current.

There is no recognized single standard that integrates all the components of security (administrative procedures, physical safeguards, technical security services, and technical mechanisms) that must be in place to preserve information confidentiality and privacy as defined in the law. There are numerous security guidelines and standards in existence today, focusing on the different techniques available for implementing the various aspects of security.

There are several concepts on which a standard should be based:

1. The standard must be comprehensive. A standard must address all aspects of security in a concerted fashion.
2. The standard must be technology neutral. Because security technology is changing quickly, there should be flexibility in choosing technical solutions. A standard dependent on a

specific technology or technologies would not be flexible enough for future advances.

3. The standard must be scalable. The standard must be implemented by all the affected entities, from the smallest firm to the largest government agency. A single-sized approach would be neither economically feasible nor effective in safeguarding data.^[3]

Outsourcing and Privatization

Insourcing and outsourcing, usually called privatization in the public sector, describe a multitude of information technology options available in the marketplace. Total outsourcing is the decision to transfer assets, leases, staff, and management responsibility for information technology service delivery to a single third-party vendor that represents more than 80% of the information technology budget. Total insourcing is the decision to retain management and provision of more than 80% of the information technology budget internally. Selective sourcing is the decision to source selected information technology functions from external providers while providing between 20 and 80% of the information technology budget internally.

General outsourcing can be divided into selective, value-added, and cooperative outsourcing. Selective outsourcing is when one area of information technology, such as data center operations, is chosen to be turned over to a third party. Value-added outsourcing is when an area of information technology activity is turned over to a third party thought to be able to provide a level of support or service that adds value to the activity that could not be cost-effectively provided by the internal information technology group. Cooperative outsourcing is when a third-party provider, public or private, jointly performs some targeted information technology activities with the internal information technology department.

Information technology research has produced three major findings:

1. Conflicting stakeholder expectations place information technology managers in the precarious position of providing Rolls Royce service at a Chevrolet price. In general, senior management perceives the entire information technology function as a utility, and therefore, sets the information technology agenda to be cost minimization. Agencies and end users perceive that information technology critically contributes to operations, and therefore, sets the information technology agenda to be service excellence. Information technology managers are frustrated, because senior managers and

users set conflicting agendas for information technology, expecting information technology managers to deliver Rolls Royce service at a Chevrolet price.

2. The published literature portrays an overly optimistic view of information technology outsourcing. The optimism exists, because reports are made during the honeymoon period when clients first sign an outsourcing contract, and clients and vendors possess high outsourcing expectations. Second, the literature only reports projected savings, not actual savings. Third, public reports underrepresent outsourcing failures, because few managers wish to advertise a mistake. Therefore, the literature misrepresents the spectrum of outsourcing experiences by focusing on the success stories.
3. Organizational members initiated outsourcing evaluations for a variety of reasons. Financial reasons included cutting costs, improving cost controls, and restructuring the information technology budget. Business reasons included returning to core competencies and devolving organizational and management structures. Technical reasons included improving technical services, gaining access to new technical talent, providing access to new technologies, and focusing the internal information technology staff on core technical activities. Political reasons included reacting to the efficiency imperative, acquiring or justifying additional resources, reacting to the positive outsourcing media reports, reducing uncertainty, eliminating a burdensome function, and enhancing personal credibility.
4. Successful sourcing evaluations compared vendor bids against a newly submitted internal bid, not against current information technology performance. When vendors submit bids that indicate savings, managers must question how the vendor proposes to reduce costs and not just attribute savings to economies of scale. In many cases, vendor bids were based on efficient managerial practices such as standardization, consolidation, centralization, and tighter controls—practices that internal information technology managers may be able to replicate if empowered to do so. Implementing cost savings measures required information technology managers to adopt policies that may not be readily agreed to by the agency units or by the agency's clientele.
5. Internal information technology departments possess superior cost advantages from some information technology cost drivers. Outsourcing vendors have an inherent data center

operating cost advantage over small information technology departments, but they are comparable to large information technology departments. Internal information technology departments have comparable hardware purchase volume discounts with outsourcing vendors, but small departments have an inherent disadvantage. Internal information technology departments and outsourcing vendors have comparable software licensing costs. Internal departments have inherent advantages on agency expertise and organizational, marketing, and transaction costs. Outsourcing vendors have inherent advantages on technical expertise, research and development, and opportunity costs.

6. When managers decide to outsource, detailed contracts are more likely to be successful than relational contracts. Vendors are not partners unless profit motives are shared, and the profit motive is not prominent in the public sector. Many public and non-profit organizations used vendors to satisfy their information technology needs for hardware and software upgrades, installation and maintenance of area networks, staff training, expert solutions, data management of payroll and accounting, and general consultation about technology and available options. Many governments have learned that the small software firms lack the staying power to improve upon or even support their products beyond the first few years. These problems show that choosing a vendor and consultant must be done carefully and with objectivity. Choosing a

vendor begins with a needs assessment that should be based on multiple users and experts.

System maintenance contracts require great detail with regard to software development and enhancement, purchasing procedures, the expertise of service personnel, staff training, and responsibilities with regard to system reliability and security. Access, data integrity, and the protection of information assets are important, because public and non-profit organizations increasingly rely on their information systems for operations and decision making, and because information technology systems are increasingly subject to the threat of viruses, hackers, and sabotage.^[4] Outsourcing vendors and information technology suppliers compound the data integrity issues and vulnerabilities, especially when hardware or software is removed from the public administrator's or the agency's control.

REFERENCES

1. http://searchDatabase.techtarget.com/sDefinition/0,,sid13_gci518970,00.html (accessed October 2001).
2. <http://www.itd.umich.edu/cgi-bin/otfdoc?ID=r1099> (accessed October 2001).
3. <http://www.radcs.com/SESSpr.html> (accessed October 2001).
4. Lacity, M.C.; Hirschheim, R. Information technology outsourcing: what problems are we trying to solve? In *Rethinking Management Information Systems: An Interdisciplinary Perspective*; Currie, W., Galliers, B., Eds.; Oxford University Press: New York, 1999; 326–360.

Data Mining as a Decision-Making Aid

William J. Mead

Consultant, White Plains, New York, U.S.A.

INTRODUCTION

Data mining (DM) is a relatively new concept by which it has become possible and feasible to dig through huge collections of data looking for previously unrecognized trends or patterns. The resulting information can potentially be highly useful. The name is a clever takeoff on mining for precious metals or gems, where it is necessary to laboriously plow through masses of rubble to find a few pieces of material of great value.

Data mining is also called “knowledge discovery in databases,” although some workers in this field draw slight distinctions between this and data mining. The definitions have not been standardized or universally agreed upon.

Most organizations, such as government agencies, commercial and industrial businesses, health care facilities, educational institutions, etc., generate large volumes of data in the course of conducting day-to-day operations. These data come from a variety of sources, examples of which include the following: in business situations, from order processing, inventory control, production planning and billing; in financial institutions, from customer account transactions; in the telecommunications industry, from keeping track of billions of phone calls and Internet connections for billing purposes; in health care, for patient records and responses to medications and procedures; in educational institutions, for student records; in government, from such things as tax receipts and expenditures, and the activities in various agencies.

For many years, computer mainframes and personal computers have been grinding out data, much of which has gone to inactive storage. But, within the past decade, it has been realized that by going back and carefully studying these banks of data, interesting and useful knowledge can sometimes be obtained. Software has been developed for this specific application, and many stunning successes have been reported. This has resulted in the evolutionary development and commercialization of more sophisticated data mining tools using applications of statistical analysis, modeling, artificial intelligence, and neural networks. As favorable results were made known, many agencies, firms, and institutions picked up interest in the

concept, and more software firms became involved in developing improved data mining tools. It is said that there are now more than 100 firms offering such software, consulting, and seminars.^[1]

As pointed out by Simon,^[2] data mining is a highly useful technique to use to help gain deeper insight into the overall picture and general trends described by the data. However, these methods can also uncover patterns that have no real-world meaning or use, so human judgment remains a critical factor. This is sometimes overlooked in the glowing reports of successes with data mining—it is indeed a good tool but not necessarily a panacea.

THE APPLICATION OF DATA MINING TO DECISION-MAKING

As mentioned by Cratochvil,^[3] the whole point of data mining is to reveal hidden information as an aid for prompt and effective decision-making. Action results from the knowledge obtained through the data mining process by discovering non-obvious patterns in existing collections of data. This process is a tool that can be of immense importance to decision makers.

In public administration, data mining or knowledge discovery can be helpful to aid auditors in detecting faulty tax returns, by establishing efficient profiles of non-compliant taxpayers by using models created from historical information. This can help identify the tax returns worth investigating, out of the millions filed. This has proven far more efficient than the previous somewhat hit-and-miss techniques of selecting the returns to be audited.^[4]

In some instances, the operation of court systems has been improved by “mining” the profiles of potential jurors. Similarly, data mining has proved useful in helping establish the relative toxicity of chemicals released into the air, and in predicting the effectiveness of new drugs and medical tests, both of which have obvious regulatory applications.

Data mining is useful also in isolating the factors involved with fraud and abuse in social assistance programs, Medicare and Medicaid, and procurement. Digging into historical records to detect patterns can

prove far more efficient in controlling fraud and recouping money than the previously used human intuition methods of approaching these topics.^[5]

Government departments and agencies can use data mining techniques to help determine how well programs under their jurisdiction are really performing. Law enforcement agencies have made effective use of data mining to find patterns and trends in crime, to help resolve open cases, to target repeat offenders, and to take steps to help prevent crimes. Data mining is also useful in analyzing and predicting traffic flows and growth. These techniques can also be used to help assure that government programs and services are more accessible to the public, and to develop and monitor programs to help meet the needs of stakeholders in the communities served. These are merely a few examples of the potential usefulness of data mining in public management.

There are interesting uses of data mining in commerce and industry, such as determining buying habits of consumers, tracking the effectiveness of advertising, credit risk assessment, and many others. One application that many of us have encountered is the ability of credit card companies, and telephone companies, to detect unusual activity on a specific account, which enables them to quickly contact the holder of the card in question to see if the use was really authorized or if the card had been misappropriated.

DATA WAREHOUSING

Most information systems are designed to facilitate day-to-day operations. However, as pointed out by Gupta,^[6] it is impractical to keep data in operational systems indefinitely. Moreover, the sheer volume of data generated by organizations and agencies makes it difficult at best to effectively extract, analyze, and identify previously unrecognized relationships and trends in historical information. Operational databases usually are not designed for long-term storage of data, or for easy access for queries. As a result, it is now common practice to transfer consistent, cleaned, organized data from operational systems into special archives called data warehouses designed and maintained for historical queries and for data mining.

Straightforward answers to simple or complex questions can often be obtained from archived data through the employment of what are called query tools, the most conventional of which is called structured query language (SQL). This is defined in American National Standard ANSI X3-135-1992 entitled *Database Language SQL*, and the international standard ISO/IEC 9075:1992 entitled *Information Technology-Database Languages-SQL*.^[7,8] A query

defines a set of rules to find the desired information. SQL is a way to communicate directly with a database. The user can specify the type of information that is wanted and can also control which set or sets of data that are to be accessed. Moreover, through SQL, the database can be modified by adding or deleting records, or by creating or modifying tables or other views of information contained within the database.

Beyond the types of questions for which SQL is useful, data warehouses also allow more sophisticated techniques to find otherwise hidden knowledge and patterns through data mining. Query tools and data mining tools are complementary, both employed to extract useful information from archives of historical data but in different ways for different purposes.

Data warehouses can be fed not only from the output of mainframe computers but also from files and spreadsheets generated on desktop and laptop personal computers. Obviously, the design of the data warehouse must take into account the sources of input and the ways the stored information is likely to be used. A variety of hardware and software exists for building, maintaining, managing, and using data warehouses. Also, a wealth of literature exists on data warehousing.^a

According to Siegel,^[14] making data warehousing integral to an organization's daily operational fabric tends to be far more difficult than designing and building the system. Effective data warehousing requires a considerable amount of training on the part of all involved.

In summary, data warehouses serve two goals, one of decision support through ready access to historical information via query tools, and the other the possibility of recognizing otherwise hidden patterns and trends in the accumulated data via data mining.

ONLINE ANALYTICAL PROCESSING (OLAP)

One of the several tools used in data mining is termed on-line analytical processing (OLAP). OLAP is a category of software that permits the use of "multidimensional" databases, i.e., it permits a wide variety of ways of looking at accumulated data to help give meaning and reduce confusion. For example, the dimensions might be time, location, the weather at various points in time, who was on duty, or other such factors that might be significant. OLAP facilitates navigating through large and complex databases. A spreadsheet is considered to be two dimensional, having rows and columns. However, more complicated scenarios

^aExamples of books on data warehousing include Refs.^[9-13].

of data can be considered as cubes composed of cells of information. In the language of OLAP, it is possible to view “cubic” data in many ways, by “drilling down” from higher to lower levels, by “slice and dice” techniques, by pivoting or rotating the cube, etc. These methods of viewing historical accumulated data enable complex queries involving multiple but interacting variables, for effective decision support.^[15] Similarly, OLAP allows decision-makers to ask “What if” questions of the data. There are many vendors of software for variations on the OLAP theme, all of which enable great flexibility in complex analysis of information stored in data warehouses.

VISUAL DATA MINING

It has long been known that seeing data graphically presented allows better and faster comprehension than looking at rows and columns of numbers. Tufte^[16] said that graphical presentations were used long before the advent of computers, dating back to the 1700s, when statistical graphics and Cartesian coordinates were originated. Such graphics as scatterplots, histograms, time-series charts, bar charts, line graphs, pie charts, Pareto charts, boxplots, control charts, and many other versions of two-dimensional representations of data have been and continue to be used to get away from what many consider to be boring tabular presentations. Many of these presentations are easy to achieve on computers. Data maps have also been used extensively for powerful presentation of facts.

The way the data are presented can impact the ease with which decisions are made. At the same time, however, graphical presentations can deliberately or inadvertently mislead the decision maker through deceptive distortions of scale and other visual tricks.

The use of color can make graphical presentations more interesting and can focus attention on important features. Color can also be used to emphasize a point or to identify recurring themes.

One of the major purposes of data mining is to detect unexpected trends and relationships. The human eye and brain are excellent at recognizing patterns, so the use of graphics with data mining is an entirely logical step. As pointed out by Thearling,^[17] useful relationships between variables that are non-intuitive are the jewels data mining hopes to locate, and visualization techniques can greatly aid this process.

Visualizing the data mining models is also a good way to help the decision maker understand and trust the complex models involved. Within the past few years, software developers have come up with many approaches to showing the results of data mining in interesting and helpful graphic presentations. Some of

these tools enable the user to manipulate the variables to better demonstrate the concepts under consideration. Lawler^[18] added that not only can complex patterns be easily spotted, but also, an aesthetically pleasing image can help spark the creativity and curiosity of viewers otherwise unmoved by simply seeing pages of numbers. In addition, visual data can help stimulate discussions and communication between individuals and groups, which can even help bridge gaps when people from different disciplines or different parts of an organization are jointly involved in decision making and problem solving.

In some instances, vivid three-dimensional images of data mining results are possible, yielding a virtual-reality environment. As Eliot^[19] said, visualization presents the data landscape in a natural and intuitive form, making full use of the human capacity to absorb and interact with complex images.

CONCLUSION

Due to the incredible increase in computing power and the accumulation of large quantities of data over recent years, combined with the development of specialized software making data mining feasible, this has become an evolving tool for decision support that has great possibilities in a number of different types of applications. However, data mining requires establishing a well-designed “data warehouse,” which is time consuming and costly to properly set up and maintain. Moreover, because human involvement is required in the interpretation of what is unearthed, considerable training and experience is needed. The successes of data mining are sometimes overstated and exaggerated, making it difficult to determine whether (in any specific circumstances) it is truly worth the effort and the money. But, the possible benefits can indeed be real, making this a topic worthy of careful consideration.

REFERENCES

1. <http://www.acm.org/sigkdd/charter.html> (accessed August 2001).
2. Simon, H. A data mining primer. *Today's Chem. Work* March 2001, 17–21.
3. Cratochvil, A. *Data Mining Techniques in Supporting Decision Making*; <http://cs.unimaas.nl/~kuper/scripties/andac> (accessed August 2001).
4. <ftp://ftp.spss.com/pub/web/wp/TAXWP-1000.pdf> (accessed August 2001).
5. <http://www.spss.com/industries/government/fraud.htm> (accessed August 2001).
6. <http://system-services.com/dwintro.asp> (accessed August 2001).

7. http://www.ora.com/reference/dictionary/terms/S/Structured_Query_Language.htm (accessed August 2001).
8. <http://www.cs.unibo.it/~ciaccia/COURSES/RESOURCES/SQLTutorial/sqlchl1.htm>.
9. Mattison, R.; Alask, R. *Data Warehousing: Strategies, Technologies, and Techniques*; McGraw-Hill: New York, 1996.
10. Inmon, W.H. *Building the Data Warehouse*; John Wiley & Sons: New York, 1996.
11. Agosta, L. *The Essential Guide to Data Warehousing*; Prentice Hall: New York, 1999.
12. Inmon, W.H.; Hackathorn, R.D. *Using the Data Warehouse*; John Wiley & Sons: New York, 1994.
13. Watson, H.J.; Gray, P. *Decision Support in the Data Warehouse*; Prentice Hall: New York, 1997.
14. Siegel, P. Data warehousing. *Chem. Eng.* **2001**, 55, April.
15. <http://dssresources.com/glossary/olaptrms.html> (accessed August 2001).
16. Tufte, E.R. *The Visual Display of Quantitative Information*; Graphics Press: Cheshire, CT, 1998.
17. <http://www3.shore.net/~kht/text/dmviz/modelviz.htm> (accessed August 2001).
18. Lawler, A. New imaging tools put the art back into science. *Science* **2001**, 292, 1044.
19. http://www.labs.bt.com/library/archive/btej_v14_1/part1.htm (accessed August 2001).

Debt Affordability

Bill Simonsen

*Center for Policy Analysis and Management, University of Connecticut,
Storrs, Connecticut, U.S.A.*

Mark D. Robbins

University of Connecticut, Storrs, Connecticut, U.S.A.

Raymond Brown

University of Oregon, Eugene, Oregon, U.S.A.

INTRODUCTION

Municipal bonds are an important way that state and local governments finance their infrastructure improvements. An understanding of how much the government can afford to pay for debt and how much debt it can safely carry is a crucial element in capital and debt planning. The purpose of this article is to describe and analyze the different ways that local governments can measure how much debt they can afford.

BACKGROUND

Debt affordability assessments are intertwined with capital improvement programs (CIP) and debt policy formulation. A CIP is a financial plan that includes an analysis, prioritization, implementation, and financing of infrastructure and other capital needs. The CIP is a multiyear plan that prioritizes and identifies capital projects (as compared to the capital budget, which is the annual spending plan). Governments may have multiple CIPs, each with a different time horizon. The CIP “provides for an orderly and routine method of planning and financing a government’s required capital improvements.”^[1] Borrowing via the municipal bond market is the primary form of capital financing, and planning for it is a key ingredient to a CIP.

Municipal borrowings are generally called municipal bonds, even though some are not technically bonds (e.g., certificates of participation). Municipal bonds are often retired over 15, 20, or even 30 years. The municipal market is very large, with from \$300 billion to \$400 billion sold each year. With this much debt, even small interest rates savings can have a substantial financial impact. The potential financial impact on future generations adds salience to municipal debt decisions. Therefore, debt policies that provide a set of guiding

principles for management decision making are important to the CIP. The Government Finance Officers Association (GFOA) and others have espoused the use of debt policies as an important management tool for state and local governments.

“A carefully crafted debt policy can enhance a government’s credit quality and improve access to both taxable and tax-exempt credit markets.”^[2] Debt policies include such items as when to use a request for proposals; the appropriate conditions for competitive or negotiated sales; proper disclosure procedures; and including citizens in the decision process through public meetings, a vote, or other means, etc.

Larkin and Joseph^[2] advocated that a government’s debt policy should begin with a comprehensive review of affordability. As Larkin and Joseph pointed out, just as when homebuyers make a selection on which home they want based on how much they can afford on a monthly basis, governments need to have a good understanding of debt affordability prior to forecasting needs and issuance. Debt affordability analysis is conducted by calculating a series of ratios and comparing them with established standards that provide guidance about the amount of debt the government can afford. Fig. 1 shows how the CIP, the debt policy, and the debt affordability analysis relate: a debt policy is an important part of a CIP, while debt affordability analysis is an important part of a debt policy.

Debt Affordability and Municipal Credit Analysis

Debt affordability analysis is closely related to the municipal credit analysis conducted by bond-rating agencies. Bond-rating agencies seek to determine the riskiness of the debt based upon a government’s ability to call forth and manage their resources, while governments need to understand how much debt they can comfortably issue. The bond-rating agencies assess

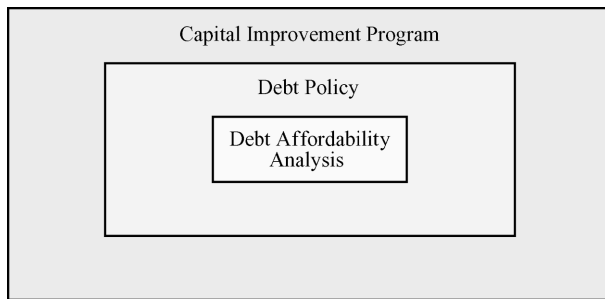


Fig. 1 The relationship between capital improvement programs, debt policies, and debt affordability analysis.

the following four factors to determine credit quality of municipal tax-backed debt:

- *Economy.* Analysis of economic factors consists of an examination of the current and projected future economic profile. This includes analysis of the breadth and depth of the local economy. Measures include trends in unemployment rates; business diversification and industry concentration; trends in the issuance of building permits; and the mix of residential, commercial, and industrial property. Demographic trends such as population growth or decline are also examined. In addition, wealth indicators are analyzed, such as per capita or personal income trends, property value per capita, percent below poverty level, and household income.
- *Financial Condition.* A key to the analysis of fiscal condition is the flexibility of the financial environment. Are revenues limited? What percent of expenditures is mandated or otherwise fixed? Are tax rate adjustments implemented as needed, or do rate increases tend to be traumatic events that are only implemented after reaching crisis conditions? Some broad factors indicative of financial condition are compliance with generally accepted accounting principles (GAAP); diversity and stability of revenue sources; a trend of operating surpluses; and fund balance as a percent of revenues.
- *Debt Position.* Analysis of debt position focuses on the structure and security of the bonds in the context of future borrowing needs and the resources available to pay the current and future debt service. Analysis of debt position is key to debt affordability analysis.
- *Administrative Factors.* The rating agency, Fitch IBCA, views management as a crucial component when evaluating the creditworthiness of a community as they report having seen management salvage, as well as exacerbate, difficult financial situations. "In its future rating assignments, Fitch IBCA will place greater and more specific weight on management practices, both good and bad, that

are employed by issuers in running their financial operations."^[3] Past management performance, in balancing budgets, managing capital needs, etc., is examined. Appropriate capital improvement programs, accurate capital and operating budgets, and prudent investment policies are indicators of sound management practices.

Debt affordability is generally understood in terms of tax-supported debt. A separate type of analysis is conducted for debt backed by other forms of security. General obligation bonds, lease bonds paid from general revenues, certificates of participation, special assessment bonds, and tax-increment bonds are forms of tax-backed securities.

Measuring Debt Affordability

Three key notions comprise debt affordability analysis: 1) ratio calculation; 2) comparison to norms or standards; and 3) forecasting. We will discuss each of these notions in turn.

Ratio calculation

The GFOA recommends that "issuers undertake an analysis of their debt capacity prior to issuing bonds."^[4] The purpose of such a suggestion, according to the GFOA, is that a comprehensive and routine analysis ensures that the government will know all new debt issued is affordable, there is balance between current and future capital needs, and there is the ability to pay for them.^[4] The GFOA suggests that a comprehensive debt analysis should contain the following:

- Measures of the debt service obligation.
- Evaluation of trends relating to the government's performance (revenues, expenditures, and unreserved fund balance).
- Measures of debt burden on the community.
- Statutory or constitutional limitations on debt issuance.
- Market factors affecting tax-exempt interest costs (interest rates, market receptivity, and credit rating).^[4]

Other obligations to consider include unfunded pension liabilities and capital leases.

More specifically, typical measures of debt affordability include the following:

- *Debt service.* Debt service consists of principal and interest payments. It is usually stated in terms relative to total revenues or expenditures.

- *Direct debt (or net direct debt).* Direct debt is a measure of a municipality’s debt paid for by its general revenues including taxes. Bonds that are self-supporting by means other than taxes, such as user fees or charges, are not included in direct debt. As mentioned earlier, tax increment bonds, special assessment bonds, and certificates of participation are included in direct debt, as they are backed by taxes or general revenues.

Using these measures, a variety of ratios can be calculated to put these figures into relative, comparable, terms to other governments. Some of the common ratios are listed below:^a

- *Debt per capita.* Debt per capita is a measure of total indebtedness divided by the population. This standardizes the debt by the population size of the government—ten million dollars might be a lot of debt for Mansfield Connecticut, but a small amount for New York City.
- *Debt as a percentage of real market value.* Debt as a percentage of the real market value is derived by dividing the direct debt by the real market value of all taxable property within the jurisdiction. Because property taxes are typically a primary source of revenues, this is an important measure, as it highlights a municipality’s ability to support its current, and future, debt through the taxing of its property.
- *Debt as a percentage of personal income.* This statistic compares the debt outstanding to a measure of community wealth. This measure is limited in that personal income statistics are often unavailable for smaller jurisdictions.
- *Annual debt service as a percentage of general and debt service fund expenditures.* This measure identifies the portion of the general and debt service funds that goes to debt service.
- *Annual debt service as a percentage of operating expenditures.* This concept is broader than the last, as it includes expenditures from all operating funds, such as the general fund, special revenue funds, and debt service funds.
- *Impact of the debt on the tax rate.* This measures the portion of the tax rate that is dedicated to servicing tax-supported debt.
- *Impact of the debt on the average household.* This statistic measures the monetary burden on the average household in the community.

It is important to also include overlapping debt when calculating the principal amount outstanding. Overlapping (or sometimes underlying) debt is the

tax-supported principal owed by other governments that overlap (or underlie) the geography of the jurisdiction conducting the debt affordability analysis. For instance, if it is a city conducting the analysis, the overlapping debt would include the proportionate debt of the county, school districts, and special districts. Overlapping debt is calculated as follows:

$$OD = \frac{MV_O}{MV_T} * TD$$

where:

OD = Overlapping debt.

MV_O = Market value of the overlapping property.

MV_T = Total market property value of the overlapping government(s).

TD = Tax-supported debt outstanding for the overlapping government(s).

The notion behind overlapping debt is that the taxpayers of a jurisdiction are responsible for the debt of all their governments, not just the one conducting the debt affordability analysis. Debt as a percentage of market value that includes overlapping debt is called debt burden.

Table 1 summarizes the different statistics that typically can be used to measure debt affordability.

Comparison to norms or standards

The debt statistics and ratios are difficult to interpret on their own. Is a debt burden of 5% high, low, or

Table 1 Debt affordability statistics

Debt service
Direct debt
Debt per capita
Debt per capita including overlapping debt
Debt as a percentage of real market value
Debt as a percentage of real market value including overlapping debt (also called debt burden)
Debt as a percentage of personal income
Debt as a percentage of personal income including overlapping debt
Debt service as a percentage of general and debt service fund expenditures
Debt service as a percentage of operating expenditures
Impact of the debt on the tax rate.
Impact of the debt on the average household

^aThis section draws significantly from Ref.^[4], pp. 15–18.

moderate? Is a debt per capita of \$5000 a lot, or just a little? It helps to interpret the statistics by comparing them to some norm or standard. Examples of standards are rules of thumb, medians, and benchmarks.

Rules of thumb are industry standards for certain ratios. Some rules of thumb standards used by Standard and Poor's^[5] are as follows:

- Debt per capita (including overlapping debt) is low if it is below \$1000, moderate between \$1000 and \$2500, and high above \$2500
- Debt burden is low if less than 3%, moderate from 3 to 6%, and high if over 6%
- Debt service as a percent of operating expenditures is low if 5% or less, moderate if 10%, and high if over 15%.

Debt statistics outside of the range of the rules of thumb are not necessarily dangerous, because governments have different service responsibilities. For instance, a fire district with volunteer firefighters may have a high debt service as a percent of operating expenditures, because there are very low labor costs.

Medians can help refine government's understanding of its debt position. Medians are generally calculated by population and government type, e.g., median debt burden for counties with a population from 100,000 to 249,999. In the past, Moody's Investors Service published medians for selected debt statistics, but they no longer do so.

Another way to refine the debt affordability analysis is to compare debt statistics with those of similar governments, or benchmarks. While the use of benchmarks can provide important comparison information, there are two basic difficulties with this approach. First, no two communities are exactly alike, so the choice of comparable governments must be done carefully. Important characteristics to match when choosing comparison governments are population, revenue mix, service responsibilities, bond rating or credit quality, and whether the laws governing the use and issuance of debt are comparable.

Second, ratios are not consistently measured across organizations. Some governments deduct funds set aside for principal payment from the outstanding debt prior to calculating ratios. For instance, some governments calculate ratios based only on general obligation debt and ignore other tax-supported debt. One of the reasons for this is that state statutes or constitutions often define the way debt is to be calculated for the purpose of state debt limits. These calculations generally exclude lease debt, COPs, or any debt considered self supporting. This may be appropriate for this purpose, but provides an inadequate

and incomplete view of the debt picture for debt affordability analysis.

Forecasting future capacity

Much of the value of debt affordability analysis is to estimate how much debt a government can afford to issue into the future. Estimating debt affordability into the future requires multiyear projections for the debt statistics. This consists of forecasts of the information that are part of the ratios, along with projections of future capital needs, such as the following:

- Population.
- Operating revenues.
- Operating expenditures.
- Real market value of property.
- Personal income (if available).
- The size of the potential issues (annual dollar amounts to be issued).
- Debt maturity schedules.
- Type of debt service structure; (level debt service, equal principal, other)
- Interest rate schedule.

Different scenarios can be modeled, given different capital expenditure requirements or debt issuance structures. Constant annual growth rates applied to a base year is the way many of these parameters were estimated by Miranda and Picur^[4] for the city of Pittsburgh. Sometimes the baselines for the ratios, such as revenues, expenditures and real market value, are forecasted for other budgeting purposes.

There is uncertainty anytime projections are made. With this in mind, it is often prudent to include sensitivity analysis for varying rates of growth for the relevant factors. Once the forecasts are accomplished, the government can decide whether it can afford to pay for the debt issuance, or if the capital program needs to be scaled back. In this way, the debt affordability analysis provides important information for the government's CIP. Spreadsheet programs, such as Excel, are well suited for this future analysis.

CONCLUSION

In many ways, deciding upon debt affordability criteria is an art and not a science. Ratios can be calculated, but they need to be interpreted to be useful. Benchmarks can provide an indication of what constitutes high, or low, ratios for key debt statistics. However, because there are multiple ratios, it is possible to be low for some, but high on others. In this case, the government must decide whether that is acceptable,

or whether efforts must be made to lower the troublesome ratios.

Ratios or benchmarks alone do not tell the whole story. Other factors, such as strong economics or history of strong management, may offset high ratios. That is, a wealthy or diverse tax base and a history of sound management practices may provide a government with a better chance to withstand and recover from problems despite poor ratios.

Debt affordability analysis is an important part of a debt policy, which in turn, is an important part of a capital improvement program. Debt affordability analysis, debt policies, and capital improvement programs are an important part of any government's financial planning.

REFERENCES

1. Joseph, J.C. *Debt Issuance and Management: A Guide for Smaller Governments*; Government Finance Officers Association: Chicago, IL, 1994; 5 pp.
2. Larkin, R.; Joseph, J.C. Developing formal debt policies. *Gov. Finance Rev.* **1991**, 7 (4), 11–14.
3. Larkin, R.C. *Impact of Management Practices on Municipal Credit*; Fitch IBCA: New York, 2000; 1 pp.
4. Miranda, R.A.; Picur, R.D. *Benchmarking and Measuring Debt Capacity*; Government Finance Officers Association: Chicago, IL, 2000; pp. 5, 6.
5. Standard & Poor's Public Finance Ratings Criteria. 2000; 29. Downloaded 9/5/01 from <http://www.standardandpoors.com/ResourceCenter/RatingsCriteria/PublicFinance/index.html#a>.

Debt Issuance

Beverly S. Bunch

*Director of Public Administration Program, PAC 324, University of Illinois at Springfield,
Springfield, Illinois, U.S.A.*

INTRODUCTION

Debt issuance occurs when a government borrows money and agrees to repay the money with interest at a later time. This practice is similar to when an individual borrows money to pay for a house and then makes monthly mortgage payments.

State and local governments borrow money for a variety of purposes. One of the major uses of debt is to finance large capital projects, such as the construction of a street, a public building, or a water treatment plant. Using debt to finance these types of projects allows the costs of the project to be spread over time rather than being paid in a single year. This can help make a project more affordable and allow the government to avoid large fluctuations in tax rates. The use of debt also helps facilitate intergenerational equity in that there is a closer match between those who use the facility over time and those who pay for the facility.

Some governments also issue debt when there is a mismatch between the timing of revenues and expenditures. For example, a government may not receive the majority of its tax revenues until the third month of the year; however, expenditures, such as payroll, will have to be made prior to that time. The government may borrow funds to pay for its expenditures and then repay the debt when the tax revenues arrive. This type of debt is referred to as tax anticipation notes (TANS) or tax and revenue anticipation notes (TRANs), if other types of revenues also are used to repay the debt.

Historically, some governments have issued debt to address a deficit situation in which revenues are less than spending. However, this practice is not considered financially prudent, because future generations have to pay for services that were provided in the past. Investors also may be reluctant to lend funds to a government that is having financial difficulties.

Governments may also issue debt to finance a loan program, such as a housing mortgage program or a student loan program. In this case, the government issues debt and then uses the money to make loans to private individuals or organizations. The government uses the funds from the loan repayments to repay the original debt.

TYPES OF DEBT

State and local government debt can be classified in a variety of ways. One form of classification is based on the length of time the debt is outstanding. Short-term debt typically refers to debt that will be repaid in one year or less, while long-term debt refers to debt that will be outstanding for a longer time period, often 20 years. Short-term debt is referred to as notes, while long-term debt is called bonds.

Another way to categorize debt is based on the security associated with debt. The most secure type of debt is general obligation debt, which is guaranteed by the full-faith-and-credit of the government. This means that investors legally can require the government to utilize its taxing authority to generate sufficient funds to repay the debt. In contrast, revenue bonds are secured solely by a specified revenue source, such as water user charges or tolls collected on a toll road. Because revenue bonds do not have the full-faith-and-credit backing of the government, they tend to have higher interest rates than general obligation bonds. However, many governments find revenue bonds attractive for the financing of projects that are expected to be self-sustaining.

Another way to classify governmental debt is by the purpose for which the bond proceeds will be used. Public purpose debt is used to finance infrastructure that will be owned by the government, such as public roads or buildings. Alternatively, a state or local government can issue private activity debt on behalf of a private organization or group of individuals who will be the primary beneficiaries from the bonds or will have primary responsibility for generating the funds to repay the bonds. Examples of private activity debt include housing mortgage revenue bonds, student loan bonds, and industrial development bonds.

Governments may issue debt to refinance existing debt. These bonds are referred to as refunding bonds, while debt issued to finance new projects is referred to as new money bonds. Similar to the refinancing of a house mortgage, a government may refinance debt when interest rates are significantly lower than when the debt was originally issued. A government also may choose to refinance debt when it wants to change some feature of the debt, such as the timing of the

principal repayments or the legal requirements associated with the debt.

TAX-EXEMPT STATUS OF DEBT

Most debt issued by state and local governments is tax exempt. This means that the interest payments made to investors are not subject to federal and, in some cases, state income taxes. As a result, investors are willing to lend funds to state and local governments at lower rates than if the interest was taxable. This feature of the federal income tax code results in significant interest savings to state and local governments.

The federal government has imposed a number of constraints to avoid abuse of the tax-exempt provision. One of the major provisions relates to restrictions on arbitrage earnings. Arbitrage is the difference between the interest rate paid on the bonds and the interest rate earned on the borrowed money until it is spent. The federal government also has imposed a state cap on the amount of tax-exempt private activity bonds that can be issued each year within each state.

Some governments issue taxable debt when they want to finance a project or program that does not satisfy the eligibility requirements for tax-exempt debt.

VOTER REFERENDUMS AND DEBT LIMITS

Most state and local government general obligation debt is subject to approval by the voters and also may be subject to some form of legal debt limitation. At the local level, the limit may be imposed by the state government or a local charter. For example, some states restrict the amount of local debt outstanding to a specified percentage (e.g., 5%) of the taxable property value within the jurisdiction. At the state level, a constitutional or statutory debt limit may restrict the amount of debt to a particular dollar amount or to a specified percentage of state revenues. Alternatively, the debt limit may indicate the maximum percentage of general revenues that can be used to repay the debt.

Revenue bonds usually are not subject to voter approval requirements or debt limitations, because they are not backed by the taxing power of the government.^[1] A special type of revenue bond, called a lease revenue bond, also is not usually subject to voter approval and debt limitations, even though the debt may ultimately be repaid with tax revenues. With this type of debt, a government utilizes a separate agency or a non-profit entity to issue bonds that are used to acquire or construct a facility to be used by the government. The facility is then leased to the government, and the agency or non-profit entity uses the lease payments to repay the bonds. The lease revenue bonds typically

include a non-appropriation clause that allows the government to terminate the lease if funds are not appropriated for the lease payments. As a result of this clause, most state and local governments' laws do not legally classify lease revenue bonds as long-term debt.^[2]

STRUCTURE OF DEBT SERVICE

Debt service refers to the payment of principal and interest on bonds. The structure of debt service addresses issues such as the timing of principal and interest payments and the process by which interest rates are set.

Most government debt is structured as serial bonds, in which a portion of the principal is repaid each year over the life of the debt. The interest rate on serial bonds varies by maturity (the date on which the principal is repaid to investors), with the shortest-term series usually having lower interest rates than the longer-term series. Some governments also issue term bonds, in which the principal is due in one or two lump sum payments in the final years the bonds are outstanding. Throughout the life of the term bonds, the government makes interest payments and transfers funds into a sinking fund to generate sufficient funds for the eventual repayment of the principal. Governments also may issue bonds that are a combination of serial and term bonds.

Government bonds are usually structured as level debt service or level principal payments.^[3] Level debt service is similar to how mortgage payments are structured in that the total amount of principal and interest paid each period is equal. The advantage of a level debt service structure is that it provides stability in the budget, because the same amount is due each year. In contrast, under a level principal structure, the amount of principal repaid each period is equal, and the debt service payments decrease over time.

Bonds can be structured with fixed-rate interest rates or variable-rate interest rates. With fixed rates, the interest rate on each serial payment is set at the time the bonds are issued, while with variable rates, the rates typically vary over the life of the bonds. In the latter case, the interest rates can be tied to some type of market index and are reset at specified intervals, such as daily, weekly, monthly, or annually.

Bonds also may be structured with a call option, which allows a government the option of repaying the debt prior to maturity. This may be helpful for a government that would like to be able to refund bonds in the future to take advantage of lower interest rates. However, the inclusion of a call option typically results in a slightly higher interest rate on the original bonds.

DEBT ISSUANCE PROCESS

Professional Advisors

State and local governments typically receive assistance in the debt issuance process from a team of professionals, including a financial advisor, a bond counsel, and underwriters.

The role of the financial advisor is to assist the state or local government throughout the debt issuance process. This may include helping the government decide what type of debt to issue, how much debt to issue, and the necessary steps to obtain the appropriate legal approvals. The financial advisor also may help the issuer structure the debt and prepare the official statement. The official statement is a document for investors that describes the bond issue and the financial condition of the government issuing the bonds. The financial advisors also may assist in selecting the other team members, obtaining a bond rating, and evaluating whether bond insurance is financially advantageous.

The bond counsel helps the government ensure that all legal provisions related to the issuance of debt are followed. This includes making sure that the debt is legally authorized, drafting legal documents, obtaining the necessary approvals, and helping the government adhere to all relevant legal procedural provisions. The bond counsel also is responsible for evaluating and certifying that the debt satisfies the legal requirements for being considered tax exempt.

The role of the underwriters is to identify investors willing to lend the government money. The underwriter purchases the bonds from the government and then sells them to investors. In a negotiated sale (explained below), underwriters may also be involved in many of the same types of activities as the financial advisor.

Method of Sale

Bonds typically are sold through competitive bid or a negotiated sale. In a competitive bid sale, underwriters or groups of underwriters called syndicates submit concealed bids specifying an interest rate for each serial maturity of the bonds. The bonds are then awarded to the underwriter or syndicate who submitted the bid that will result in the lowest overall interest cost for the government.

The bonds are typically awarded based on the calculation of the true interest cost—the TIC. The TIC takes into account the amount of interest paid, as well as the timing of those interest payments. Some governments used to (and some still do) award bonds on the basis of the net interest cost (NIC), which is based on the total interest payments but does not take into account

the timing of the interest payments. The TIC is preferred over the NIC, because the TIC takes into account the time value of money.^[4]

In a negotiated sale, an issuer selects an underwriter or syndicate well in advance of the sale date to help the issuer with the marketing and sale of the bonds. The issuer negotiates with the underwriter or syndicate to determine the underwriter's spread, which is the underwriter's compensation, and an interest rate for each maturity in the series. After an underwriter or syndicate purchases a government's bonds, the underwriter then sells the bonds to investors.

There has been an ongoing debate regarding whether it is better for a government to sell bonds through competitive bid or through a negotiated sale. Some professionals believe that the competitive bid process is more likely to result in a lower interest cost, especially for bonds that are "plain vanilla."^[5] This means that the issuer is known in the market and that the bonds do not have a complicated structure or security. Competitive bid also has an appearance of fairness, in that the bonds are awarded through a sealed bid process to the underwriters that submit the bid with the lowest interest cost.

However, in some cases, critics note that a negotiated sale may be preferable. With a negotiated sale, the underwriters know in advance of the actual sale that they will be awarded the bonds and, therefore, may be more willing to invest in presale marketing of the bonds. If they are able to find investors for most of the bonds prior to the setting of the interest rates, they may be less inclined to incorporate a risk component into the negotiated interest rates.^[6] Negotiated sales also allow underwriters flexibility in structuring the bonds (such as maturity schedules and call features) and in selecting a sale date depending on the condition of the market. In a competitive bid sale, the structuring of the bonds and sale date have to be decided and announced in advance. Another potential advantage is that a negotiated sale tends to allow a government more flexibility in pursuing policy objectives, such as making bonds available to citizens as investments or facilitating participation by locally-owned businesses or minority- or women-owned businesses.^[7]

However, the increased discretion in selecting an underwriter for a negotiated sale raised some concerns about the relationship between underwriters' contributions to political campaigns and the awarding of bonds by elected officials who received those funds. In response, the Municipal Securities Rulemaking Board established Rule G-37, which states that municipal securities brokers and dealers generally are prohibited from engaging in municipal securities business with an issuer within 2 years after having made a contribution to that official.

Bond Ratings

Most governments that sell bonds in the national market obtain a bond rating prior to the sale. A bond rating is the assignment of a grade by a private bond rating firm as to the creditworthiness (risk of default) associated with a particular bond issue. The bond rating firms utilize grades ranging from AAA to C. Higher-rated bonds typically receive lower interest rates than lower-rated bonds due to the lower perceived probability of a default.

Analysts for the bond rating companies assess the structure of the particular bond issue and the financial security backing the bond. For a general obligation bond, the bond rating analyst also addresses the issuer's managerial capabilities, debt burden, financial and operational factors, and the economic conditions of the government's jurisdiction. For a revenue bond, the bond rating analyst considers other factors such as the coverage ratio (the ratio of net revenues to annual debt service) and the structure of a reserve fund to pay for debt service if sufficient revenues are not available.

Bond Insurance

In some cases, a government may find it advantageous to purchase bond insurance from a private company. In exchange for an insurance premium, the bond insurance company agrees to pay the debt service if the government is unable to make the payments. For insured bonds, the bond rating is based on the bond rating for the private bond insurance company rather than the government issuing the debt. Historically, the major municipal bond insurance companies have been rated AAA, the highest bond rating. In assessing whether bond insurance is prudent, the government should weigh the cost of the insurance premium versus the present value of the expected decrease in interest costs associated with a higher bond rating.

Use of Technology

Governments are beginning to increase their use of technology for activities related to the issuance of municipal bonds. Recently, some governments have sold their bonds through a bond auction held on the Internet. This has allowed governments to try innovative features such as selling serial bonds maturity-by-maturity rather than on an all-or-none basis and selling bonds directly to large institutional investors. The City of Pittsburgh, one of the pioneers in this area, notes that it has obtained significant savings by using the Internet for bond sales.^[8]

DISCLOSURE

State and local governments that issue debt are obligated to provide investors with relevant and timely information about the bonds and the financial condition of the government. Although there have been few government bond defaults, situations such as the Washington Public Power Supply System bond default in the 1980s and the financial difficulties of Orange County, California, in the 1990s, have drawn increased attention to the importance of having appropriate disclosure.^[9]

The Government Finance Officers Association (GFOA), which is a membership organization of state and local budget and finance officials, has established *Disclosure Guidelines* to assist state and local governments with the disclosure of appropriate information related to the issuance of debt. The Securities and Exchange Commission Rule 15c2-12 also requires underwriters of state and local government bonds to adhere to certain reporting requirements. In addition, some governments are utilizing their web sites to make relevant information available to investors before and after the sale of bonds. The latter is especially important, because many state and local government bonds are traded in a secondary market after the initial sale.

REFERENCES

1. Petersen, J.E.; McLoughlin, T. Debt policies and procedures. In *Local Government Finance: Concepts and Practices*; Petersen, J.E., Strachota, D.R., Eds.; Government Finance Officers Association: Chicago, 1991; 263–292.
2. Aguila, P.R., Jr.; Petersen, J.E. Leasing and service contracts. In *Local Government Finance: Concepts and Practices*; Petersen, J.E., Strachota, D.R., Eds.; Government Finance Officers Association: Chicago, 1991; 321–338.
3. Robbins, M.D.; Simonsen, B.; Jump, B., Jr. Maturity structure and borrowing costs: the implications of level debt service. *Munic. Finance J.* **2000**, *21* (3), 40–63.
4. Mikesell, J.L. *Fiscal Administration: Analysis and Applications for the Public Sector*, 5th Ed.; Harcourt Brace College Publishers: Fort Worth, 1999; 533–584.
5. Simonsen, W.; Kiltredge, W.P. Competitive versus negotiated municipal bond sales: why issuers choose one method over the other. *Munic. Finance J.* **1998**, *19* (2), 1–29.
6. California Debt Advisory Commission. *Competitive Versus Negotiated Sale of Debt*; Issue Brief; California Debt Advisory Commission, September, 1992; Vol. 1.
7. Bunch, B.S. Increasing women and minority participation in municipal debt issuance: an analysis of Texas' experiences. *Munic. Finance J.* **1997**, *18* (3), 1–20.
8. Hennigan, P. Municipal bond sales via the internet: the story of Pittsburgh's electronic auction. *Gov. Finance Rev.* **1998**, *14* (4), 23–27.
9. Fairchild, L.M. Are federal disclosure requirements beneficial for the municipal bond market? *Munic. Finance J.* **2000**, *21* (3), 65–82.

Decentralization in Developing Countries

Scott Fritzen

Lee Kuan Yew School of Public Policy, National University of Singapore, Singapore

Patrick W. Ong

Peace and Equity Foundation, Quezon City, Philippines

INTRODUCTION

Decentralization as an approach to development administration occupies an important conceptual position within the development discourse. Manor sums up the situation nicely, in so doing emphasizing the rhetorical importance of the concept to developing country governments themselves:

[Decentralization] is being considered or attempted in an astonishing diversity of developing and transitional countries—by solvent and insolvent regimes, by democracies (both mature and emergent) and autocracies, by regimes making the transition to democracy and by others seeking to avoid that transition, by regimes with various colonial inheritances and by those with none. It is being attempted where civil society is strong, and where it is weak. It appeals to people of the left, the center and the right, and to groups which disagree with each other on a number of other issues.^[1]

Even officials in Myanmar, one of the world's most centralized states, have expressed the need for reforms involving a degree of decentralization.^[2] Decentralization has been applauded for its supposed potential to improve levels of public participation, bureaucratic accountability, administrative efficiency, and responsiveness to local needs, among other goals.

Yet, the assessment of decentralization in a significant share of the academic and practitioner literature has shifted from marked optimism to one of caution, even pessimism. Since the 1990s, except for some overly sanguine donor reports, the “decentralization-is-problematic” argument predominated in the academic literature. Analysts typically point to one or more dangers such as increasing inequality, the empowering of local elites, political instability, and general ineffectiveness.

This review begins by presenting a means of classifying the different forms of decentralization and their ideological underpinnings. The roots of decentralization's renaissance are then traced to a complex set of political realities and motivations in the centralized governments that undertook this reform. Some problematic issues and controversies surrounding decentralization are explored in the concluding section.

DISSECTING DECENTRALIZATION

The most common typologies of decentralization attempt to distinguish between the various functions and/or resources being decentralized. One often-cited classification presumes four broad categories of decentralization: administrative; fiscal; political; and market.^[3]

Administrative decentralization refers to the transfer of policy making, planning, and management responsibilities from central to local levels. Distinctions are often made between two subtypes:

- **Deconcentration**—in which branches of the central government are geographically dispersed, but no real authority is transferred to lower levels. If the Ministry of Finance opens offices at the district level, in an attempt to improve tax collection in problematic districts, this would be an example of deconcentration.
- **Delegation**—in which authority for certain functions are transferred to lower levels, which remain substantially accountable to (but not directly controlled by) the central level.

Fiscal decentralization refers broadly to efforts to change the distribution and sources of resources available to local governments. Such efforts can take many forms, including transfers between levels of government, authorization of local borrowing, cost recovery, and changes to revenue sources available to local governments through taxes, user fees, and contributions.

Political decentralization (or devolution) refers to attempts to devolve powers to democratically elected local governments or, in much weaker forms, attempts to make local governments more accountable to communities through the establishment of oversight boards or the introduction of new forms of citizen participation in development projects and policy making.

Finally, *market decentralization* involves attempts to transfer substantive control over resource allocation to non-state actors. Privatization is the obvious example, and it can apply not only to state-owned enterprises but also to broad swathes of the economy.

Table 1 Nature and basis of decentralization

Nature of delegation	Basis for delegation	
	Territorial	Functional
Within formal political structures	Devolution (political decentralization, local government, and democratic decentralization)	Interest group representation
Within public administrative or parastatal structures	Deconcentration (administrative decentralization and field administration)	Establishment of parastatals and quangos
From state sector to private sector	Privatization of devolved functions (deregulation, contracting out, and voucher schemes)	Privatization of national functions (divestiture and economic liberalization)

(From Ref.^[4])

Another useful typology formulated by Turner and Hulme looks at the nature and basis of the transfer of authority innate in any form of decentralization.^[4] Normally, decentralization is based on territorial considerations, with the aim of bringing authority geographically closer to both the front-line bureaucrats and the public they service. In less common instances, transfers can be made on a functional basis, which results in transferring authority to specialized agencies. Thus, decentralization could have either a territorial or a functional basis.

Across these two categories, the nature of authority transfer could be any of three manners: delegation within formal political structures; transfer within public administrative or parastatal structures; or transfers to nonstate agencies. This classification yields the six forms of decentralization shown in Table 1.

The two sets of typologies presented here help clarify the concepts surrounding decentralization and enumerate the various ways in which it can be implemented. Not all experiences of decentralization, however, will fit tightly into these categories, as decentralization has often been implemented as a hybrid of these forms. The different manifestations of decentralization can be rooted in a wide range of ideologies, goals, and political contexts, which have served as distinct motivations among the central governments and multilateral aid agencies advocating decentralization.

PUTTING DECENTRALIZATION INTO CONTEXT

The 1980s saw the beginning of a wave of decentralization that had swept across the globe in the following two decades. What accounts for this growing prominence of decentralization as both a reform program and category of analysis? One factor is that interest in decentralization had, and still has, varied roots, as seen in the strong ideological bend of much writing on decentralization. Fig. 1 points to some of the aims

typically associated with different types of decentralization reforms. These include “economistic” interest in efficiency, classic “leftist” concerns with participation and democracy, and “rightist” interest in reducing state intervention in the economy.

Various regimes have used the ideologies presented above to explain their push toward decentralization, but, in addition to efficiency or empowerment rationales, central authorities have of course also driven by more self-serving motivations.

In his analysis of the political economy of decentralization, Manor^[1] pointed out that many centralized regimes had to engage in some form of decentralization, as a response to the public’s loss of confidence in their governments. Such a loss was often the result of the degeneration of existing patronage systems and political parties. “Regimes faced mounting demands from organized interests, yet sluggish economic growth and increasing corruption by political activists at all levels undermined their ability to respond to those demands” (on page 27 of Ref.^[1]). Decentralization thus became widely appealing to politicians, as a means of coping with an eroding centralized system, as it pushed the responsibility of responding to some of the public’s mounting demands to lower levels of authority.

Furthermore, the lack of conceptual clarity in the forms of decentralization within the earlier decentralization discourse debate enabled authoritarian regimes to borrow rhetoric associated with political



Fig. 1 A spectrum of ideological underpinnings of decentralization.

forms of decentralization, while pursuing administrative reforms requiring little systemic change and tending to strengthen central monitoring of peripheral regions.

ISSUES AND CONTROVERSIES IN DECENTRALIZATION

Does Decentralization Improve Service Delivery and Accountability?

Decentralization is at the heart of a range of reforms, seeking to improve service delivery through paving a “short road to accountability.”^[5] Citizens under a centralized regime normally have to voice their demand for better services to central authorities who, in turn, direct local-level bureaucrats to respond. By bringing policy makers closer to the public, decentralization makes it easier for citizens to voice their demands on government, as well as better monitor the performance of the responsible politicians or bureaucrat. Policy makers are also better able to monitor their subordinates who are in the front lines of service delivery.

There are several successful experiences, wherein decentralization played an instrumental role in improving basic service delivery, such as the experience of 16 municipalities in Columbia, where water, education, and road infrastructure improved. Another example is the City of Porto Alegre in Brazil, which experienced a doubling of enrollment rates and improved access to basic sanitation.^[6] Yet, while the available empirical evidence suggests a positive impact of decentralization on service delivery, the evidence is insufficient to conclude causality, as decentralization often coincides with other reform programs.^[7] Furthermore, since these studies did not collect household-level data, the impact of decentralization across different socioeconomic classes cannot be determined.

Accountability is also intended to improve with decentralization, as local political competition and the public's greater ability to monitor local officials can improve performance and lessen corruption. But these positive effects can be offset by several risks faced in the decentralization reform process.^[8] In places where powerful local elites exist, decentralization could serve as a vehicle for greater consolidation of their power and influence, leading to the eventual capture of the state. There is also the risk of expanding and further enrooting clientelist networks and patterns of patronage politics. Local politicians could, instead, choose to cater to a small subsection of the electorate who could be counted on for loyalty and political support, come election time. There is a growing consensus that context plays an important role in how decentralization affects accountability. The strength and influence of competing interest groups, levels of

economic and social inequity, and the capacity of institutional counterchecks on government (local legislature, judiciary, media, and civil society), among other factors, affect the impact of decentralization on accountability relationships.

Indonesia offers a good example of the ambiguity of a major decentralization reform. While its experience of “big bang” decentralization has clearly opened up new spaces for popular participation in political debate and for the watchdog groups in civil society, there has also been evidence of every one of the major problems associated with decentralization, such as “decentralizing corruption,” increased inequalities between resource-rich and resource-poor regions, and failure of the central government to successfully set and enforce minimum service standards in critical areas of national priority.^[9]

The World Bank suggests two conditions that allow political decentralization to bring about the promised improvements in basic social services: “First, voters must be more likely to use information about the quality of local public goods in their voting decision. Second, local political promises to voters must be more credible than regional and national promises” (on pages 89–90 of Ref.^[10])

Threat of Macroeconomic Instability

Fiscal decentralization entails a careful balancing act between the revenues and expenditures transferred from central to local governments. Decentralization can undermine the fiscal position of local governments with respect to transferred functions, leading to a deterioration of government service quality. On the other hand, increased central-level transfers can undermine local revenue-raising effort.

Aside from affecting service delivery and local taxation effort, fiscal decentralization also affects the macroeconomic stability of a country, as poorly planned fiscal decentralization can lead to uncontrolled local government budget deficits. The experiences of four Latin American countries underline the necessity of careful macroeconomic fiscal management. The transfer and expansion of local-government expenditure in Argentina, Brazil, Columbia, and Mexico, amidst fiscal decentralization, have made it more difficult for the central governments to control government budget deficits and stabilize inflation during the Latin American economic crisis in 1994–1995.^[11]

Litvack et al. argued that proper decentralization design could mitigate these problems.^[12] Some of the recommendations they offer include:

- Expenditure responsibilities among different levels of government should be clearly stated.

- Central-government transfers should be made transparent and predictable, ideally based on a simple and standard distribution formula.
- Local government should face restrictions in borrowing.
- Local-government borrowings should not be subsidized or guaranteed by the central government.

Does Decentralization Increase Fiscal and Capacity Disparities?

There is substantial evidence that significant administrative, political, or administrative decentralization can increase fiscal and capacity differentials between administrative units and local governments. Local governments differ greatly in their ability to raise revenue; hence, greater reliance on local sources will, in the absence of significant central-government transfers, raise disparities. This was the case in Indonesia, where even after central-government transfers, the poorest region has only 20% of the average regional per capita revenue.^[9] Hofman and Guerra also identified the existence of large fiscal disparities in Vietnam and China, linked to changing patterns in the financing of subnational governments.^[13] Their analysis shows significant but weak correlation between the levels of expenditure and the level of service delivery (using health and education indicators as a proxy). While their study does not show that decentralization had actually increased these disparities in expenditure and service levels, this is a reasonable conclusion.

Political decentralization expands the responsibilities of local government from just program implementation and service delivery to include policy-making. This necessitates a greater range of skills and capacities that local governments need to employ in order to fulfill their functions. These include technical and managerial skills, of course, but also extend to the sense of the legitimacy of local government in the eyes of the citizenry and its capacity to engage local populations in problem solving (which can depend both on government capacity and attitudes and on the density and quality of civil society locally). The point is that in all of these capacity areas above, we can expect local governments to vary fairly dramatically. This may not be a problem for large urban governments, as resources are more available to them, but it may pose a serious problem for small and far-flung local governments with limited financial and human resources.

This does not necessarily imply that service delivery performance itself, let alone development outcomes, diverged as dramatically; correlation between levels of expenditure, service delivery patterns, and outcome indicators in health and education, for instance, has

often been shown to be weak. Central-government efforts to address this issue have revolved around attempts to equalize fund transfers to local governments. Fiscal disparities across local governments decreased (to varying degrees) after fund transfers from central governments, as data from China, Indonesia, Philippines, and Vietnam indicate.^[13]

These considerations should not be taken to mean that decentralization always or necessarily fails in relatively poor areas of a country.

- One of the poorest provinces in Brazil was able to achieve impressive boosts in performance through creative local management, which Brazil's federal structure had, in turn, enabled.^[14]
- Fritzen^[15] modeled some of these complexities for Vietnamese provinces and districts, showing how the responses of, and the effects on, local governments of a central-government initiative to decentralize social policy functions varied dramatically. Fiscal, administrative, and political capacities, together with local socioeconomic conditions, all helped to predict the success of the scheme, but some poor local governments were able to find creative ways to overcome resource constraints.
- In Columbia, service delivery improved or remained the same in all of the 16 municipalities, where case studies of decentralization were undertaken following a decentralization initiative, notwithstanding well-documented, widening disparities.^[16]

Asymmetric decentralization has been proposed as an alternative framework to deal with capacity disparity problems associated with decentralization.^[5,12] The concept refers to the possibility of decentralization being applied to different degrees across subnational units, depending both on demand for the reform and on local-government capacity. The weakness of such a proposal is the complexity inherent in its implementation, which may be beyond the capacity of many central governments to effectively carry out.

How to Strengthen Local-Government Capacity

An important segment of the decentralization debate focuses on local capacity development framework. This involves specific attempts to develop theories about local government, which will facilitate central and donor interventions to support local capacity and institutional development. The major distinction among approaches to capacity building, which the emergence of structural and New Development Administration theorists has highlighted, is that between “supply-” and “demand-driven” orientations.

Supply-driven theorists tend to view local-government deficiency in capacity as a function of limited human and financial resources, which should be remedied, in Rondinelli's revealing phrase, "tutorially," i.e., gradually through patient central government and donor intervention, if possible *prior* to initiating decentralization.^[3]

In contrast, theorists such as Litvack think that the problem of whether local governments are capable of responding to central requirements and tasks is related at least as much to program design. "If bureaucratic requirements imposed by the center were more appropriate," argue Litvack and colleagues, "they might be less of a problem." Korten and Ostrom, from significantly different theoretical starting points, both emphasize the need to conceive of central level as creating *facilitative conditions* for local capacity development, as opposed to "supplying" the capacity itself.^[17,18] Fiszbein argues that local capacity development involves the promotion of "innovative and responsible leadership and civic involvement."^[16] These writers thus emphasize local *demand* for capacity development; this is basically a way of saying that "shifting responsibilities to lower-tier governments may provide the incentive for public officials to invest in capacity building or seek creative ways to tap into existing sources of capacity."^[12] Another aspect of this demand-driven approach is information sharing among local governments in service of a learning-by-doing strategy.^[19]

The Center's Role in Decentralization

The central government has a strong role to play in all systems, decentralizing or not. It provides overall policy direction, defines minimum standards of service delivery, transfers technical and fiscal resources to assist local governments, guards against local overspending, and monitors subnational performance. Aside from these, the central government should also support information sharing among different local governments. This could be done through recognition of exemplary local governments (an example of which is the "Gawad Galing Pook" awards annually given to innovative local governments in the Philippines), the articulation of minimum service standards, and support of local-government associations, among many other alternatives. These functions are crucial in mitigating some of the problems associated with decentralization.

How central-government actors, more used to a "command and control" style of government, adjust to their new roles is thus critical to reform success and is also contingent on the political configurations and impetus behind decentralizing reforms. Central-level readiness is an important input in the design of

the reform process. At least three types of problems related to the center's role are cited in the literature. An often-cited problem in literature is the tendency for central governments to abdicate important responsibilities in the process of decentralization. Second, complicated and cumbersome monitoring system designs may be beyond a central government's capacity to implement effectively, weakening the attempt to hold local governments accountable in areas of high national priority. Third, some powerful central-level actors may oppose decentralization due to powerful interests in managing economic activities, as when ministries in Vietnam resisted decentralization of control over state-owned enterprises that had long proved a source of revenue and authority. Such resistance may even lead to currents of effective "recentralization," even when the process of ostensibly decentralizing rhetorically continues, as documented for Thailand by Mutebi.^[20]

CONCLUSION

Decentralization is likely to remain a key part of the rhetoric and practice of public sector reforms, but there is a need for the discourse to be more nuanced. It is not a one-way street and it is not irreversible, and attempts to recentralize can be expected. But whether or not recentralization occurs, the center will continue to play an important role in decentralizing systems and, thus, needs to be given further attention. It may be useful to think of a pendulum in center-local relations, wherein ultimately no set equilibrium will ever be reached.

Beyond center-local relations, there is also a need for greater care in defining and contextualizing decentralization. It is not enough to simply categorize the reforms being implemented. Critical attention should be given to factors underpinning the "demand" for (or appropriateness of) decentralization—such as the degree of socioeconomic development and size of a country—and the ability of a country to "supply" functioning decentralized institutions—such as the existing degree of responsive governance. The immediate political circumstances underlying decentralization reforms also must be characterized, in order to understand how they are changing the constellation of stakeholders in a country.

Finally, understanding the national context of decentralization is not enough. Literature should also pursue a greater understanding of the subnational context. There is little understanding of how policy processes and decision making take place at the local level. Anthropology of the local state must be developed, to help gather empirical data on the variations among local political structures and institutions.

This would bring about a more complete picture of the problems and prospects of decentralization reforms in developing countries.

REFERENCES

1. Manor, J. *The Political Economy of Democratic Decentralization*; World Bank: Washington DC, 1999.
2. Stepan, A. *Multi-nationalism, Democracy and "Asymmetrical Federalism"*; Technical Advisory Network of Burma, 2002; working paper.
3. Rondinelli, D.A. Assessing decentralization policies in developing countries: the case for cautious optimism. *Dev. Policy Rev.* **1986**, 4, 3–23.
4. Turner, M.; Hulme, D. *Governance, Administration & Development: Making the State Work*; McMillan Press Ltd: Hong Kong, 1997.
5. World Bank. *World Development Report 2004: Making Services Work for the Poor*; World Bank: Washington, DC, 2004.
6. Santos, B.D. Participatory budgeting in Porto Alegre: toward a redistributive democracy. *Polit. Soc.* **1998**, 26 (4), 461–510.
7. Bardhan, P. Decentralization of governance and development. *J. Econ. Perspect.* **2002**, 16 (4), 182–205.
8. Campos, J.; Hellman, J. Governance gone local: does decentralization improve accountability. In *East Asia Decentralizes*; World Bank: Washington, DC, 2005.
9. Hofman, B.; Kaiser, K. The making of the big bang and its aftermath. Can decentralization help rebuild Indonesia? International Studies Program, Andrew Young School of Policy Studies; Georgia State University, Atlanta, GA, 2002.
10. World Bank. *World Development Report 2004*; World Bank: Washington, DC, 2004.
11. Dillinger, W.; Perry, G. Macroeconomic management in decentralized democracies: the quest for hard budget constraints in Latin America. In *Development in Latin America and the Caribbean: Decentralization and Accountability of the Public Sector*; World Bank: Washington, DC, 1999.
12. Litvack, J.; Ahmand, J.; Bird, R. *Rethinking Decentralization in Developing Countries*; World Bank: Washington, DC, 1988.
13. Hofman, B.; Guerra, S. Fiscal disparities in East Asia: how large and do they matter. In *East Asia Decentralizes*; World Bank: Washington, DC, 1999.
14. Tendler, J. *Good Government in the Tropics*; Johns Hopkins University Press: Baltimore, 1997.
15. Fritzen, S. Decentralization and local government performance: a comparative approach with application to social policy reform in Vietnam. Ph.D. Dissertation, Princeton University, 2000.
16. Fizbein, A. The emergence of local capacity: lessons from Colombia. *World Dev.* **1997**, 25 (7), 1029–1043.
17. Korten, D.C. People-centered development: toward a framework. In *People-Centered Development: Contributions Toward Theory and Planning Frameworks*; Korten, D.C., Klauss, R., Eds.; Kumarin Press: West Hartford, CT, 1984.
18. Ostrom, E. Crossing the great divide: coproduction, synergy and development. *World Dev.* **1986**, 24 (6), 1073–1087.
19. Borges-Mendez, R.; Vergara, V. The participation-accountability nexus and decentralization in Latin America. In *Development in Latin America and the Caribbean: Decentralization and Accountability of the Public Sector*; World Bank: Washington, DC, 2000.
20. Mutebi, A.M. Recentralising while decentralising: center-local relations and “CEO” Governors in Thailand. *Asia Pacific J. Public Admin.* **2004**, 26, 33–53.

Decentralization in Southeast Asia

Ledivina V. Cariño

National College of Public Administration and Governance, University of the Philippines,
Quezon City, Philippines

INTRODUCTION

Decentralization is an important political process that holds the promise of enhancing any society's enjoyment of democracy and development. This article focuses on how its two principal types, devolution and deconcentration, have been manifested in Southeast Asia. Specifically, it discusses how colonialism and the later struggle against home-grown authoritarian regimes have put a stamp on the efforts of the countries of the region to close the gap between the rulers and the ruled. The present trend is a movement from centralization to devolution, with even those states choosing deconcentration doing so within a larger governance framework of a limited state and popular participation. The article ends with warnings about possible pitfalls of decentralization and how these may be contained.

PRINCIPAL DECENTRALIZATION CONCEPTS: DEVOLUTION AND DECONCENTRATION

"Decentralization" refers to the transfer of powers, functions, and resources away from the central government. It is supposed to bring government closer to the people. How it does that depends on the type of decentralization chosen. Greater physical access to government offices and services can be accomplished through "deconcentration," which is the creation of a system of field units staffed by civil servants who draw their salaries from, and implement programs drawn up by, the central government. Deconcentration, also called "administrative decentralization," is more strictly defined as the transfer of functions and resources alone, because power remains at the center and the transfer is effected to agents of the central government that remain responsible to, and are controlled by, it. The expected main strengths of deconcentration are greater efficiency in service provision and the equal treatment of all localities within the national territory.

Government can also be brought closer to the people by granting them power to select their own officials, and, through them or directly, make the decisions for their own localities. This type of decentralization is called "devolution," where the central government

transfers powers to elected local governments. Devolution, also called "political decentralization," is usually accompanied by local autonomy and is often regarded as the higher form of decentralization. In fact, "decentralization," unqualified, is usually used as a synonym of devolution.

Devolution shares with deconcentration the advantage of improved access and efficiency. However, the principal strengths of devolution lie in its accountability and responsiveness to the people. Although the center continues to see to it that all local governments maintain national standards and less developed ones are assisted in the performance of their functions, the accountability of local governments is primarily to its citizens. Because of this, it is expected to fashion programs more in tune with the particular demands of its local inhabitants, instead of being bound by the homogenizing ethic that deconcentration follows.

Deconcentration and devolution are two ends of a continuum rather than mutually exclusive categories. For instance, devolution in the Philippines is widely accepted as the most evolved in the region, with local governments elected for the last half-century and major functions transferred to these local units since 1991. However, education, the biggest sector (accounting for 18% of the state budget and a third of the civil service force), remains a deconcentrated function. Besides, the central government continues to perform some of the devolved functions instead of concentrating on national standard setting, monitoring, and incentives provision, its appropriate roles in a regime of devolution.

Also, federalism, being a union of autonomous sub-national units called states, is usually associated with devolution. However, despite having the only federal structure in the region, Malaysia maintains a deconcentrated system of governance. Local chief executives of the states are either hereditary or appointed by the Malaysian king, the *Yang di Pertuan Agong*. In turn, the state governments appoint the heads of local bodies. Moreover, the federal government can make and has made binding decisions on matters in the State List, with nary a complaint from the local units. Besides, close to 85% of total government revenues accrues to the federal government, giving the states very little room to fund their own programs.^[1]

The wide variety of arrangements and processes that fall under the rubric of decentralization conveys a sense of the complexity of the concept. No nation has stuck to only one option, and the decentralization regime of each country is a product of its own history and culture; the pressures of its political, economic, and social forces; and even the demands of the global environment.

IMPACT OF COLONIALISM ON SOUTHEAST ASIAN DECENTRALIZATION

Western colonialism to a large extent drew the boundaries of the nations of Southeast Asia, and its legacy continues to be manifested in the decentralization status and struggles of these nations. What is now Malaysia consists of individual kingdoms until the 18th and 19th centuries. The British consolidated them as a federation under a British resident-general, with each territory ran by their respective sultans with a British adviser. Thus the federal structure of Malaysia has its roots in colonial times. Meanwhile, French Indo-China consisted of Cochinchina, Tonkin, Annam, Laos, and Cambodia. Upon independence, Laos and Cambodia claimed the land of their precolonial kingdoms, while the first three territories formed the present Vietnam.

Perhaps the most problematic colonial bequests are the archipelagoes of Indonesia and the Philippines whose component islands were separate kingdoms until they were ruled as single colonies by the Netherlands and Spain, respectively. The idea of a single nation was born practically only in their respective struggles for independence.

The Netherlands imposed a federation of 15 republics on newly independent Indonesia in 1947. This was reversed with the creation of a unitary state in 1950. At present, most of the former Dutch East Indies have accepted the idea of national unity under one Indonesia, except for Aceh and Papua whose demand is for independence. Indonesia has responded with Law 18/2001 and Law 22/2001, granting them greater autonomy. East Timor, which was recently granted independence from Indonesia, used to be a Portuguese, and not a Dutch colony. For the rest of the country, resistance has been directed instead against the centralizing policies of Jakarta and the loss of regional identity and autonomy. Similarly, the Philippine territory is defined as it was when the United States (successor colonizers to Spain) left in 1946, but secessionist movements raged in areas not completely colonized by the West. The Philippine government countered with proposals to give autonomous status to Muslim Mindanao and the Cordilleras, upon approval by their inhabitants in a special plebiscite. The organic laws were accepted

in parts of Mindanao but not in the Cordilleras; thus the former enjoys devolved powers while the latter remains simply an administrative region. In the case of both Indonesia and the Philippines, the autonomy afforded these territories is an exception to the general level of devolution provided to the rest of the local governments in these nations.

While there is no necessary historical sequence to decentralization, most countries experience it first as deconcentration. It has been said that in Thailand, anyone wishing to ignore the government simply disappears into the hinterlands.^[2] However, a field unit makes that less possible. Thus deconcentration extends the reach of the State. Moreover, it also attempts to maintain equity and justice by treating all areas as homogeneous parts of the territory. Nevertheless, with no accountability to the people it serves while being so distant from the center to which it is supposed to be responsible, that unit may simply intone, as the Spanish colonial field officials used to say in the Philippines, “obedeusco pero no cumplo” (“I obey but I do not comply”).^[3] Thus rules and regulations that stream from the center may be followed to the letter, but the spirit of service and equity that animates them may be lost as they move down through several layers of the bureaucratic hierarchy.

GROWTH OF DEMOCRACY AND DEVOLUTION

The last decade of the 20th century has seen a clear trend in Southeast Asia toward devolution and a loosening of central control even in deconcentration. Devolution has commended itself as a result of the growth of democracy, or at least, the idea of democracy, in the region, and the rise of an awakened citizenry that demands a say in governance. These factors are associated with economic growth and an expanding middle class. They are complemented by the demonstration effect of experiences in neighboring countries, made visible and immediate by mass media, technological developments, intraregional information exchanges, and the support of international donor organizations for decentralization. The effects of these factors can be seen in the devolution experiences of the Philippines, Indonesia, Thailand, and Cambodia, all of whom emerged from authoritarian rule in the last 15 years of the last century.

The Philippines started the trend with the overthrow of Martial Law President Ferdinand E. Marcos in a peaceful 4-day “People Power Revolution” in 1986. In the redemocratization period that followed, it wrote into the Constitution of 1987 the principles of local autonomy and people’s participation in governance. Although the Philippine decentralization movement has its roots in the 1960s with the election of councils

at village level, the landmark devolution law was enacted only in 1991, with people power euphoria still in the air. The Local Government Code devolved major powers to local government units, and increased their revenues as well as their taxing powers. It also provided for citizen participation in local governance through the inclusion of people's representatives in five local special bodies.

Similarly, devolution in Indonesia is a component of *Reformasi*, the movement for reform that toppled the three-decade-long rule of President Suharto in 1996. Post-Suharto amendments to the constitution provide for wide-ranging autonomy to regional governments and democratically elected local legislatures at provincial, town, and city levels. Law 22/1999 and Law 25/1999 provide for devolution and its concomitant, fiscal decentralization, respectively. These laws are based on the principles of democracy, community participation, equity and justice, recognition of the potential and diversity within regions, and the need to strengthen local legislatures. Law 22 gives the provinces a dual status, as autonomous regions and as administrative arms of the center. Meanwhile, it gives full autonomy to districts and municipalities (rural and urban local governments, respectively), devolving powers to them except in security and defense, foreign policy, monetary and fiscal matters, and justice and religious affairs, similar to a federalistic allocation of powers.^[4-7]

A peaceful revolution of the Thai middle class overthrew the military government in 1992, and gave rise to the new democratic Constitution of 1997. That constitution was both democratically written—because even ordinary citizens could submit their suggested drafts—and democratic in substance and spirit. It sought to enhance public participation in governance and to promote new channels for democracy, including a National Decentralization Committee that includes private citizens.^[8] Nine articles on local self-rule and decentralization have served as the basis for a series of decentralization laws and policies passed since 1999. The most important of these is the National Decentralization Act, which specifies devolution in 4 years and change of the ratio of expenditure between central and local government from 91:9 in 1999 to 65:35 in 2006. Meanwhile, the Municipality Act of 2000 mandates that mayors of metropolitan municipalities and cities would be directly elected at the end of their current term.^[9]

Cambodia, meanwhile, emerged from a bloody civil war that transformed it from a socialist state into a multiparty liberal democracy in 1993. Decentralization has been at the heart of its rehabilitation, being a feature of the *Seila* (Stone Foundation) Program and other reform initiatives of the Royal Government of Cambodia. A poverty-alleviation pilot program supported by the United Nations Development Program,

Seila attacks poverty by strengthening local government structures from the commune level upward. The government has written Seila's decentralization mechanisms into the 2001 Law of Administration of Communes. In 2002, elected commune officials replaced state appointees and started to wield powers in order and security, health, economic and social development planning, cultural and environmental property, and general welfare.^[10]

For the other countries, deconcentration remains the prevailing decentralization mode. However, recognizing the problems of central control that it represents, administrative decentralization is now embodied within a larger governance framework that involves providing field units more autonomy, the participation of the citizens, and fledgling efforts toward elected local units. The size of the city-state of Singapore gives it little option except deconcentration, but it is nevertheless moving away from classic central control to allowing more autonomy to be exercised by administrative units such as schools and health centers. This has happened even in Vietnam, a socialist State that recognizes the Communist Party as the force leading State and society and practices democratic centralism. The Government is the executive body of the National Assembly and the highest administrative body of the State. It directs the work of ministries at the central level as well as People's Committees (PCOMs), the executive bodies at the local levels. Although People's Councils are elected at provincial, district, and commune levels, they are supervised and guided by the Standing Committee of the National Assembly and the Government and are thus not autonomous local bodies. Nevertheless, as the program of *doi moi* (renovation) is moving the country from a centrally planned to a more market-oriented economy, Vietnam's Public Administration Reform (PAR) program embodies a complex decentralization agenda. Launched in 1995, PAR specifies the centralization of all regulation-making tasks and the decentralization of economic and social decision making. Thus the Grassroots Democracy Decree (No. 29, passed in 1998) gave elected commune-level administrations the task of ensuring that the citizens exercise their rights, and that government be accountable to households for information about local activities and finances.

CONCLUSIONS: PROBLEMS AND CHALLENGES FOR DECENTRALIZATION

The popularity of decentralization in all the countries of the region does not imply that it is a panacea for all the ills of governance and development. There is a need for complementary policies and measures to be

able to achieve poverty reduction, people's participation, and economic development.

Moreover, there are problems associated with decentralization itself. If power is transferred to officials who care more for their self-interest than the general welfare, decentralization can conceivably nurture fiefdoms of corruption and abuse. This suggests the need for decentralization to be accompanied by strong accountability mechanisms, such as local media and civil society, and impartial central oversight agencies. Many analysts also point to the issue of fiscal decentralization, which do not necessarily move concomitantly with power decentralization. Some countries decentralize power but keep the control of funds at the center, making the local units incapable of adequately serving their constituents. On the other hand, a generous center may keep local units dependent on its largesse and unable or unwilling to generate local revenues that could make them unpopular with the people. Other problems may arise because of the weakness of the new decentralized structures, the lack of competent officials and staff in many localities or their greed and power-hunger, and the absence of active citizens and institutions that can help both in capability building and in exacting accountability from the field units and local governments. Decentralization is a means of moving government closer to the people. Yet as has been seen, in many Southeast Asian countries, this is not a one-way street as the people have gotten new decentralization policies through their own demands and struggles. A similar militancy is necessary so that its gains are constantly monitored and defended and its ills minimized.

REFERENCES

1. Celestino, A.B. Malaysia: does it really need decentralization? In *Decentralization and Power Shift: An Imperative for Good Governance (A Sourcebook on Decentralization Experiences in Asia)*; Brillantes, A.B., Jr., Cuachon, N.G., Eds.; Asian Resource Center for Decentralization: Diliman, Quezon City, 2002; Vol. 1, 137–150.
2. Siffin, W.B. *The Thai Bureaucracy: Institutional Change and Development*; East-West Center Press: Honolulu, 1966.
3. Endriga, J.N. Historical notes on graft and corruption in the Philippines. *Philipp. J. Public Adm.* **July–October 1979**, 23 (3–4), 241–254.
4. Asanuma, S.; Brodjonegoro, B. *Indonesia's Decentralization Policy: Origins, Issues and Policy Directions*; 2003. http://www.gtzsfdm.or.id/documents/dec_ind/o_pa_doc/Shinji_Brodjonegoro_DecentralizationPolicy_Sept2003.pdf.

5. Esden, B. Indonesia: rising above challenges. In *Decentralization and Power Shift: An Imperative for Good Governance (A Sourcebook on Decentralization Experiences in Asia)*; Brillantes, A.B., Jr., Cuachon, N.G., Eds.; Asian Resource Center for Decentralization: Diliman, Quezon City, 2002; Vol. 1, 115–126.
6. Mokhsen, N. *Decentralization in the Post New Order Era of Indonesia*; 2003. <http://www.commonwealthseminar.org.nz/Papers2003/Nuraida%20Mohksen%20Indonesia%20Decentralisation%20Case%20Study.pdf>.
7. Rocamora, J. *Legal and Policy Frameworks for 'Participation' in Southeast Asia*; 2003. LogoLink Research. http://www.ids.ac.uk/logolink/resources/downloads/regionalreports/RegionalReportSoutheastAsia_20fmal.pdf.
8. Mutebi, A.M. *Thailand's Decentralization Experiment: Evolution, Dimensions, and Challenges*; 2003. <http://www.fas.nus.edu.sg/ppp/wp/wp46.pdf>.
9. Cuachon, N. Thailand: the continuing quest for local autonomy. In *Decentralization and Power Shift: An Imperative for Good Governance (A Sourcebook on Decentralization Experiences in Asia)*; Brillantes, A.B., Jr., Cuachon, N.G., Eds.; Asian Resource Center for Decentralization: Diliman, Quezon City, 2002; Vol. 1, 165–172.
10. Tumanut, M. Cambodia: hitting the road. In *Sourcebook on Decentralization in Asia*; 2002; 94–101. <http://www.decentralization.ws/srcbook/cambodia.pdf>.

BIBLIOGRAPHY

1. International Development Research Center. *Some Socio-Economic Features Related to Decentralization*; 1997. <http://www.idrc.ca/socdev/pub/vietnam/ch1.html>.
2. Mok, K.-h.; Chan, D. *Decentralization and Marketization of Education in Singapore: A Case Study of School Excellence Model*; 2002. http://personal.cityu.edu.hk/~samiclee/middle_activities_501.html.
3. Quan, N.M. *Decentralization in Vietnam: Key Issues and Possible Solutions*; 2002. http://www.decentralization.ws/icd2.decent_vietnam.htm.
4. Rasyid, E.; Sugiyanto, C.; Ozeki, Y. *Local Government Capacity Constraints and Decentralization in Indonesia*; 2002. http://www.gtzsfdm.or.id/documents/dec_ind/o_pa_doc/LocGovtCapacity_Rasyid%20et.al.%20Sept2003.pdf.
5. Sutherland, H.; Raben, R.; Locher-Scholten, E. *Rethinking Regionalism: Changing Horizons in Indonesia 1950s–2000s*; 2000. <http://www.knaw.nl/indonesia/transition/workshop/chapter3sutherlandrabenlocher-s.pdf>.
6. Suwondo, K. *Decentralization in Indonesia*; 2002. <http://www.infid.be/INFID%20Background%202002%20Decentralisation.pdf>.
7. The Commune Council Support Project. *NGO Statement to the 2001 Consultative Group Meeting on Cambodia*; 2001. http://www.bigpond.com.kh/users/ngoforum/cg2001/commune_admin.htm.

Decision Making in Administrative Organizations: Theories

James S. Guseh

*Department of Public Administration, North Carolina Central University,
Durham, North Carolina, U.S.A.*

INTRODUCTION

Decision making is the process of choosing from among alternative courses of action. It is both an objective and a subjective process. The objective process involves setting goals and formulating an efficient process for attaining the goals. The subjective process may involve the values of decision makers, which may lead to the acceptance of a satisfactory rather than an optimum solution. Decision making is central in public administration because it is one of the core managerial functions in organizations. How decisions are made, what criteria are used, what costs are involved, and who benefits are some of the important questions that characterize the decision-making process. Thus, how managers are judged depends on the decisions they make.^[1]

There are various theories or models of decision making. This article focuses on the five dominant models of decision making: the classical, bounded rationality, incremental, mixed scanning, and garbage can. An emerging model of decision making, known as win-win analysis, which seems to address some of the limitations of the other models, is also discussed. The analysis includes the approaches utilized by each model, the criticisms, and other dimensions of the process.

and analysis of which means of attaining them are most reasonable, i.e., require the least input of scarce resources. ... Thus, whenever economists refer to a "rational man" they are not designating a man whose thought processes consist exclusively of logical propositions, or a man without prejudices, or a man whose emotions are inoperative. In normal usage all of these could be considered rational men. But the economic definition refers solely to a man who moves toward his goals in a way which, to the best of his knowledge, uses the least possible input of scarce resources per unit of valued output.^[2]

Although the steps involved have been outlined in various ways by different authors, the essential elements involved the following steps:

1. Identify the problem to be solved.
2. Establish the evaluation criteria for problem resolution.
3. Propose the various alternatives for solving the problem.
4. Analyze the likely consequences of each.
5. Compare and evaluate the results of each alternative.
6. Choose the best alternative.
7. Implement the chosen alternatives.
8. Evaluate the results.

THE RATIONAL MODEL

The rational model of decision making, also known as the classical or economic model, has held a prominent place in the literature in public administration. The model is derived from economic theory that people seek to maximize utility and pursue their self-interests. The model is based on the assumption that people are economically rational and attempt to maximize outputs for a given input or minimize input for a given output in an orderly and sequential process. Political economist Anthony Downs presents the economic analysis of the rational model as follows:

Economic analysis thus consists of two major steps: discovery of the ends a decision maker is pursuing

Details of each step can be addressed as separate topics but are not discussed in this article.

Although, in actuality, the decision-making process is rarely this orderly, the rational model has had a significant influence on decision making in bureaucracy. For example, it is reported to have been applied in some form by President John F. Kennedy's administration to address the Cuban Missile Crisis in 1962.^[3] It has also been employed in addressing many urban problems in American cities such as Cleveland, Dallas, Dayton, Fort Worth, Long Beach, Philadelphia, Phoenix, Pittsburgh, San Antonio, San Francisco, Virginia Beach, Roanoke, and Greensboro, to name but a few.^[1]

Despite its widespread application in organizations, the rational model has been challenged since the late 1950s. One of the leading critics of this model is Nobel

laureate Herbert Simon. While the rational decision of the classical model assumes that the decision maker is completely informed about both the organizational goals and all the possible alternatives and acts to maximize utility, the administrative decision maker can rarely approximate the kind of rationality assumed by the classical economic model. Rarely can an administrator count on full knowledge of the situation, including the consequences of the various alternatives. According to Simon:

It is impossible for the behavior of a single, isolated individual to reach any high degree of rationality. The number of alternatives he must explore is too great, the information he would need to evaluate them so vast that even an approximation to objective rationality is hard to conceive. Individual choice takes place in an environment of “givens”—premises that are accepted by the subject as bases for his choice; and behavior is adaptive only within the limits set by these “givens.”^[4]

Thus, Simon put forth a more realistic model of decision making known as the bounded-rationality model.

THE BOUNDED-RATIONALITY DECISION-MAKING MODEL

This model represents the decision maker as an “administrative man” with limited information-processing ability. According to Simon, the decision maker’s ability to analyze the alternatives and objectives is limited or bounded to varying degrees. He explains the concept as follows:

When the limits to rationality are viewed from the individual’s standpoint, they fall into three categories: he is limited by his unconscious skills, habits and reflexes; he is limited by his values and conceptions of purpose, which may diverge from the organizational goals; he is limited by the extent of his knowledge and information. The individual can be rational in terms of the organization’s goals only to the extent that he is *able* to pursue a particular course of action, he has a correct conception of the *goal* of action, and he is correctly informed about the conditions surrounding his action. Within the boundaries laid down by these factors, his choices are rational—goal-oriented.^[5]

Unlike the “economic man” of the classical economic model who seeks to maximize, the “administrative man” seeks to “satisfies,” i.e., find satisfactory solutions rather than maximize. While the economic man seeks rationality, the administrative man recognizes the limits of his capability for rational behavior. The administrative man is content with a simplified

and incomplete view of the world that can never, because of human limitations, approximate the complexity of the real world. Simon highlights the characteristics of the administrative man as follows:

First, because he satisfies rather than maximizes, administrative man can make his choices without first examining all possible behavior alternatives and without ascertaining that these are in fact all the alternatives. Second, because he treats the world as rather “empty,” and ignores the “interrelatedness of all things (so stupefying to thoughts and action), administrative man is able to make his decisions with relatively simple rules of thumb that do not make impossible demands upon his capacity for thought.^[6]

Thus, while the administrative man cannot achieve the ideal behavior of the classical economic man, he does the best with what he has. He is willing to settle for an adequate solution in contrast to an optimal one.^[7]

The steps in the bounded-rationality model are as follows:^[8]

1. Identify the problem to be solved or goal to be defined.
2. Determine the minimum level or standard that all acceptable alternatives will have to meet.
3. Choose one feasible alternative that will resolve the issue.
4. Appraise the acceptability that will resolve the issue.
5. Determine if it meets the minimum levels that have been established.
6. If the alternative is not acceptable, identify another and put it through the evaluation process.
7. If the alternative is acceptable, implement it.
8. After implementation, determine how easy (or difficult) it was to discover feasible alternatives and use this information to raise or lower the minimum level of acceptability on future problems of a similar nature.

How accurate is the bounded-rationality model? Empirical studies have shown that the model is very accurate.^[9,10] A comparison of the classical economic model and the bounded-rationality model shows that the latter presents a more realistic view of decision making.

INCREMENTAL MODEL

Political scientist Charles Lindblom has also criticized the rational model. He argues that the rational model suffers from a number of limitations, such as:^[11]

1. Human intelligence or problem-solving capacity is too limited to encompass all of the options

and potential outcomes of the alternatives generated by the synoptic approach.

2. Information is not adequate to assess all options accurately.
3. Comprehensive analysis requires too much time.
4. Comprehensive analysis is too expensive.
5. Facts and values cannot be neatly separated as required by the comprehensive approach.

Lindblom offers another theory of decision making, variously called incrementalism, muddling through, and successive limited comparisons. Incrementalism, in contrast to the rational model, stresses decision making through a series of limited, successive comparisons, with a relatively narrow range of alternatives rather than a comprehensive range; it uses the status quo, not abstract goals, as the key point of reference for decisions. Incrementalism focuses primarily on short-term rather than long-term effects, on the most crucial consequences of an action rather than on all conceivable results, and on less formalized methods of measuring costs and benefits.^[12]

The essential elements of this model include the following:^[11,13]

1. Goals are not isolated and determined before analysis begins. Goal determination and analysis are closely intertwined, even simultaneous. The means often affect the ends, and vice versa.
2. Decision makers usually consider only a limited number of alternatives—ordinarily only those that differ marginally from existing policy.
3. All consequences, even of the more restricted options, are not evaluated. All consequences cannot be known, and the time and effort required for comprehensive assessment normally are unavailable.
4. Since means and ends are inseparable, problem redefinition is continuous, analysis is never ending, and policy is never made once and for all but remade endlessly.

Lindblom has presented several arguments to support this model. Errors resulting from employing incremental changes from the status quo may not have serious implications and may be easier to correct. Politically, incremental solutions are easier to agree upon, since this change may fall somewhere between extreme positions, thereby getting plurality support. Lindblom further maintains that making continual incremental adjustments in both the problem definition and proposed solutions is a reasonable and effective method of solving problems and making decisions.

Like the rational model, the incremental model has also come under attack. Two of the leading critics,

Yehezkel Dror and Amitai Etzioni, have identified serious shortcomings of the model, and the latter has proposed an alternative model of decision making. According to Dror, the incremental changes may not address real-world problems, such as a declaration of war that may require innovative solutions that are bolder than those of the incremental approach. Thus, the incremental approach does not facilitate or provide a basis for organizational motivation or change but, instead, facilitates the maintenance of the status quo. According to Dror, incrementalism works best when:^[14]

1. The results of present policies are satisfactory.
2. There is a high degree of continuity in the nature of the problems.
3. There is a high degree of continuity in available means for dealing with the problem.

Expanding on Dror's criticisms, Etzioni argues that the incremental approach fails to distinguish between fundamental and nonfundamental decisions. He suggests that the incremental approach is suitable for nonfundamental matters but not for fundamental ones, where a wider perceptual horizon is needed.^[15] Etzioni has proposed an alternative approach to decision making, known as mixed scanning.

MIXED SCANNING

This approach combines the best features of both the rational and incremental approaches to decision making. As Etzioni describes decision making:

Fundamental decisions are made by exploring the main alternatives the actor sees in view of his conception of his goals, but—unlike what rationalism would indicate—details and specifications are omitted so that an overview is feasible. Incremental decisions are made but within the context set by fundamental decisions (and fundamental reviews). Thus, each of the two elements in mixed-scanning helps to reduce the effects of the particular shortcomings of the other; incrementalism reduces the unrealistic aspects of rationalism by limiting the details required in fundamental decisions, and ... rationalism helps to overcome the conservatism slant of incrementalism by exploring longer-run alternatives.^[15]

According to Etzioni, one of the instances in which this model applies is in the court system. With many cases to decide, lower courts tend to apply the incremental approach by maintaining or slightly deviating from the status quo. When cases are appealed, higher courts tend to analyze issues in more

detail, using the rational approach, in order to make fundamental (and sometimes bold) decisions.

Despite this example of when this model is applied, the major criticism is that it provides few guidelines for how and when it should be employed.

THE GARBAGE-CAN MODEL

While the approaches to decision making discussed earlier may suggest that organizations make decisions under conditions of much certainty and knowledge, Michael Cohen, James March, and Johan Olson have examined decision making under conditions of unclear organizational setting. That is, there are circumstances in which organizations have “goals that are unclear, technologies that are imperfectly understood, histories that are difficult to interpret, and participants who wander in and out.”^[12] Cohen et al. have termed these conditions as “organized anarchies.” Such organizations operate under conditions of pervasive ambiguity, with so much uncertainty in the decision-making process that traditional theories of coping with uncertainty do not apply.^[12] Thus, they consider an organization as “a collection of choices looking for problems, issues and feelings looking for decision situations in which they might be aired, solutions looking for issues to which they might be the answer, and decision makers looking for work.”^[16] According to March et al., decision making in these circumstances is not regarded as a crisis or transitory. Instead, it is regarded as “characteristic of any organization in part,” and is “particularly conspicuous in public, educational, and illegitimate organizations.”^[7] Decision making under these conditions is referred to as the “garbage-can theory of organizational choice.”

The garbage-can model views:

a choice opportunity as a garbage can into which various problems are dumped by participants. The mix of garbage in a single can depends partly on the labels attached to the alternative cans; but it also depends on what garbage is being produced at the moment, on the mix of cans available, and on the speed that garbage is collected and removed from the scene.^[17]

Which problems get attached to which solutions is a major issue. “This is largely determined by chance—by what participants with what goals happened to be on the scene, by when the solutions or the problems entered, and so on.”^[7] Although the model may seem bizarre or pathological when compared to the traditional rational decision-making model, “it does produce decisions under conditions of high uncertainty: that is some solutions do get attached by some participants to some problems.”^[7]

According to Cohen and March, decision making in this model is made in three ways:^[18]

1. By oversight, without any attention to existing problems and with a minimum of time and energy.
2. By flight (or avoidance), postponement, or buck-passing.
3. By resolution, which occurs generally (on) when flight is severely restricted or when problems are relatively minor or uncomplicated.

In terms of application, this model seems to describe the decision-making processes by college and university presidents as follows:

Opportunities for choice in higher education can easily become complex “garbage cans” into which a striking variety of problems, solutions, and participants may be dumped. Debate over the hiring of a football coach can become connected to concerns about the essence of a liberal education, the relations of the school to ethnic minorities, or the philosophy of talent.^[7]

Despite the unstable environment or unusual nature of organized anarchies, leaders in these organizations can still improve their performance in decision making. Scott suggests that:

By carefully timing issue creation, by being sensitive to shifting interests and involvement of participants, by recognizing the status and power implications of choice situations, by abandoning initiatives that have become hopelessly entangled with other, originally unrelated problems, by realizing that the planning function is largely symbolic and chiefly provides excuses for interaction, leaders in organized anarchies can maintain their sanity and, sometimes, make a difference in the decision made.^[7]

In light of recent developments in many public organizations, this model of decision making may provide the most appropriate perspective from which to understand the complexity of government decision making in the 21st century.^[12]

WIN-WIN POLICY ANALYSIS

In each of the models of decision making discussed, there seems to be some trade-off among the alternative choices. For example, in the rational model, all decision makers involved may not agree on the evaluation criteria. The policy alternative selected may not be the choice of some decision makers. Thus, the evaluation criteria selected can determine the policy alternative that is chosen. In the case of the incremental

model, the incremental choice tends to be a compromise, whereby all major sides have to give up something to gain something.

An emerging decision-making model that seems to address the win-lose phenomenon in these models of decision making by presenting a win-win policy is referred to as win-win policy analysis or superoptimizing solutions. Although the model received initial treatment in the 1980s by various researchers, Stuart Nagel has expanded upon the model since the 1990s through various worldwide applications.^[19,20]

Win-win analysis is defined as formulating policies that exceed the best initial goals of expectations of all major groups involved in the policy dispute.^[20] The groups could be conservatives, liberals, and political parties. A win-win or superoptimum solution is different from a compromise. While in a compromise situation all sides lose something and gain something in order to arrive at an agreement, a win-win analysis is “capable of achieving conservative goals more than the conservative policies” and simultaneously capable of achieving the liberal goals more than the liberal policies.^[19,20] Win-win analysis has applications in any field of public policy, such as economic, social, political, legal, or environmental.

Nagel provides the basic steps in the win-win policy analysis and various ways to facilitate the development of win-win solutions. There are five basic steps to win-win policy analysis:

1. What are the major goals of conservatives, liberals, or other major groups who are disputing what policy should be adopted for a given policy problem?
2. What are the major alternatives of those groups for dealing with the policy problem?
3. What are the relations between each major alternative and each major goal? In their simplest form, these relations can be expressed in terms of a minus sign (relatively adverse relation), plus sign (relative conducive relation), and a zero (neither adverse nor conducive relation).
4. What new alternative there is might be capable of
 - a. achieving the conservative goals even better than the conservative alternative and
 - b. simultaneously capable of achieving the liberal goals even more than the liberal alternative? Whatever new alternative meets these two criteria is a win-win alternative or a superoptimum solution.
5. Is the proposed win-win alternative capable of getting over various hurdles that frequently exist? These hurdles may be, in random order, political, administrative, technological, legal, psychological, and economic.

There are various ways of facilitating ideas for win-win solutions, such as:

1. Expand the resources available so that both liberal and conservative goals can be achieved.
2. Modify the liberal alternatives so that they also achieve the conservative goals and/or modify the conservative alternatives so that they also achieve the liberal goals.
3. Redefine the problem to emphasize goals, rather than alternatives.
4. Fully combine alternatives that are not mutually exclusive.
5. Make use of a mind-stimulating matrix that shows goals on the column, alternatives on the rows, and scores or directions of relations in the cells. Have an empty row at the bottom of the matrix for inserting words and data on possible win-win solutions.

CONCLUSIONS

Decision making is important in administrative organizations. Although the various models of decision making discussed have their strengths and limitations, they provide insights into organizational decision making. Moreover, since the choices made can have positive or negative implications for the organizations, the type or mode of decision making employed is important. Given the importance of decision making in organizations, more research is needed to improve organizational decision making.

REFERENCES

1. Morgan, D.R.; England, R.E. *Managing Urban America*; Chatham House Publishers: Chatham, 1986; 116, 121.
2. Downs, A. *An Economic Theory of Democracy*; Harper & Row: New York, 1957; 4 pp.
3. Allison, G. *Essence of Decision: Explaining the Cuban Missile Crisis*; Little, Brown: Boston, 1971.
4. Simon, H. *Administrative Behavior*; Macmillan: New York, 1976; 79 pp.
5. Simon, H.A. *Administrative Behavior*; The Free Press: New York, 1976; 241 pp.
6. Simon, H.A. *Administrative Behavior: A Study of Decision-Making Process in Administrative Organization*; The Free Press: New York, 1957; xxvi pp.
7. Scott, W.R. *Organizations: Rational, Natural, and Open Systems*; Prentice-Hall: Englewood Cliffs, NJ, 1992; 45, 297–299; Cohen, M.D.; March, J.G.; Olsen, J.P. People, problems, solutions and the ambiguity of relevance. In *Ambiguity and Choice in Organizations*; March, J.G., Olsen, J.P., Eds.; Universitetsforlaget: Bergen, Norway, 1976; 24–37.

8. Hodgetts, R.M. *Organizational Behavior: Theory and Practice*; Macmillan: New York, 1991; 374 pp.
9. Cyert, R.M.; March, J.G.; Eds. *A Behavior Theory of the Firm*; Prentice-Hall: Englewood Cliffs, NJ, 1963.
10. Clarkson, G.P.E.; Cyert, R.M.; March, J.G.; Eds. A model of trust investment behavior. In *A Behavior Theory of the Firm*; Prentice-Hall: Englewood Cliffs, NJ, 1963; 253–267.
11. Lindblom, C.E. *The Intelligence of Democracy*; Free Press: New York, 1965; 138–143.
12. Milakovich, M.E.; Gordon, G.J. *Public Administration in America*; Bedford/St. Martin's: Boston, 2001; 202, 227, 228, note 51.
13. Lindblom, C.E. The science of muddling. *Public Admin. Rev.* **1959**, *19*, 79–88.
14. Dror, Y. Muddling through—‘science’ or inertia? *Public Admin. Rev.* **1964**, *24*, 154.
15. Etzioni, A. Mixed scanning: a ‘third approach to decision making.’ *Public Admin. Rev.* **1967**, *27*, 385–392.
16. Cohen, M.D.; March, J.G.; Olsen, J.P. A garbage can model of organizational choice. *Adm. Sci. Q.* **1972**, *17*, 2.
17. Cohen, M.; March, J.; Olsen, J. *Organization Theory*; Brooks/Cole: Pacific Grove, CA, 1989; 263 pp. (as quoted in Harold Gortner, Julianne Mahler, and Jeanne Nicholson).
18. Cohen, M.; March, J. *Leadership and Ambiguity*; McGraw-Hill: New York, 1974; 83–84.
19. Nagel, S. *Super-Optimum Solution and Win–Win Policy: Basic Concepts and Principles*; Greenwood Publishers/Quorum Books: Westport, CT, 1997.
20. Nagel, S. *Policy-Analysis Methods and Super-Optimum Solutions*; Nova Science: New York, 1994.

Decision Making, Incrementalism and Transformational Change

Keith Carrington

*Criminal Justice and Legal Studies Division, St. John's University,
Jamaica, New York, U.S.A.*

INTRODUCTION

Before the publication of Herbert Simon's text, *Administrative Behavior*, in 1957,^[1] most scholars showed little interest in decision making at the low levels in organizations. This surprisingly was due to the influence of the politics-administration dichotomy, which caused the focus to be on decision making at the highest level in organizations where policy was formulated. Simon's work, however, adjusted the focus when he asserted that decision making constituted the core of administration, and that the process was the same throughout the organization.^[2]

Decision making is one of most frequently discussed topics in public administration. The quest to understand the decision-making process is not new, and the process of "deciding" has frustrated many decision makers. Public sector managers are inundated with decision-making models. Despite these models and a proliferation of literature, no one model has been proffered as the "one best" method. Notwithstanding, the models, each with its strengths and weaknesses, present an array of choices for public managers.

DECISION MAKING: A DEFINITION

Starling^[3] defined decision making as the process of "selecting from various alternatives one course of action." Starling further postulated that the decision process involves several distinct steps: 1) identifying the problem; 2) gathering facts; 3) making the decision; and 4) implementing and evaluating the decision.^[3] An essential component in the definition is the final step, which emphasizes that the decision-making process does not conclude with the selection of a particular course of action. Simon^[4] reinforced this point when he stated, "the process of decision does not come to an end with the general purpose of an organization has been determined. The task of 'deciding' pervades the entire administrative organization quite as much as does the task of 'doing'."

The development of an ideal decision-making model with universal application seems impractical. Hence,

public sector managers utilize various models to select a particular course of action. This paper will address a few decision-making models.

DECISION-MAKING MODELS

There are several decision-making models, each based on a variety of premises and assumptions. The *Essence of Decision*^[5,6] provides frameworks for explaining decision making. Using three models, Rational Actor (Model I), Organizational Process (Model II), and Government Process (Model III), the authors outlined approaches to decision making by using the Cuban Missile Crisis to illustrate their application.

MODEL I: THE RATIONAL ACTOR MODEL

Stevenson et al.^[7] described the Rational Actor Model as a prescription that tells decision makers how to "decide." This model is based on two assumptions: all information is available to the decision maker, and the decision maker would maximize the decision to benefit the organization.^[8] The central concepts of the Classical Model are:

1. *Goals and objectives*: The rational decision maker ranks all possible sets of consequences in terms of his or her values and objectives.
2. *Alternatives*: The decision maker must choose among a set of available alternatives.
3. *Consequences*: To each alternative, there is a set of consequences or outcomes to be considered.
4. *Choice*: The decision maker selects the alternative with the highest payoff function.^[9]

The behavior of the actor—an individual, agency, or group—is intentional, goal seeking, and value maximizing. The decision maker employs rigor, purposiveness, and consistency—among goals and objectives.^[9] The comprehensive rationality component of this model infers that the decision maker has the ability to consistently rank and review all alternatives and

their consequences and then choose the alternative that achieves the highest utility.

MODEL II: THE ORGANIZATIONAL PROCESS MODEL

The Organizational Process Model's premise is that government decisions stem from the output of several agencies that provide information and advice. This model proffers decision making as less deliberate and the function of government's "vast conglomerate of semi feudal, loosely allied organizations each with a substantial life of its own," functioning according to standard patterns of behavior.^[9] Each organization is given individual aspects of a problem to attend to.

This model emphasizes that the organization's major concern is to operate within the framework with a set of constraints negotiated through various components of the organization. This helps them reduce uncertainty in the environment and set out those alternatives that are immediate, available, and related to the problem at hand. According to Allison and Zelikow,^[9] the decision-making process then can be understood with knowledge of the organizations' operations.

MODEL III: GOVERNMENT PROCESS MODEL

Allison's Governmental Politics Model stipulates that decisions are made through bargaining among individuals and groups with diverse interests and varying degrees of power to support those interests. Decisions reached are sometimes a result of compromise and the power and skill of proponents and opponents. Politics plays a major role in decision making. Decisions are made through the involvement of many actors who focus on many problems, numerous goals, some of which are personal, and by "pulling and hauling that is politics."^[9]

CRITICISMS OF ALLISON'S MODELS

The Classical Model appears to be oversimplified. It ostensibly makes unrealistic demands on the decision maker, who, according to Simon, must have "powers of prescience and capacities for computation resembling those usually attributed to God."^[5] In fact, decision makers experience considerable stress in attempting to deal with their limited cognitive abilities. President Warren G. Harding articulated the cognitive stress in the following statement:

John, I can't make a damn thing out of this tax problem. I listen to one side and they seem right, and then God! I talk to the other and they seem just as right, and there I am where I started. I know somewhere there is a book that would give me the truth, but hell, I couldn't read the book. I know somewhere there is an economist that knows the truth, but I don't know where to find him, and haven't the sense to know him and trust him when I did find him God what a job.^[10]

The cookie-cutter approach of the prescriptive classical model epitomizes logic, focus, and rationality, which, misleadingly, may convey the idea that the approach guarantees successful decision making. However, the model does not enjoy widespread use. According to the article, *The Wisdom of Solomon*,^[11] rational decision-making techniques are used "less than 20% of the time."

Models II and III have several weakness. According to Bendor and Hammond,^[12] the two models are "sufficiently ambiguous that it is difficult to discern the models' defining properties." Model III depicts decision making as a process involving many actors engrossed in "pulling and hauling." At the end of this process, one group triumphs over the others. Przeworski^[13] reminds us that during these political discussions, "reasons and facts are exhausted ... issues are decided by voting, which is an imposition of one will upon a resisting will ... Politics is an antagonistic relationship between 'us' and 'them' ...". This approach to decision making creates winners and losers. Although the process produces a decision that may be based on a mixture of alternatives, it might be a patchwork decision that can have disastrous consequences.

Bendor and Hammond^[12] articulated that the three models unquestionably have several important methodological points. However, the lessons learned from the models may lack foundation and, in some cases may be incorrect. Finally, it seems that the models do not provide good frameworks for low-level decision makers.

THE ADMINISTRATION MODEL

The Administration Model was proposed by March and Simon^[14] as a descriptive model that explains how decision makers actually make decisions. A decision maker using this model defines the problem in terms of his or her understanding of the situation, and does not consider all information, alternatives, and their consequences. This model may require decision makers to follow steps similar to those of the classical model, such as generating alternatives

and reflecting on their consequences. However, there is no impractical quest to glean all information. Instead, the decision maker collects relevant information based on a definition of the situation.^[15] Simon^[16] stated that there are limits to a decision maker's cognitive ability, and it is this bounded rationality that forces them to satisfice rather than optimize.^[14]

THE INCREMENTAL MODEL

Lindblom's^[17] Incremental Model offers another approach to decision making. Lindblom suggests that incrementalism is more likely used in practice. This method of decision making involves making sequential (incremental) decisions. Successive limited comparisons may be the most appropriate approach when problems are complex. Using the incremental model, the decision maker goes through the following steps:

1. An objective is established.
2. Few alternatives immediately relevant are outlined and then compared.
3. Select an alternative. This would be a combination of the choice among values and the choice among the instruments for achieving the objective.

With this approach, the decision maker focuses on incremental values and analyzes only those demonstrating the differences between the alternatives. According to Lindblom,^[17] the decision maker's "need for information on values or objectives is drastically reduced ... and his capacity for grasping, comprehending, and relating values to one another is not strained beyond the breaking point." The decision maker expects to have partial success with this decision, so he or she would expect to repeat the process endlessly, as they will be building on what has already been done. A major advantage of incrementalism is that it safeguards against errors, as the previous steps would have informed the decision maker of the possible consequences of similar steps.

Tarter and Hoy^[18] state that the incremental model has the following five distinctive features:

1. Means-ends analysis is inappropriate, because objectives and alternatives emerge simultaneously.
2. Good solutions are those upon which decision makers agree regardless of objectives.
3. Options and outcomes are dramatically reduced by considering only alternatives similar to the current state of affairs.
4. Analysis is limited to differences between the existing situation and proposed alternatives.

5. The incremental method eschews theory in favor of successive comparisons of concrete, practical alternatives.

The authors believe that incrementalism simplifies decision making, which does not make it whimsical.

Incrementalism is not without its critics. True^[19] articulates that despite incrementalism's widespread use, it has several flaws. They include the model's failure to provide comprehensive explanation of the expectations from policy and budget decisions; it says little about competition; it lacks a method to differentiate between incremental change and non-incremental change; it does not offer an understanding of value conflicts in the three levels of government when there is disagreement about who gets what; and it only gives a partial description of what really happens.

Lindblom^[17] stated that the model had several imperfections; for example, the model did not have built-in safeguards for the relevant values; and that several important policies might be overlooked, because steps taken earlier may not point in their direction. Despite the shortcomings, incrementalism has remained the dominant model used to understanding government decision making since its introduction in the mid-1960s.^[20]

TRANSFORMATIONAL CHANGE

Why would an organization choose to change its operations, and what factors would determine the type of change an organization should undertake? This question can be answered in a variety of ways. Most organizations do not voluntarily change their operations; and change does not occur by chance. Instead, organizational change occurs in response to some phenomenon from within or outside the organization or both. Organizations may undergo change to deal with inertia, chart a new course for the organization to follow, introduce new technology or handle some unexpected outcome. In addition, shrinking budgets, increased demands for improved services, and pressure from clientele may force management to do things differently to deliver better quality services.

Organizations are susceptible to many factors, as they do not operate in a stable, protected environment shielded from internal and external pressures. Notwithstanding, many organizations resist change because management may see the system as being stable. According to Kiel,^[21] "When change is desired, demanded, or inevitable, changing such a stability creates wrenching situations for employees and management."

Management scholars no longer associate order, stability, and consistency with the operations of

organizations. In fact, according to Kiel,^[21] public organizations operate in a non-linear and dynamic system full of surprises. Consequently, in response to turbulence in the organization's environment, an organization may be forced to change.

Jurow^[22] defines transformational change as "doing something different by creating new structures and new processes to fit a new objective." It necessitates a change in the status quo and the creation of new frameworks unmistakably different from what existed previously. The typical view of organization transformation theorists is that the process is a "holistic, ecological, humanistic approach to radical, revolutionary change in the entire context of an organizational system."^[23] Some theorists believe that organizations do not deliberately implement transformational change and that the change may be crisis-driven and necessary for the organization's survival.^[24] However, some organizations may change, because it is rational to do so. Dyck^[25] found that transformational changes based on value-rationality and not the status quo were the most likely to be implemented.

The literature on organizations provides different theories on how organizational change takes place. However, Sastry^[26] postulates, "Despite the important theoretical and practical implications of understanding organizational change, the organizational processes involved in transformational change have not been fully explored." Porras and Robertson^[27] concur and articulate that the theories' failure to fully explain the dynamics of the planned change is a major weakness of the field.

It is difficult to disagree that theories have not fully explained the implications of organizational change. However, Kiel^[21] articulates several principles drawn from non-linear dynamics that can provide public managers with insights into the dynamics of transformational change. There are three essential processes in transformational change. First, there will be fluctuations within the organization, and they should not be viewed as negatives, because these fluctuations can drive and energize the change process in the organization. The unusual behaviors that occur within the organization encourage the search for new problem-solving methods and service provision.

Second, managers should be aware that there would be instability within the organization. This, however, should serve as a prerequisite to change, thereby allowing the appropriate responses to environmental demands. Managers must exercise caution when dealing with instability, as only a certain level is required to realize the potential for change.

The third essential process in transformational change is disorder. This element is critical, as it facilitates the exploration of several possibilities that may lead to new operating procedures or new structures or both.

Like any other process, transformation change may fail to have the intended impact on the organization. Backer^[28] identifies management's failure to address the human side of the organization as a major factor. Managers shepherding transformation change in organizations must involve all employees in the process, as "Personal change is the prerequisite to organizational change... organizational transformation occurs one individual at a time."^[28]

CONCLUSION

The presence of fluctuations, instability, and disorder signals the formidable challenge that confronts public sector managers during the change process. These processes are neither surprising nor unexpected, because public agencies function in a non-linear environment fraught with uncertainty, chaos, and unintended consequences.

The process of change may present agencies with a seemingly insurmountable challenge. Notwithstanding, both management and other employees must have the discipline to facilitate emerging procedures and structures. A successful transformed public agency would benefit employees and clientele in several ways. Nutt and Backoff^[29] articulate that such agency would be able to serve a greater variety of clients in several new ways and improve the efficacy of the agency's operations.

REFERENCES

1. Simon, H.A. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*, 3rd Ed.; The Free Press: New York, 1976.
2. Denhardt, R.B. *Theories of Public Organization*; Wadsworth, Inc.: Belmont, CA, 1984.
3. Starling, G. *Managing the Public Sector*, 5th Ed.; Harcourt Brace and Company: Orlando, FL, 1998.
4. Simon, H.A. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*, 3rd Ed.; The Free Press: New York, 1976.
5. Allison, G.T. *Essence of Decision: Explaining the Cuban Missile Crisis*; Little, Brown and Company: Boston, MA, 1971.
6. Allison, G.; Zelikow, P. *Essence of Decision: Explaining the Cuban Missile Crisis*, 2nd Ed.; Addison-Wesley Educational Publishers Inc.: New York, 1999.
7. Stevenson, M.K.; Busemeyer, J.R.; Naylor, J.C. Judgment and Decision-Making Theory. In *Handbook of Industrial and Organizational Psychology*, 2nd Ed.; Dunnette, M.D., Hough, L.M., Eds.; Consulting Psychologists Press: California, 1990; 283-374.
8. Edwards, W. The theory of decision making. *Psychol. Bull.* **1954**, *51*, 380-417.

9. Allison, G.; Zelikow, P. *Essence of Decision: Explaining the Cuban Missile Crisis*, 2nd Ed.; Addison-Wesley Educational Publishers Inc.: New York, 1999.
10. Janis, I.L.; Mann, L. *Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment*; The Free Press: New York, 1977.
11. McCormick, J. The wisdom of Solomon. *Newsweek* **1987, August 17**, 110, 62–63.
12. Bendor, J.; Hammond, T.H. Rethinking Allison's models. *Am. Polit. Sci. Rev.* **1992, June**, 86 (2), 300–320.
13. Przeworski, A. *Democracy and the Market: Political and Economic Reforms in Eastern Europe and Latin America*; Cambridge University Press: New York, 1991.
14. March, J.G.; Simon, H.A. *Organizations*; Wiley: New York, 1958.
15. George, J.M.; Jones, G.R. *Organizational Behavior: Understanding and Managing*; Addison-Wesley Publishing Company: New York, 1996.
16. Simon, H.A. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*, 3rd Ed.; The Free Press: New York, 1976.
17. Lindblom, C.E. The science of muddling through. *Public Adm. Rev.* **1959**, 19 (2), 78–88.
18. Tarter, C.J.; Hoy, W.K. Toward a contingency theory of decision making. *J. Educ. Adm.* **1998**, 36 (3), 212–228.
19. True, J.L. Avalanches and incrementalism: Making policy and budgets in the United States. *Am. Rev. Public Adm.* **2000, March**, 30 (1), 3–18.
20. True, J.L. Avalanches and incrementalism: Making policy and budgets in the United states. *Am. Rev. Public Adm.* **2000, March**, 30 (1), 3–18. Retrieved January 3, 2002 from EBSCOhost Full Display on-line database on the World Wide Web: <http://ehostvgw12.epnet.com/ehost.asp?>
21. Kiel, L.D. *Managing chaos and complexity in government: A new paradigm for managing change, innovation, and organizational renewal*; Jossey-Bass Inc.: California, 1994.
22. Jurow, S. Change: The importance of the process. *Educom Rev.* **1999, Sept./Oct.**, 34 (5), 60, 1 p. Retrieved August 22, 2001 from EBSCOhost Full Display on-line database on the World Wide Web: <http://ehostvgw12.epnet.com/ehost.asp?>
23. Limerick, D.; Passfield, R.; Cunnington, B. Transformational change: Towards an action learning organization. *Learn. Organ.* **1994**, 1 (2), 29–40. Retrieved January 3, 2002 from EBSCOhost Full Display on-line database on the World Wide Web: <http://ehostvgw12.epnet.com/ehost.asp?>
24. Edelman, L.F.; Benning, A.L. Incremental revolution: Organizational change in highly turbulent environments. *Organ. Dev. J.* **1999, Winter**, 17 (4), 79. Retrieved July 27, 2001 from ProQuest on-line database on the World Wide Web: <http://proquestmail@bellhowell.infrolearning.com>.
25. Dyck, B. Understanding configuration and transformation through a multiple rationalities approach. *J. Manag. Stud.* **1997, September**, 34 (5), 793–823. Retrieved January 3, 2002 from EBSCOhost Full Display on-line database on the World Wide Web: <http://ehostvgw12.epnet.com/ehost.asp?>
26. Sastry, M.A. Problems and paradoxes in a model of punctuated organizational change. *Adm. Sci. Q.* **1997, June**, 42 (2), 237–275. Retrieved July 27, 2001 from ProQuest on-line database on the World Wide Web: <http://proquestmail@bellhowell.infrolearning.com>.
27. Porras, J.I.; Robertson, P.J. Organizational Development: Theory, Practice and Research. In *Handbook of Industrial and Organizational Psychology*, 2nd Ed.; Dunnette, M.D., Hough, L.M., Eds.; Consulting Psychologists Press: Paola Alto, CA, 1992.
28. Backer, T.E. Managing the human side of change in VA's transformation. *Hosp. Health Serv. Adm.* **1997, Fall**, 42 (3), 433–459. Retrieved July 27, 2001 from ProQuest on-line database on the World Wide Web: <http://proquestmail@bellhowell.infrolearning.com>.
29. Nutt, P.C.; Backoff, R.W. *Facilitating Transformational Change*; The Journal of Applied Behavioral Science **1997**, 33 (4), 490–508. Retrieved July 27, 2001 from ProQuest on-line database: <http://proquestmail@bellhowell.infrolearning.com>.

Decision Making, Open System and Non-Equilibrium

Keith Carrington

*Criminal Justice and Legal Studies Division, St. John's University,
Jamaica, New York, U.S.A.*

INTRODUCTION

Making decisions is a normal part of our daily routine. Most of our decisions are made with very little processing or reviewing of alternatives and their consequences. While we make most decisions singularly, we, however, do not act as individual decision makers on all matters affecting our life. There are instances when we make decisions as members of a group. Whether we are involved in the decision-making process as individuals or members of a group, many of the decisions we make have far-reaching consequences.

Decision making is an important organizational function. It is important to public sector organizations because of their important social function. In organizations, individuals and groups make decisions. Many of these decisions not only affect the lives of the agency's clientele but also determine the organization's viability. Not long ago, decision making in organizations was considered the domain of the upper echelon. This notion has since been dispelled with the realization that workers at all levels are involved in the decision-making process. At the lower levels of an organization, decisions are made routinely about operational and other job-related issues. At the upper echelon, however, decision making takes on an additional element as managers deal with complex problems that require a high level of expertise.

Decision making is a "rational, deliberate, purposeful action, beginning with the development of a decision strategy and moving through implementation and appraisal of results."^[1] The process consists of three important elements: intelligence, design, and choice. Intelligence involves scanning the environment and arriving at a decision point. The second element is design, which involves developing alternative courses of action. Finally, choice is selecting the alternative with the best chance of success.^[2]

Many types of decisions are made in organizations. These decisions can be placed into two categories: programmed and non-programmed decisions. Programmed decisions are made in response to recurring, predictable problems. Decision makers respond to recurring problems by automatically implementing performance programs worked out beforehand (see Fig. 1). Performance programs are standard procedures

of behavior that employees follow. An organization may have several performance programs, as some problems may recur. Many programmed decisions evolve from non-programmed decisions implemented to deal with problems that occurred earlier.^[2-4] These programs are not "written in concrete," and should be reviewed to determine when changes might be necessary.

There are advantages to using programmed decisions. First, time is not wasted searching for information to develop responses to problems that are predictable and recur frequently. Second, these decisions allow the decision maker to respond quickly to threats or problems from the environment, which, if not resolved promptly, could lead to negative consequences. Third, programmed decisions allow decision makers to implement specifically designed strategies, thereby avoiding errors in the decision-making process.

Non-programmed decisions are the second type of decisions. These decisions address novel, unstructured, complex, and non-programmable programs encountered for the first time.^[2-4] This type of decision making requires superior skills and expertise. Decision makers follow particular steps that may include: identifying the problem; collecting information; examining alternatives and consequences; selecting a course of action; implementing the decision; and evaluating.

Framing a strategy to address a novel problem is not an easy task. Two points are worth mentioning. First, decision makers must accurately identify the problem, which would make the underlying causes unambiguous. This minimizes errors in framing and selecting an appropriate response. Second, decision makers must know the upper and lower limits of the decision. According to Sorenson^[5] and Starling,^[6] the upper limits are as follows:

1. The limits of permissibility (Is it legal? Will others accept it?).
2. The limits of available resources.
3. The limits of available time.
4. The limits of previous commitments.
5. The limits of available information.

The lower limit of a decision "refers to what, at least, must occur for the problem to be solved."^[6] The lower limits inform the decision maker of the

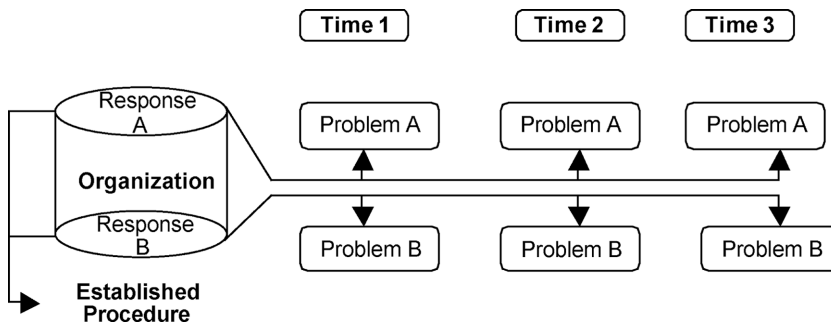


Fig. 1 Programmed decision response to recurring problems.

necessary factors that must be present for the decision to address the problem or situation identified.

There are several decision-making models, which give decision makers an array of choices. A few of the models are listed below:

1. Incrementalism.
2. The Rational Actor Model.
3. The Organizational Process Model.
4. Government Process Model.
5. The Administration Model (see *Decision making, Incrementalism and Transformational Change* for a discussion of these models).

No model has been proffered as the “one best” model for decision making, and the use of a model does not guarantee success. The best model would be the one that resolves the problem, and it may comprise elements from several models. Adhering steadfastly to one approach to make a decision may preclude the successful resolution of a problem. Tarter and Hoy^[1] remind us,

None of the models is a substitute for reflective thinking. None should be used rigidly as to prevent the recognition of an elegant solution in search of a problem ... When models are depicted in diagrams or schemata, it is tempting to view them as lock-step procedures to follow blindly. This is wrong as seeing all decision making as so idiosyncratic as to deny patterns.

Many of the models provide a theoretical approach to making decisions. Many scholars and practitioners propound that some models project a “cookie cutter” approach to decision making, and many of their features have very little applicability to real-world decision making. The lesson for decision makers is that while the models provide guidelines, there are differences in how decisions are made in the real world. This criticism is not sounding the death knell for the models, as several of their premises are used in real-world decision making. Real-world decision making offers decision makers several important lessons. First, government’s most important decisions are made by groups of people and not individual decision makers

working alone. Second, decision making is not a completely rational process. Third, people involved in the decision-making process possess different levels of analytical skills. Fourth, there are real limitations to applying strict analytical methods to all aspects of any problem.^[6]

While Simon advocated rationality in the decision-making process, Allison and Lindblom’s work shifted the focus to the influence of environmental factors on the decision-making process. This shift showed organizations not as isolated units concerned with only internal factors, but entities subject to the influence of the environment.^[7]

OPEN SYSTEMS

The open-system perspective has some of its roots in the intellectual ferment that emerged shortly after World War II. The works of von Bertalanffy^[8] fostered interest in open systems by showing that systems theory could be applied to many entities, including organizations.^[9] The system’s perspective views an organization as a complex entity with several intertwined and interconnected elements that continuously interact with the environment.^[10] Open-systems organizations are “systems of interdependent activities linking shifting coalitions of participants; the systems are embedded in—dependent on continuing exchanges with and constituted by—the environments in which they operate.”^[9] An open-systems organization not only has interchanges with the environment but also needs the vital interchanges to maintain its viability.^[11]

Organizations are open systems because of their energetic input–output system, interaction, and dependence on the external environment. Organizations were not always considered open systems. Before the application of the open-systems perspective, social scientists used one of two approaches to examine social structures. First, they regarded the structures as closed systems and applied the laws of physics and ignored environmental factors. The second approach endowed a self-determining concept to the structures and attributed their functionality to some magical purposiveness.^[12]

In a closed system, each part contributes to the overall working of the organization, the primary concern is efficiency, and the agency is considered insulated from environmental influences. The organization then is in full control and therefore focuses on functionality. Taylor's scientific management and Weber's concept of bureaucracy are examples of such systems.^[7,9,13]

In contrast to closed systems, where structures were thought to be relatively self-contained and independent, the open-systems perspective views the organization as acutely dependent upon the external environment. The organization has continuous exchanges with the environment. It uses the energy and materials from the environment to produce goods and services that are delivered to its clientele. Such organizations are susceptible to challenges and uncertainties, as managers do not possess full knowledge of the factors that could influence the organization's functions or have the ability to predict and control the factors that can influence the organization.^[12]

Notwithstanding the perilous environments that are fraught with formidable challenges, the open-systems organization continues its relationship with the environment, which is crucial to its survival. This interaction can facilitate organizational adaptation and change. Failure to change can lead to an organization's demise. Neumann, Jr.^[14] articulates that the major challenge facing modern organizations is how they could effectively adapt to their rapidly changing environments. He argues that modern organizational structures have been devised throughout history to meet the demands of the increasingly complex external environments, and expounds:

As with species that fail to adapt to changes in their environmental niches, so too non-adaptive organizations do not survive. The key to coping with external environments generally requires that the created structures of the organizations be compatible with the existing dynamics of the external world. Such a fundamental requirement should be intuitive, but in practice, many organizations appear not to understand this necessity.

Trist and Emery^[15] concur that some type of adaptation or evolution is necessary for an organization to survive in an increasingly turbulent environment. They believe that the evolution to open systems benefits the organization. Heckman^[16] also agrees that adopting an open system "maximizes the flexibility and adaptive potential of all its members..." thereby allowing the organization to sustain itself in a turbulent and challenging environment.

How can a complex organization adapt and sustain itself in an unpredictable environment? Philip Selznick's^[17] study of the Tennessee Valley Authority

(TVA) provides insights into some of the methodologies employed to deal with environmental threats. His work on the TVA outlined an open-systems approach to organizational analysis. The analysis emphasized the importance of the relationships between the organization and its environment. The study highlights the intricate system of relationships, which is important to an organization's stability and continuity. These relationships emerge as organizations endeavor to meet their needs and handle numerous problems.

One focus of the study was on the major strategic decisions to deal with threats in the environment. The mechanism used to handle environmental threats was co-optation, which safeguarded the organization and built relationships with social forces in the environment. For example, the TVA leadership co-opted the agricultural leadership—a threat from its environment—into its policy-making structure. The goal of this strategic move was to glean support for the organization's policies. Despite the benefits that could redound through the open-systems approach, many organizations attempt to operate in a closed system. This choice may be foolhardy, as organizations pursuing such isolation never achieve their goals.

NON-EQUILIBRIUM

One view of organizations is that they operate in a state of equilibrium for a very long period, where activities would be consistent with their dictates and structures. Public sector agencies are seen by some as examples of organizations striving to maintain equilibrium through stability. Some scholars regard stability and equilibrium as the "normal" and preferred conditions of organizations.^[18] Stability does not prevent change. Occasionally, punctuated change is made incrementally to maintain equilibrium. Some believe that the quest for equilibrium is dysfunctional and could lead to an organization's demise.

An opposing view is that organizations function under non-equilibrium conditions, and due to the environmental threats, are constantly adapting to sustain themselves. Non-equilibrium conditions are seen as prerequisites for an organization's growth and adaptation.

Limerick, Passfield, and Cunnington^[19] stated that the field has seen stable, incremental change and rapid organizational change. They reconcile the two types of changes by proffering Gould's model of "punctuated equilibrium." He stated:

Speaking most generally, our usual concept of time and history, abetted by traditional gradualism, views change as continuous and intrinsic, and structure as a temporary incarnation of the moment. The alternative

view, the basic vision underlying catastrophism, if you will, sees stable structure and organization as the usual state of things, and change as a rare and disruptive event provoking rapid shifts from one configuration to another. I believe that we are now witnessing, in field after field, a growing recognition of this punctuational style for change at all scales (cited in Limerick, Passfield, and Cunningham).

This perspective advocates that there are long periods of equilibrium punctuated by short periods of non-equilibrium, which resulted in the reconfiguration of organizations. Some believe that equilibrium and non-equilibrium conditions are interrelated.^[19]

Not all scholars and practitioners support the concept of punctuated change. Many believe that the unpredictable environment creates the non-equilibrium conditions in which organizations operate. Many suggest that non-equilibrium conditions prevail in an unpredictable, uncertain and, according to Doyle,^[20] “an unforgiving and, at times even hostile environment.” These conditions, which are fraught with pitfalls, compel organizations to change. Arthos^[21] articulates that non-equilibrium conditions have emerged as one of the natural features of open systems, and the non-equilibrium conditions encourage ongoing attempts to transform organizations. These transformational change attempts are made frequently; however, most of these attempts are rejected.^[22]

The equilibrium paradigm no longer holds firmly in the public sector. Practitioners and scholars now acknowledge the non-linear nature of public sector organizations and the instability and turbulence these organizations confront.^[18] These conditions present a formidable challenge to public agencies in their quest to deliver services. Government agencies are open systems, as they extract (taxes and fee) and release energy (services) through interaction with the environment. There may be demands from citizens and employees for a different method of service delivery. These pressures may force the organization to develop a novel service delivery.^[23]

The non-equilibrium perspective views the public administration environment in two ways. First, the environment comprises a mosaic of organizations at different stages of their evolution. This montage contains declining, self-regulating, and extremely unstable organizations; the latter may experience symmetry breaks and novel reconfigurations. The second view holds that the intrusion of politics on administration can enhance instability, which may lead to symmetry-breaking behavior. He argues against the politics/administration dichotomy, because politics may be needed to push public sector organizations beyond stability. Consequently, politics and administration must be inexorably linked.^[18] The non-equilibrium

conditions in which public sector organizations function demands continued renewal and change to remain viable. To ensure survival, it is necessary for organizations to change to continue operations in a non-linear environment.

NON-EQUILIBRIUM THEORY

Change is an accepted feature of organizations, and knowledge of the change process is important to practitioners and scholars. It is, therefore, important that the explanatory models used should be appropriate to bring to light invaluable lessons. Incremental and equilibrium models are still being used to explain how organizations change to adapt to their environment. However, these models give only a variation on the issue. Due to the non-equilibrium conditions and chaos organizations encounter while interacting with the environment, these models do not properly explain the change process. These models have failed to encompass the complexity of development of social systems and do not “incorporate those instances when wholesale, dramatic system changes occur.”^[18]

The inability of models to explain the changes that occur highlights the need for a new approach. Kiel^[18] proffers non-equilibrium theory, which he believes would provide an in-depth understanding of the evolutionary nature of the “process of change in which instability, disorder, and unpredictability serve as the central features in the development of new forms organization and complexity.” In addition, the theory gives insights into the change process leading to new configurations and can complement incremental models, thereby elucidating the entire evolutionary processes in systems relevant to public administration.

CONCLUSION

The cataclysmic events of September 11, 2001, underscore the unpredictable and perilous environmental conditions that can befall public and private sector organizations. The fallout from the events has affected all sectors. This has demonstrated several important lessons. First, organizations cannot isolate themselves from their environments. Even if they are not affected directly, ripple effects from turbulence in their environments could have an impact on an organization's overall performance. Second, although the September events are not typical of the environments in which organizations operate, they indicate that organizations would always face uncertainties and challenges. Third, change and adaptation must be a feature of organizations if they are to survive and remain viable in non-equilibrium conditions.

Many may argue that an open-systems organization is better equipped to deal with environmental threats. An advantage a decision maker has in an open systems organization is that he or she can use several methods to achieve the organization's goals; which is a defining characteristic of open systems.^[12] Kast and Rosenzweig^[24] argued that the dichotomization of systems has led to the belief that open systems are good, and closed systems are bad. They, however, believe that most social organizations are "partially open" and "partially closed," and that both approaches could be used under certain circumstances.

REFERENCES

1. Tarter, C.J.; Hoy, W.K. Toward a contingency theory of decision making. *J. Educ. Adm.* **1998**, *36* (3), 212–228. Retrieved January 4, 2002 from EBSCOhost Full Display on-line database on the World Wide Web: <http://ehostvgw12.epnet.com/ehost.asp?>
2. Simon, H.A. *The New Science of Management Decision*; Harper & Brothers Publishers: New York, 1960.
3. Browne, M. *Organizational Decision Making and Information*; Ablex Publishing Corporation: Norwood, NJ, 1993.
4. George, J.M.; Jones, G.R. *Understanding and Managing Organizational Behavior*; Addison-Wesley Publishing Company: New York, 1996.
5. Sorenson, T.C. *Decision Making in the White House*; Columbia University Press: New York, 1963.
6. Starling, G. *Managing the Public Sector*, 5th Ed.; Harcourt Brace College Publishers: New York, 1998.
7. Denhardt, R.B. *Theories of Public Organization*; Wadsworth, Inc.: Belmont, CA, 1984.
8. von Bertalanffy, L. *General System Theory: Foundations, Developments, Applications*, Revised Ed.; George Braziller: New York, 1968.
9. Scott, W.S. *Organizations: Rational, Natural, and Open Systems*; Prentice Hall: Upper Saddle River, NJ, 1998.
10. Shafritz, J.M.; Ott, J.S. *Classics of Organizational Theory*, 3rd Ed.; Brooks/Cole Publishing Company: Pacific Grove, CA, 1992.
11. Swanson, D. Dysfunctional conglomerates: an explanation provided by linking ontological individualism to social relationships within an open system. *Behav. Sci.* **April 1992**, *37* (2), 139, 14 pp. Retrieved from Retrieved August 8, 2001 from EBSCOhost Full Display on-line database on the World Wide Web: <http://ehostvgw12.epnet.com/ehost.asp?>
12. Katz, D.; Kahn, R.L. Organizations and the system concept. In *Classics of Organization Theory*, 3rd Ed.; Shafritz, J.M., Ott, J.S., Eds.; Brooks/Cole Publishing Company: Pacific Grove, CA, 1992; 270–280.
13. Navatorava, V.N.; Crompton, J.L. A revised conceptualization of marketing in the context of public leisure services. *J. Leis. Res.* **2000**, *33* (2), 160–185. Retrieved August 22, 2001 from ProQuest on-line database on the World Wide Web: <http://proquestmail@bellhowell.infrolearning.com>.
14. Neumann, F.X., Jr. Organizational structures to match the new information-rich environments: lessons from the study of chaos. *Public Prod. Manage. Rev.* **September 1997**, *21* (1), 86–100. Retrieved from Retrieved January 4, 2002 from EBSCOhost Full Display on-line database on the World Wide Web: <http://ehostvgw12.epnet.com/ehost.asp?>
15. Trist, E.L.; Emery, F.E. The causal texture of organizational environments. *Hum. Relat.* **1965**, *18*, 21–32.
16. Heckman, F. Designing organizations for flow experiences. *J. Qual. Partic.* **March 1997**, *20*, 2. Retrieved from Retrieved January 4, 2002 from EBSCOhost Full Display on-line database on the World Wide Web: <http://ehostvgw12.epnet.com/ehost.asp?>
17. Selznick, P. *TVA and the Grass Roots*; Harper & Row: New York, 1949.
18. Kiel, L.D. Nonequilibrium theory and its implications for public administration. *Public Adm. Rev.* **1989**, *49*, 544–551. Retrieved October 8, 2001 from ProQuest on-line database on the World Wide Web: <http://proquestmail@bellhowell.infrolearning.com>.
19. Limerick, D.; Passfield, R.; Cunnington, B. Transformational change; Towards an action learning organization. *Learn. Organ.* **1994**, *1* (2), 29–40. Retrieved from Retrieved January 4, 2002 from EBSCOhost Full Display on-line database on the World Wide Web: <http://ehostvgw12.epnet.com/ehost.asp?>
20. Doyle, M. Organizational transformation and renewal: a case for reframing management development? *Personal Rev.* **1995**, *24* (6), 6–18. Retrieved from Retrieved January 4, 2002 from EBSCOhost Full Display on-line database on the World Wide Web: <http://ehostvgw12.epnet.com/ehost.asp?>
21. Arthos, J. Locating the instability of the topic places: rhetoric, phronesis and neurobiology. *Commun. Q.* **Summer 2000**, *48*, 3. Retrieved October 4, 2001 from ProQuest on-line database on the World Wide Web: <http://proquestmail@bellhowell.infrolearning.com>.
22. Dyck, B. Understanding configuration and transformation through a multiple rationalities approach. *J. Manag. Stud.* **1997**, *34* (5). Retrieved from Retrieved January 4, 2002 from EBSCOhost Full Display on-line database on the World Wide Web: <http://ehostvgw12.epnet.com/ehost.asp?>
23. Kiel, L.D. *Managing Chaos and Complexity in Government: A New Paradigm for Managing Change, Innovation, and Organizational Renewal*; Jossey-Bass: San Francisco, 1994.
24. Kast, F.E.; Rosenzweig, J.E. General systems theory: applications for organization and management. In *Classics of Organization Theory*, 3rd Ed.; Shafritz, J., Ott, J.S., Eds.; Brooks/Cole Publishing Company: Pacific Grove, CA, 1992; 294–307.

Decision Making, Rational

E. J. Woodhouse

Science and Technology Studies, Rensselaer Polytechnic Institute,
Troy, New York, U.S.A.

INTRODUCTION

In everyday life, an action is considered rational to the extent that it embodies good sense, including logic, proportionality, and appropriate fit with its context. In the mid-twentieth century, however, the term “rational decision making” came to be appropriated by systems theory, multiattribute utility analysis, and other hypercognitive approaches to figuring out what to do in complex situations. This article summarizes aspects of that specialized usage but also points to many other traditions of political thought that contribute to understanding the requisites for reasoned social choice.

Complex, ideologically contested choices typify governmental decision making, and analysis alone cannot determine how to cope sensibly with such situations. Calculation can assist in negotiating viable policy, but cannot take the place of ordinary reasoning and political deliberation. Hence, fairer and wiser governance depends on improved political processes more than on better analysis, and I argue for reappropriating the term “rational” so as to recapture its underlying spirit and refocus on the political.

This article analyzes aspects of that usage in the context of other traditions of thought bearing on reasoned social choice. I argue for reappropriating the term “rational” so as to recapture its original spirit.

CALCULATION AS PART OF POLITICS

In everyday usage, the term “rational” typically contrasts a reasoned mode of thought or action with one that is “irrational,” in the sense of being unreasonably distorted by emotion, bias, or ideology. Thus, a housing bureaucrat asks others in his profession: “Is our only choice between ‘solutions based on cool, rational decision-making by the experts’ and solutions based on the ‘passions’ of a mobilized mass movement?’” His answer, of course, is that “Rational decision-making in the production and rehabilitation of affordable housing is not incompatible with community organizing and direct action.”^[1] Likewise, working within a commonsensical meaning of the term is a professor at the Royal Military Academy in Brussels, Belgium, advocating what he calls “rational military

decision making,” choices that are “both ethical and efficient.”^[2]

The first systematization of rational decision making in political science was Dahl and Lindblom’s *Politics, Economics, and Welfare*, which drew on a long tradition of political thought recognizing that “Discussion is perhaps the oldest and most universal process for facilitating rational calculation in social action.” They analyzed “an extensive variety of social processes that facilitate rational calculation,” including two contrasting approaches to reducing the number and complexity of variables: quantification and delegation.^[3] Sensible collective action, they were saying, depends on a diverse combination of methods, some analytic and some social, and these methods typically intertwine, as when tasks are delegated to experts.

More generally, a variety of social science scholarship clarifies barriers facing intelligent collective action while illuminating possibilities for circumventing or lowering those barriers. Building on the old saying attributed to Clement Atlee, “He who controls the agenda controls the meeting,” research on agenda setting arguably attacks one of the most important barriers to intelligent governance.^[4–6] Even more fundamental than governmental agenda setting are the psychological, sociocultural, and political processes by which people come to frame how they think about issues.^[7,8] If the attack on Pearl Harbor inherently constituted an act of war, then launching World War II against the Japanese was a foregone conclusion, and there merely remained a few details about how to do it. In contrast, if international relations theorist Bruce Russett was correct in determining that Pearl Harbor did not constitute a clear and present danger to U.S. national security, then a much wider array of options might usefully have been considered.^[9] Framing is key.

Social and political psychologists who study conflict resolution and group and organizational choice observe situational factors leading to patterned, less-than-intelligent behaviors such as groupthink.^[10] Among the approaches offered to ameliorate problems that are complex and ill-structured and which change as problem solving advances, there is “double-loop learning.” This emphasizes learning to change underlying values and assumptions^[11] by interacting with others to induce questioning of one’s “theory of

action,’’^[12] resulting in better acceptance of failures and mistakes, yielding a more realistic approach to problem solving.^[13,14] Much of the voluminous literature on organizations bears on improving their functioning and, hence, on rational decision making.

ENTER HYPERRATIONALITY

The mix of analysis and politics assumed in the bulk of social science scholarship on decision making is not well reflected in the treatments of decision theory with which the label “rational” has been most closely associated for the past several decades. Hyperrational decision approaches attempt to describe and sometimes prescribe how individuals (or groups making a joint decision) do and should choose among alternatives in accord with their preferences and understandings.

The basic elements of utility theory were initiated long ago by Bernoulli,^[15] Bentham,^[16] and Pareto,^[17] and translated in the mid-twentieth century to the goal of maximizing “expected utility.” That sounds sensible, but operationalizing the intention turns out to be pretty difficult, as some critics recognized quite early.^[18] Step one in figuring out the expected utility of an action typically is to multiply the probability of a predicted outcome by the utility of that outcome. Because most political actions produce multiple outcomes, step two is to add the utilities of all the outcomes. That sounds cumbersome, but it is not too bad until one realizes that political participants hardly ever really know the probabilities or the utilities, which begins to undermine one’s confidence in the whole operation. The hyperrationalists are smart folks, of course, so they have partial responses to these problems; at bottom, however, the dividing line between the hyperrationalists and most social thinkers is that the majority just does not buy the basic story that utility maximization is possible or that it can be achieved via calculation.

von Neumann and Morgenstern’s^[19] work on game theory sometimes is considered a starting point for contemporary formal decision theory, but attention to ways of dealing with risk and uncertainty go back much farther through Charles Knight to earlier economists, including Marshall. Luce and Raiffa^[20] popularized the term “decision theory,” and Raiffa^[21] formalized a workable approach to “decision analysis” on which continuously refined work continues on topics such as revealed probabilities and the structure of individual utility functions.^[22] Cost–benefit, risk–benefit analysis and probabilistic risk assessment were offshoots, the latter focusing on probabilistic analysis of risky endeavors, such as nuclear power plants.^[23]

Rational decision advocates often attempt to break choices down into a series of steps: define the problem, identify criteria, weight the criteria in terms of their relative importance, generate alternatives, rate each alternative on each criterion, and “compute the optimal decision.”^[24] There may have been a time when some analysts were naïve enough to assume this process unproblematic, but most now recognize lots of problems with it. Because human memory and other cognitive capacities are so limited, we proceed by bounded rationality, meaning we tend to focus on just one or two criteria instead of the full range that might be sensible.^[25] Criteria are tough to compare, and relative importance depends on the situation: lower taxes are attractive, but at the expense of driving science teachers into industry? How often does one actually know the probability that a given action will have a given effect? And what about unintended consequences?

Emotion and interpersonal cuing tend to be extremely important to most humans, and trying to squeeze it all out of decision choices actually can lead to counterproductive outcomes. Thus, probabilistic risk analysis was deployed with great skill at high expense by nuclear power advocates, but many people just did not feel comfortable with giant power plants that could suffer even a small risk of catastrophic meltdown.^[26] Even economists are increasingly questioning their field’s conventional assumptions about human choice.^[27,28] Thus, it no longer is weird to read in the *Journal of Economic Literature* that:

Because we are social animals, the collective behavior of humans must be understood in terms of the cultural forces at work. World views, mores, taboos, and sanctions are among the social influences that shape and guide our behavior, so much so that characterizing our activities as “choice” and “decision making” is often inappropriate. This is no less true in market settings than it is elsewhere.^[29]

Behavioral Decision Theory or Heuristic Judgment Theory is generally regarded as a more realistic approach to understanding human choice, in part, because it is based on empirical observation of the guidelines and heuristics people actually use when they make decisions.^[30] For example, when particular images are highly “available” cognitively (World Trade Center collapse, shark attack, fire), people’s thinking will be more influenced than when dangers are less vivid (flood, radon in the basement, stress from gradually accumulating sleep deprivation). Individuals, groups, and organizations depart rather significantly from the choices experts say would be most in line with real-world probabilities.^[31] We have weak capacities for estimating probabilities and general cognitive limits for processing information.^[32] Simon^[33] argued that

making a choice that is good enough, satisficing, usually makes more sense than trying to optimize; and in everyday practice, this certainly is how most humans and their organizations behave (when they behave sensibly, which, of course, is far from routine). The psychology-based approach, therefore, is more helpful in understanding real life than are the various cognitive approaches derived from conventional economic assumptions about “rational man”; but the psychological approach also promises a good deal less than does rational choice (for a direct comparison, see Ref.^[34]).

RATIONAL CHOICE

Differing substantially from these policy-analytic endeavors is Rational Choice Theory, founded in part by Arrow,^[35] Downs,^[36] Buchanan and Tullock,^[37] and Olson.^[38] Practitioners of rational choice are generally concerned with understanding rather than with prescription and focus more on political behavior than on the substance of public policy. But, their approach was the most prestigious and rapidly growing part of political science in the last quarter century, colonizing many subfields of the discipline by drawing on economists’ simplifying assumptions about human behavior. So, one cannot speak of rational decision making without taking rational choice theory into account.

Rational choice begins with simple assumptions about individuals having incentives to maximize income, economize on time, and otherwise get what they want, and it attempts from this microfoundation to develop understandings of why people and institutions function as they do in politics. Blending with game theory in some hands (e.g., Ref.^[39]), analysts attempt to show why most of what happens in politics is “rational.” For example, whereas most observers of Congress perceive many different factors influencing assignments of legislators to committees, rational choice theorists try to demonstrate an underlying cause: committee assignments maximize the probability of policy outcomes that suit the majority of legislators in each party.^[40]

Despite widespread professional admiration for the theoretical rigor, most social thinkers find the assumptions used so simpleminded as to undermine the believability of the claims. Moreover, practitioners of rational choice have devoted far more effort to refining their theorems than to empirical tests of their ideas.^[41]

COPING WITH UNCERTAINTY AND DISAGREEMENT

Another way to tell the story of the “rational-comprehensive” approach in decision theory is to

emphasize not its intellectual but its cultural roots.^a The receptivity to calculation as a substitute for political interaction stems from its compatibility with the mastery tradition in western thought: Techniques to conquer uncertainty have long been popular, manifested in the physical sciences and engineering as well as in religion.

Lindblom systematized and effectively debunked the folk theory of rational-comprehensiveness in “Muddling Through” and subsequent work, showing that rational-comprehensive aspirations make impossible demands for information, causal theory, time, and other resources.^[43–45] For example, as Alice Rivlin put it, “Little progress has been made in comparing the benefits of different social action programs,” and “little is known about how to produce more effective health, education, and other social services.”^[46]

Fortunately, although there is no way to calculate the “correct” approach to a complex policy issue or to analytically set the agenda or to even define what “the problems” are, political participants declare what they consider the high-priority problems, and they propose solutions. Such proposals are never proven superior to other possible problem definitions, agendas, and policy options, but fortunately, proof is unnecessary, because action is undertaken when a working majority reaches agreement.

Can mere agreement really be considered an intelligent way to run a society? To evaluate that, one needs some way of judging what constitutes an “intelligent” political process. No one has a definitive set of criteria for that, but many people might look for at least the following:

1. That concerns moderately or strongly held by any sizeable number of people are taken into account in some non-trivial way.
2. That reasonable tradeoffs are made among conflicting values.
3. That, insofar as feasible, policy actions take into account available information about social problems and opportunities, performance of existing programs, costs, and other relevant matters.

Although no political system is presently designed to achieve these elements of intelligent policy making nearly as well as can be envisioned, it turns out that negotiations and other adjustments among political partisans can often achieve a measure of all of these goals. How? Consider some of the ways partisans are likely to think about making and responding to policy proposals:

1. In crafting a proposal, partisans will tend to look for ways to get what they want without

^aThis section is drawn largely from Ref.^[42].

stirring up adverse responses or retaliations from others. So, the search for agreement simultaneously becomes a method of taking diverse views into account.

2. Knowing that they will need allies when putting forth their own proposals, others are disposed to grant what proposers ask, unless there is some good reason not to do so. Thus, a bias toward action is built into the system: I will help you get what you need, if you will help me get what I need.
3. Those with initial objections have strong motivation to come up with counterproposals, ones that will allow them to join with the original proposer in seeking a mutually beneficial outcome.
4. The need to win agreement keeps demands within the sphere likely to be considered “reasonable” or intelligent by most of those whose agreement is needed.
5. Moderation likewise is encouraged, because those putting forth a proposal will be wasting their own scarce time, energy, and other resources if they cannot win agreement; and, even if eventually winnable, immoderate proposals may not be worth the effort and the favors used in the process.

Thus, working for their private gain and for their vision of the public interest, partisans interact with each other in ways that often converge toward fairly sensible outcomes. The need to achieve agreement leads to policy making that, to some degree, aims for the three goals of responsiveness to public sentiment, sensible trade-offs, and attention to relevant information.

These interactions can be thought of as substituting politics for analysis, because no one really “understands” the social problem that is being ameliorated, inasmuch as the participants each hold somewhat different definitions of the problem. Alternatively, political interaction can be said to achieve a form of understanding that cannot be produced through analysis alone. Because understanding is generally sought as a means to improved action, whenever a working majority agrees on a new or revised policy, that policy can be thought of as embodying a new understanding. This accords with one definition of the term understanding: a shared agreement.^[42]

It is not clear that anyone in the field now really advocates rational-comprehensive techniques as the primary method of approaching complex public issues, but various scholars dissatisfied with incrementalism tried to formulate second-best approaches seeking to preserve some of the aims of comprehensive analysis,

such as making sure that important priorities are not overlooked.^[47–49] Despite all the attention to incrementalism and to Simon’s work on satisficing, despite the fact that most policy scholars got the point, those who understand analysis as an adjunct to politics perhaps have not done as well as the hyperrationalists at advancing a systematic program of conceptually and empirically refined research.

However, one strand of qualitative decision theory has built a research program around the fact that social problem solving ordinarily proceeds through trial-and-error learning. How can policy analysts help partisans figure out how to make the inevitable errors less damaging and how to learn from them more rapidly? Three of the main pitfalls in ordinary trial-and-error learning are as follows: 1) a policy trial may produce unbearably costly outcomes; 2) policy moves may retain too little flexibility, preventing errors from being corrected readily; and 3) learning about errors may be very slow. More intelligent trial and error would feature strategies for making errors less costly, for building in greater flexibility, and for speeding up the learning process.^[26,42]

Even in highly uncertain endeavors, it is often possible to partly foresee and protect against some of the worst risks, as the National Institutes of Health did in requiring early biotechnology research to be conducted in special laboratories, sealing in potentially dangerous new organisms.^[50] To prevent policies from becoming deeply enmeshed in implementers’ careers and otherwise resistant to change, in framing policy moves, partisans and analysts can improve their odds of success by developing policy options that can be altered fairly readily, should unfavorable experience warrant. For example, flexibility is higher when a policy’s costs are borne gradually, allowing expenditures to be redirected as learning develops. NASA’s space shuttle illustrates the problem: A launch regime relying on expendable rockets would have been much easier to revamp.^[51,52]

Flexibility also can be enhanced by phasing in a policy during a learning period, a common practice in business; by experimenting in a limited geographical area or for a delimited client base; by conducting simultaneous trials of two or more alternative approaches; by using an existing bureaucracy instead of creating a new, dedicated organization with permanent staff; and by many other tactics.^[53–55]

The research on intelligent trial and error exemplifies a much broader body of policy-oriented scholarship aimed at understanding how to make collective action more intelligent without the simplifying assumptions typically necessitated by the hyper-rational approaches. The spirit of that endeavor is captured by Andrews’^[56] concept of “humble analysis.”

CONCLUSION

In sum, rational decision making is a sprawling topic not defined by any simple, bounded set of literature. The term is contested, with hyperrationalists influenced by economics tending to win disproportionate (but often unfavorable) attention from policy intellectuals and from participants in policy processes. Critics of cost-benefit analysis, game theory, and other purely analytical techniques have been on the defensive for several decades, in part because their techniques are much fuzzier and more difficult to teach and deploy.

But, the critics' intellectual case is convincing: analysis simply cannot be substituted for politics, and virtually no one really argues otherwise. As Johan Olsen put it, hyperrationalists "suggest an approach that assumes away most of the complexity of political actors, the organized settings within which they operate, and institutional change, rather than make a serious effort to understand that complexity."^[57]

For the future, it is safe to assume that one basic challenge will continue to confront those who care about bringing greater intelligence to bear in social problem solving. The potential intelligence of democracy primarily requires that social institutions and processes be arranged so that interaction among diverse persons can find areas of agreement that take into account a great many considerations bearing on each choice. This partially substitutes for rational calculation and partly uses analysis in the service of partisan goals. Yet, those who exercise political and economic authority generally are far more willing to utilize narrow techniques for improving efficiency than they are to improve and redefine effectiveness by listening to social scientists calling for social-organizational changes to broaden the sphere of negotiation and otherwise strengthen the decision-making process.

Hence, "the troubled attempt to understand and shape society" goes to the heart of power relations in contemporary civilization and is by no means a purely academic dispute.^[8] The quest for more intelligent or "rational" collective decision making goes well beyond the terrain conceivably covered by rational choice, cost-benefit, multiattribute utility theory, and other logico-quantitative techniques. Such techniques usually are not aimed at how issues are framed, and they are not aimed at many other factors that actually matter when policy negotiations and impositions are going on; hence, the analytic techniques sometimes called "rational decision theory" actually are most helpful when located within the much broader universe of social thought pertaining to intelligent choice. Rethinking rational decision making to make relevant a broader swath of scholarship in public policy, political science, and social thought more generally probably is essential for narrowing the substantial

discrepancy between aspiration and achievement in moving toward the potential intelligence of democracy.

REFERENCES

1. Davis, J.E. *Rational Decision Making to Balance Ideological Extremes*; National Housing Institute, Shelterforce Online, March/April 1994; 1 pp.
2. Van Damme, G.C. Course syllabus. **2001**, 1 (<http://www.usafa.af.mil/jscope/JSCOPE00/VanDamme00.html,9/30/01>)
3. Dahl, R.A.; Lindblom, C.E. *Politics, Economics, and Welfare*; Harper and Brothers: New York, 1953.
4. Cobb, R.W.; Elder, C.D. *Participation in American Politics: The Dynamics of Agenda-Building*; Allyn and Bacon: Boston, 1972.
5. Rochefort, D.A.; Cobb, R.W. *The Politics of Problem Definition: Shaping the Policy Agenda*; University Press of Kansas: Lawrence, KS, 1994.
6. Cobb, R.W.; Ross, M.H. *Cultural Strategies of Agenda Denial: Avoidance, Attack, and Redefinition*; University Press of Kansas: Lawrence, KS, 1997.
7. Edelman, M. *The Symbolic Uses of Politics*; University of Illinois Press: Urban, IL, 1964.
8. Lindblom, C.E. *Inquiry and Change: The Troubled Attempt to Understand and Shape Society*; Yale University Press: New Haven, 1990.
9. Russett, B.M. *No Clear and Present Danger: A Skeptical View of the U.S. Entry into World War II*; Harper and Row: New York, 1972.
10. Janis, I.L.; Mann, L. *Decision Making*; Free Press: New York, 1977.
11. Argyris, C. *Increasing Leadership Effectiveness*; Wiley: New York, 1976.
12. Argyris, C.; Schon, D. *Theory in Practice*; Jossey-Bass: San Francisco, 1974.
13. Argyris, C. *On Organizational Learning*; Blackwell: Cambridge, MA, 1993.
14. Schon, D.A.; Argyris, C. *Organizational Learning II: Theory, Method, and Practice*; Addison-Wesley: Reading, MA, 1996.
15. Kendall, M.G. Daniel Bernoulli on maximum likelihood. *Biometrika* **1961**, *48*, 1-18.
16. Bentham, J. *Principles of Morals and Legislation*; Oxford University Press: Oxford, 1823.
17. Pareto, V. *Manuel d'Economie Politique*; Marcel Giard: Paris, 1927.
18. Edwards, W. The theory of decision making. *Psychol. Bull.* **1955**, *41*, 380-417.
19. von Neumann, J.; Morgenstern, O. *Theory of Games and Economic Behavior*, 3rd Ed.; Princeton University Press: Princeton, NJ, 1953.
20. Luce, R.D.; Raiffa, H. *Games and Decisions*; Wiley: New York, 1957.
21. Raiffa, H. *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*; Addison-Wesley: Reading, MA, 1968.
22. Viscusi, W.K.; Evans, W.N. Estimation of revealed probabilities and utility functions for product safety decisions. *Rev. Econ. Stat.* **1998**, *80*, 28-33.

23. Crouch, E.A.C.; Wilson, R. *Risk/Benefit Analysis*; Ballinger: Cambridge, MA, 1982.
24. Bazerman, M.H. *Judgement in Managerial Decision Making*, 5th Ed.; Wiley: New York, 2002.
25. Simon, H.A. A behavioral model of rational choice. *Q. J. Econ.* **1955**, *69*, 99–118.
26. Morone, J.G.; Woodhouse, E.J. *The Demise of Nuclear Energy? Lessons for Intelligent Democratic Control of Technology*; Yale University Press: New Haven, 1989.
27. Sen, A. Social choice theory: a re-examination. *Econometrica* **1977**, *45*, 53–89.
28. Arrow, K.J. Rationality of self and others in an economic system. In *Decision Making: Alternatives to Rational Choice Models*; Zey, M., Ed.; Sage: Newbury Park, CA, 1992; 63–77.
29. Harvey, J.T. Heuristic judgment theory. *J. Econ. Issues-March* **1998**, *32*, 47–64.
30. Kahneman, D., Slovic, P., Tversky, A., Eds.; *Judgment Under Uncertainty: Heuristics and Biases*; Cambridge University Press: Cambridge, 1982.
31. Dawes, R.M. *Rational Choice in an Uncertain World*; Harcourt Brace Jovanovich: New York, 1988.
32. Rappoport, L.; Summers, D. *Human Judgment and Social Interaction*; Holt, Rinehart and Winston: New York, 1973.
33. Simon, H.A. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*, 3rd Ed.; Macmillan: New York, 1947.
34. Hogarth, R.M.; Reder, M.W. *Rational Choice: The Contrast Between Economics and Psychology*; University of Chicago Press: Chicago, 1986.
35. Arrow, K.J. *Social Choice and Individual Values*; Yale University Press: New Haven, 1951.
36. Downs, A. *An Economic Theory of Democracy*; Harper and Row: New York, 1957.
37. Buchanan, J.M.; Tullock, G. *The Calculus of Consent: Logical Foundations of Constitutional Democracy*; University of Michigan Press: Ann Arbor, MI, 1962.
38. Olson, M. *The Logic of Collective Action*; Harvard University Press: Cambridge, 1965.
39. Brams, S.J. *Rational Politics: Decisions, Games, and Strategy*; Academic Press: Boston, 1985.
40. Kiewiet, R.D.; McCubbins, M.D. *The Logic of Delegation*; University of Chicago Press: Chicago, 1991.
41. Green, D.P.; Shapiro, I. *Pathologies of Rational Choice: A Critique of Applications in Political Science*; Yale University Press: New Haven, 1994.
42. Lindblom, C.E.; Woodhouse, E.J. *The Policy-Making Process*, 3rd Ed.; Prentice Hall: Englewood Cliffs, NJ, 1993.
43. Lindblom, C.E. The science of ‘muddling through’. *Public Adm. Rev.* **1959**, *19*, 78–88.
44. Braybrooke, D.; Lindblom, C.E. *A Strategy of Decision: Policy Evaluation as a Social Process*; Free Press of Glencoe: New York, 1963.
45. Lindblom, C.E. *The Intelligence of Democracy*; The Free Press: New York, 1965.
46. Rivlin, A.M. *Systematic Thinking for Social Action*; Brookings Institution: Washington, DC, 1971; 7 pp.
47. Dror, Y. Muddling through: Science or inertia? *Public Adm. Rev.* **1964**, *24*, 153–157.
48. Etzioni, A. Mixed-scanning: a ‘third’ approach to decision-making. *Public Adm. Rev.* **1967**, *27*, 385–392.
49. Dryzek, J. Complexity and rationality in public life. *Polit. Studies* **1987**, *35*, 424–442.
50. Morone, J.G.; Woodhouse, E.J. *Averting Catastrophe: Strategies for Regulating Risky Technologies*; University of California Press: Berkeley, 1986.
51. Logsdon, J. The decision to develop the space shuttle. *Space Policy* **1986**, *2*, 103–119.
52. Brunner, R.D.; Byerly, R., Jr. The space station programme: defining the problem. *Space Policy* **1989**, *6*, 131–145.
53. Collingridge, D. *Technology in the Policy Process: The Control of Nuclear Power*; St. Martin’s Press: New York, 1983.
54. Collingridge, D. *Management of Scale: Big Organizations, Big Technologies, Big Mistakes*; Routledge: New York, 1992.
55. Woodhouse, E.J.; Collingridge, D. Incrementalism and the future of political decision theory. In *An Heretical Heir of the Enlightenment: Science, Politics, and Policy in the Thought of Charles E. Lindblom*; Redner, H., Ed.; Westview: Boulder, 1993.
56. Woodhouse, E.J. Reframing incrementalism: a constructive response to the critics. *Policy Sci.* **1993**, *25*, 255–273.
57. Andrews, C.J. *Humble Analysis: The Practice of Joint Fact-Finding*; Praeger: Westport, CT, 2002.

BIBLIOGRAPHY

1. Olsen, J.P. Garbage cans, new institutionalism, and the study of politics. *Am. Polit. Sci. Rev.* **2001**, *95*, 191–198.

Decision Making: Administrative Dispute Resolution

Becky A. Nesbit

Lisa B. Bingham

*School of Public and Environmental Affairs, Indiana University,
Bloomington, Indiana, U.S.A.*

INTRODUCTION

Administrative agency use of alternative dispute resolution (ADR) refers to the use of methods other than traditional adjudication, rule-making, litigation, or informal administrative action to resolve conflicts that arise in public administration. This article begins by describing traditional processes for addressing conflict in public administration. Next, this article overviews the most common dispute resolution methods and their potential benefits. Then the article discusses the development and sources of authority for administrative agency use of ADR in public administration as well as the pros and cons of administrative use of ADR. Finally, the article concludes by discussing the implementation of ADR programs and their evaluation in public administration.

CONFLICT RESOLUTION IN PUBLIC ADMINISTRATION

Public agencies have a variety of means to address conflict in making, implementing, and enforcing public policy at the federal, state, and local levels of government. These include traditional processes such as rule-making, adjudication, informal agency action, and litigation as well as various forms of ADR, such as interest-based negotiation, mediation, fact finding, binding and nonbinding arbitration, and other facilitated processes.^[1]

Administrative agency action can take forms analogous to each of the three branches of government; at times, agencies act in a quasi-legislative, quasi-judicial, or executive capacity. Quasi-legislative action includes the formation of policy through the development of guidelines or through informal or formal rulemaking leading to new regulations. The rule-making process is synoptic and prospective; it entails the collection and review of information and comment generally related to creating or fleshing out public policy, after which, the agency publishes draft regulations. Quasi-judicial action includes informal or formal adjudication or contested hearings to determine the rights of specific parties. Adjudication is retrospective; it

entails the creation of a factual record of past events related to the enforcement of existing public policy either in the form of a statute or regulation. Executive action includes the remaining broad swath of informal agency decision-making generally related to implementing public policy. Agencies can also initiate or defend litigation as part of the process of enforcing policy.

Administrative use of ADR includes three main types of processes—negotiation, consensual processes involving a third-party neutral, and adjudicatory processes involving a third-party neutral. In general, the disputing parties must voluntarily consent to its use before engaging in an ADR process.

Two kinds of negotiation fall under the rubric of ADR processes—principled (interest-based) negotiation and partnering. Principled or interest-based negotiation encourages separating the people or personalities from the problem and focusing on the interests in the case as a means of facilitating creative, mutually beneficial solutions.^[2] Partnering is a method of dispute avoidance involving a retreat and conflict management training before entering into a contract. Consensual processes involving a third-party neutral include facilitation, conciliation, mediation, early neutral evaluation, minitrials, and summary jury trials. In these processes, a third-party neutral guides the disputing parties in their search for a solution but cannot impose a binding decision on them. The most common adjudicatory processes are fact-finding, nonbinding rights or interest arbitration, binding rights or interest arbitration, and med-arb. In these processes, a third-party neutral decides the outcome after hearing the disputants' cases.

ADR in public administration can also help increase public participation in policy decisions (and other decisions) of government.^[1,3] In negotiated rule-making, a committee from the public, along with a professional facilitator, convenes and tries to reach a consensus about changing an agency's rule.^[4,5] Negotiated rule-making in environmental dispute resolution has been particularly successful because the decisions are more responsive to participants' values and are more substantively robust, although the range of participants is not always representative of the greater public.^[6]

Public agencies and organizations increasingly employ ADR processes to resolve conflicts. Proponents of ADR claim that it is a good complement to decision-making in public administration, partly because ADR processes are varied and flexible enough to fit into a wide variety of organizational settings.^[1] Conflicts in public administration encompass a wide variety of disputants, and therefore, vary greatly. Examples of conflicts in public administration include intra-agency disputes (such as those between employees and management) or interagency disputes (disputes between two or more agencies). In addition, conflicts can be intergovernmental or can occur between a government agency and one or more private organizations, or between a government agency and the people or institutions it regulates.

ADR supporters claim that ADR offers potential benefits of time and cost savings and more satisfactory and stable outcomes by avoiding costly and time-consuming litigation. In response to these claims, Congress passed several pieces of legislation to encourage public sector use of ADR.

AUTHORITY FOR ADR USE IN PUBLIC ADMINISTRATION

Before 1990, federal public agencies had indirect and implicit authority to use ADR through the Administrative Procedure Act (APA), which permitted settlement as a means of addressing complaints. Many agencies also had a general enabling statute that indirectly allowed them to use a variety of dispute resolution processes in resolving contractual disputes.^[7] Binding arbitration was more controversial and was prohibited outside of labor disputes by the comptroller general. Data are not available regarding how often agencies took advantage of this indirect authority to use ADR to resolve disputes.

Since 1990, Congress has encouraged the use of ADR in the public sector by passing several laws that provide the federal sector with direct, specific statutory authority to address disputes using ADR methods. The Civil Justice Reform Act of 1990 [28 U.S.C. §§ 471–482 (1994)] required the federal courts to decrease the costs and delays of litigation, specifically recommending the use of voluntary ADR to achieve this end and authorizing judges to refer cases to ADR. Because the federal government is the biggest single consumer of federal court services, this change resulted in greater use of ADR for litigated cases to which federal agencies were parties. The US Department of Justice has collected some data from assistant US attorneys on this use of ADR and determined that it saved time and money.^[8] Subsequently, Congress enacted the ADR Act of 1998 to require that each federal court adopt a dispute

system design, that is, a plan for integrating ADR into the administration of their caseloads [28 U. S. C. § 651 (2005)].

However, this authority applied on its face only to the federal courts. The Administrative Dispute Resolution Act (ADRA) of 1990, made permanent in the ADRA of 1996 [5 USCS §§ 571 et seq. (2005)], granted broad authority to the Executive Branch to use various ADR techniques. Congress found that the excessive cost and time of litigation justified the statute; it praised the efficiency of ADR methods and adopted a policy preferring ADR to litigation. The ADRA defines ADR as “any procedure that is used in lieu of adjudication as defined in the APA, to resolve issues in controversy, including but not limited to settlement negotiations, conciliation, facilitation, mediation, fact finding, minitrials, and arbitration, or any combination thereof” [5 U. S. C. § 571 (2005)]. On its face, the ADRA applies to processes used in lieu of adjudication, a quasi-judicial procedure for making decisions that affect public policy. In practice, the ADRA has also provided authority for quasi-legislative processes for making policy using collaboration and consensus.^[9] The 1996 amendments exempted confidential communications between the disputants and the third-party neutral from disclosure under the Freedom of Information Act. They also removed disincentives to use binding arbitration; specifically, in the previous version of the act, federal agencies could reject awards issued in binding arbitration within 30 days after their issue. By eliminating the right to reject an award, the revised statute made binding arbitration equally enforceable and final for both parties.^[10]

The Negotiated Rulemaking Act (NRA) of 1990, made permanent in the NRA of 1996 [5 U.S.C. §§ 561, et seq. (2005)], encouraged federal agencies to use negotiation with interested stakeholders to develop regulations before engaging in traditional notice and comment rule-making.^[3] This statute expressly authorizes the use of consensus-based processes for a quasi-legislative agency action to make policy. There is an ongoing debate about the efficacy of negotiated rule-making and what criteria ought to be used to assess its success, for example, the time it takes to adopt rules, the frequency of litigation challenging a final rule, or the satisfaction of stakeholders with the result of the process.^[9]

Federal ADR policy also evolved through presidential mandates. President Clinton requested the formation of the Interagency ADR Working Group in 1998 to coordinate and facilitate federal agency use of ADR.^[10] Through the Interagency ADR Working Group, federal agencies participate in at least one of five sections: workplace, enforcement and regulatory, claims, contracts/procurement, and the small agency caucus. The Working Group provides federal agencies

across the federal government with information about program design, training, identification of neutrals, and evaluation systems. More information about the Interagency ADR Working Group is available on their website (www.adr.gov).

State and local governments derive authority to use ADR in three ways: as part of their power to enter into contracts; through provisions of the State Administrative Procedure Act, or through express statutory or executive order authority to use some form of ADR.^[7,11] At least 25 states have some form of express authority for state or state and local agencies to use ADR (more information is found on the website of the Policy Consensus Initiative, www.policyconsensus.org). State executive orders and statutes require state agencies to familiarize themselves with ADR methods and to use them when appropriate. Local governments also have direct ADR authority through their police powers, contracting powers, and budgetary and legislative authority. Some local governments use local community mediation programs to address conflict in the public schools, for cases of juvenile delinquency (called victim-offender mediation) or for neighborhood disputes that would otherwise consume judicial resources (the classic barking dog case).^[11] More information about ADR use by local governments is available from the National Association for Community Mediation (www.nafcm.org).

THE PROS AND CONS OF ADMINISTRATIVE AGENCY USE OF ADR

Proponents of ADR argue that it saves time and money, results in improved relationships between parties to a dispute, enhances the exercise of voice by citizens and stakeholders in matters of public policy, and can yield outcomes that are both win-win and more creative than traditional win-lose processes. Some claim that ADR helps public administrators to avoid policy stalemates by improving communication between an agency and the public. Through participation, the public perceives policy decisions as more legitimate. There is a developing body of research assessing the evidence for these claims by using evaluation criteria including time to final disposition of a dispute, parties' assessments of their transaction costs, participant satisfaction, the overall pattern of ADR outcomes, changes in parties' conflict management styles, and systemic impacts such as complaint filing at the organizational level. (A comprehensive review of the growing body of evidence that ADR proponents can use to justify their claims is outside the scope of this article, but for a review, see Refs.^[11,12].)

ADR is not appropriate for every conflict to which an agency is a party. There are both concerns about

when an agency should use ADR and critiques about its impact on the perception and reality of government legitimacy, transparency, and accountability in public law and decision-making. For example, some argue that ADR gives special interests undue influence over public policy. Institutional repeat players have advantages in both repeated contact with agency representatives and greater resources to bring to bear on the process of policy-making. They argue that the agencies become more receptive to this one voice in the policy process (agency capture). Public interest groups argue that they cannot afford to participate in lengthy negotiations regarding a proposed regulation. Others suggest that consensus-based processes coopt advocacy groups and make them less effective.

There are also concerns regarding the impoverishment of public law and precedent. Some have argued against settlement; the classic example is what if our country's leading desegregation case had been mediated in private rather than decided publicly by the US Supreme Court? Moreover, mediation and arbitration are designed to be private, confidential processes that do not create precedent. This leads to criticism that decision-making is ad hoc, not transparent, and not subject to adequate public or judicial review.

In response to these concerns, many agencies are developing guidelines for ADR use. They are instituting processes to solicit the voices of stakeholders about program design. Lastly, they are instituting systems for evaluating the efficacy and quality of their ADR programs to continue to improve how they serve the public and the public interest.

IMPLEMENTING ADR IN THE PUBLIC SECTOR

Five years after the ADRA of 1990, a review of federal agencies' implementation of ADR found that each agency had implemented ADRA to different levels.^[13] Some agencies were using multiple ADR methods on a regular basis, while other agencies restricted themselves to fewer techniques. Lopsided implementation of ADR in these agencies can be explained by difficulties in amassing human and financial resources for implementation as well as difficulties in creating institutional pressures for implementation.^[14] By far, mediation was the most widely used ADR method, followed by early neutral evaluation. The least used methods were ombudsman programs, fact finding, and facilitation.^[13] Use has continued to grow in the federal government in particular.

As experience with ADR in public administration increases, it is possible to identify some important lessons public administrators have learned about the best practices for implementing an ADR program.^[11] First, the organization should systematically assess

how it handles conflict in the present system. The agency must identify the stakeholders, current processes in use, resources available to support their use, their outcomes, and how stakeholders view them. Second, good system design requires the participation of people who will be using the system: stakeholder involvement. Third, ADR is only an option if the system is designed to focus on the interests of disputants, or their underlying needs and concerns, rather than the more traditional focus on legal rights and obligations. Fourth, a key to effective ADR is giving disputants control over the choice of process for resolving conflict, the authority to resolve their own disputes, and the training and skill-building necessary to do so effectively.

Fifth, recent thinking supports the creation of integrated conflict management systems, that is, an overarching framework for all the grievance or conflict management processes in an agency. Such a system should have multiple points of entry, multiple options for addressing the dispute (fit the forum to the fuss), with processes available at the lowest possible level within the agency or organization. It should organize options from low to high cost, with relatively inexpensive processes for determining rights if interest-based consensual processes fail.

Sixth, the best-designed system can fail if the agency fails to implement it comprehensively. Effective implementation includes training, publicity, informational materials, point people as resources, goal-setting, outreach, intake, and monitoring use. It also requires top-down and bottom-up support. Leaders at the top of the agency should provide adequate resources and public support, and should use the program. People in the trenches should provide testimonials from satisfied participants, success stories, and newsletters, and build word-of-mouth on the street. Seventh, good implementation requires a system for monitoring and improving the program. Continuous data collection can help provide objective accounts of a program's function and point to ways to improve it. Stakeholder feedback is essential to a program's success. It encourages ownership, and if feedback is positive, this encourages others to use the program. It also identifies strengths and weaknesses. Objective data about program outcomes can help create internal support for budget and human resources.

Although ADR use in the public sector has increased dramatically, obstacles prevent full institutionalization. Key among these is the difficulty in demonstrating ADR success.^[13,15] More systematic data on program performance may enable an agency to overcome obstacles such as organizational inertia, scarce agency resources, unfamiliarity with the techniques, participants' fear of appearing weak if they use ADR, and skepticism about the ability of ADR to work.^[15]

CONCLUSIONS

The policy toward a greater range of dispute resolution processes encourages ADR over traditional adversarial and adjudicative process. Congress passed several laws in the 1990s authorizing and encouraging the use of ADR in the public sector as an attempt to reduce the costs and delay associated with litigation and to improve the outcomes. Public agency use of ADR is becoming increasingly common. Alternate dispute resolution methods include such techniques as mediation, arbitration, minitrials, facilitation, and other hybrid methods. ADR processes can be applied to intra-agency disputes, interagency disputes, intergovernmental disputes, and disputes between an agency and private organizations or citizens. Principles of good dispute system design are emerging. Public administrators need to acquire a deeper knowledge of the wide range of dispute resolution processes available as part of their substantive responsibilities and bring this knowledge to bear on improving administrative agency conflict management systems.

ARTICLE OF FURTHER INTEREST

Alternative Dispute Resolution Processes, p. 80.

REFERENCES

1. Bingham, L.B.; Nabatchi, T.; O'Leary, R. The new governance: practices and processes for stakeholder and citizen participation in the work of government. *Public Admin. Rev.* **2005**, *65* (5), 547–558.
2. Fisher, R.; Ury, W.; Patton, B. *Getting to Yes*; Penguin Books: New York, 1991.
3. Manning, N.J. ADR and administrative responsiveness: challenges for public administrators. *Public Admin. Rev.* **1994**, *54* (2), 197–202.
4. Kerwin, C.M. *Rulemaking: How Government Agencies Write Law and Make Policy*, 2nd Ed.; CQ Press: Washington, DC, 1999.
5. Harter, P.J.; Pou, C., Jr. Using negotiated rulemaking effectively. In *Federal Administrative Dispute Resolution Deskbook*; Breger, M.J., Schatz, G.S., Laufer, D.S., Eds.; American Bar Association: Chicago, 2001; 111–134.
6. Beierle, T.C.; Cayford, J. Dispute resolution as a method of public participation. In *The Promise and Performance of Environmental Conflict Resolution*; O'Leary, R., Bingham, L.B., Eds.; Resources for the Future: Washington, DC, 2003; 53–68.
7. Bingham, L.B. Alternative dispute resolution in public administration. In *The Handbook of Public Law and Administration*; Cooper, P.J., Newland, C.A., Eds.; Jossey-Bass: San Francisco, 1997; 546–566.

8. Senger, J. *Federal Dispute Resolution: Using ADR with the United States Government*; Jossey-Bass: USA, 2003.
9. O'Leary, R., Bingham, L.B., Eds.; *The Promise and Performance of Environmental Conflict Resolution*; Resources for the Future Press: USA, 2003.
10. Senger, J. Turning the ship of state. *J. Dis. Resol.* **2000**, 79 (1), 79–95.
11. Bingham, L.B. *Mediation at Work: Transforming Workplace Conflict at the United States Postal Service*, Human Capital Management Series; IBM Center for The Business of Government, 2003; 1–46.
12. Jones, T.S. Conflict resolution in the field: special symposium. *Conflict Resol. Quart.* **2004**, 22 (1–2), 1–320.
13. Bingham, L.B.; Wise, C.R. The administrative dispute resolution act of 1990: how do we evaluate its success? *J. Public Admin. Res. Theory* **1996**, 6 (3), 383–414.
14. Nabatchi, T. The institutionalization of alternative dispute resolution in the federal government. *Public Admin. Rev. in press*.
15. Susskind, L.E.; Babbitt, E.F.; Segal, P.N. When ADR becomes the law: a review of federal practice. *Negotiation J.* **1993**, 9 (1), 59–75.

Default

Robert J. Eger
Andrew Haas

*Public Administration and Urban Studies Department, Georgia State University,
Atlanta, Georgia, U.S.A.*

INTRODUCTION

This chapter provides a broad overview of bond default. It includes the definition of a bond default, an explanation of default vs. technical default, bond recovery, private municipal guarantees, and a history of bond defaults.

WHAT IS A BOND DEFAULT

Default risk is the chance that the issuer of a bond may be unable to make timely principal or interest payments. During a default, bondholders generally do not lose all of their principal value invested in the bond. In most cases, a default results in the deferral of the coupon payment. Bonds in default can be purchased on the secondary market at a discounted rate. These bonds carry significantly more risk but pay a higher return if the issuer successfully emerges from bankruptcy. If payment is not restarted or protected by bond insurance, investors are sometimes able to recoup their losses through the legal system, which can involve all parties in the default issue from the issuer of the bond all the way down to the underwriters.

DEFAULT VS. TECHNICAL DEFAULT

There is a notable difference between default and technical default. Default is a failure to pay in a timely manner, according to the covenants of the bond, principal and/or interest when due. This is often the result of a bankruptcy. A technical default is a failure by the bond issuer to meet the conditions of a bond agreement. Examples of technical default are failure to maintain required fund reserves or other capitalization requirements. Losses to the bondholder are not necessarily a result of a technical default. A technical default may be reversed by changes in policy or reparations made by the bond issuer. It may serve as a warning that a default is coming but if problems are resolved in time, default can be avoided. A technical default will often lower the price of a bond in the secondary market.

BOND RECOVERY

Bonds can recover from default in several ways. The issuer may make full debt service payments through other financing or liquidate collateral used to secure the bonds. Municipal service providers often repay accumulated debt when cash flows resume normal function. Physical assets that are public property and are never pledged to bondholders back this type of debt. Bondholders have an automatic lien on revenues until the entity is able to recover normal operations. The two sectors with the highest default rates, industrial development bonds and multifamily housing bonds are often backed by collateral leading to higher than average recovery rates.

PRIVATE MUNICIPAL BOND GUARANTEES

There are two main private sector options to improve a bond rating: bond insurance and standby letters of credit. Both affect the net cost of borrowing.

Bonds insurance, written by commercial insurance companies, can improve a bond issue rating. This is in addition to the pledge of security of the issuer's revenues. The insurance provides payment to the bondholders for the life of the issue if a default should occur. Municipal bonds that would qualify as borderline investment grade will often seek insurance to increase the issue's credit rating and reduce the cost of borrowing. Municipal bonds that are rated AAA because of their insurance protection will often yield more than an issue rated triple-A without insurance protection. Most insurance of bond issues take place at the initial issue while some take place in the aftermarket. In the aftermarket, insurance is purchased for only a portion of the bond issue. Bond insurance firms and banks providing private guarantees must maintain a credit rating of AAA.

Another form of guarantee is a standby letter of credit from a commercial bank. The letter is a pledge of the bank's credit to make principal and interest payments on the bonds if the issuer is in default. It is provided for a fee to improve the rating on the bond issue and drive down the cost of capital.

JUNK BONDS

“Junk bonds” are bonds that are rated below investment grade and carry the highest risk of default. The term “junk bond” comes from the low credit rating that drives the higher yields required to attract investors. The highest quality bonds, rated AAA, carry a low yield but also a low risk of default. The lowest quality bonds, rated BB through D, carry high yields but also high risk of default. Studies have demonstrated that portfolios of high yield bonds have higher returns than other bond portfolios, suggesting that the higher yields paid by junk bonds more than compensate for their additional default risk.

HISTORY OF MUNICIPAL BOND DEFAULTS

The history of municipal bond defaults in the United States closely parallels cycles of economic growth and westward expansion. During each major depression, most defaults occurred in the western regions. A period of heightened municipal borrowing and economic development preceded each depressionary period. Fluctuating land values, financial mismanagement, and unrealistic fiscal projections all have contributed significantly to bond defaults. The post-World War II period has seen revenue bonds as a new source of default mostly due to inaccurate revenue projections.

The default of \$2.25 billion by an association of local governments in Washington State, known as the Washington Public Power Supply System (WPPSS), is perhaps the most famous bond default in history. The associations’ nuclear-plant construction program collapsed in the early 1980s under the pressure of rising costs, high interest rates, and overestimated demand for electricity. Two of the five nuclear power plants to be built were financed mainly by tax-free municipal bonds. These bonds carried a higher interest rate than the federally guaranteed financing of the other three plants. The premium interest rate made the bonds attractive to investors but also carried a higher risk of default. Only one of the five nuclear-plant projects ever began operation. As the single largest single issuer of municipal bonds in U.S. history, the default sent shock waves through the bond market. The impact was so severe that new disclosure requirements were implemented by the U.S. Securities and Exchange Commission (SEC) to help prevent future defaults.

Another large default occurred in Orange County California in December 1994. Orange County Treasurer, Robert Rron, faced with strained financial conditions needed to enhance investment returns and leveraged the county’s investment fund using short-term

notes, repurchase agreements, and derivatives. The need was driven by the county’s unwillingness to raise taxes and other fiscal constraints. On the outside, the \$7 billion debt portfolio appeared safe as it contained U.S. government bonds with relatively short-term maturities. Upon closer inspection, it was revealed that these debt instruments were highly leveraged by repurchase agreements by private firms. Citron borrowed more than \$12 billion at low interest rates, using the pool’s securities as collateral, and invested the money in high-yield securities. This investment tactic proved highly profitable and earned above market returns. However, late in the summer of 1994, the investment scheme started falling apart as the Federal Reserve increased discount rates. Soon after the debt portfolio started showing sizeable paper losses and the private firms began to demand more collateral for their repurchase agreements, Orange County was unable to meet these demands. The demands to absorb a \$1.5 billion loss forced Orange County to file for bankruptcy protection.

CONCLUSIONS

Municipal bonds are generally considered safer than corporate bonds, since a municipality is less likely to go bankrupt than a corporation. However, there have been large defaults by municipalities that have forced significant losses on investors. Protection against default is available for a fee on some municipal bonds that are insured by the private sector and is intended to improve the rating of a bond issue and reduce borrowing costs. A technical default on a bond does not necessarily mean a financial loss to an investor as it can be remedied through fiscal changes. However, it can be considered a warning of a more serious problem with the finances of a bond issue.

BIBLIOGRAPHY

1. Cohen, N.R. Municipal default patterns: an historical study. *Public Budgeting Finance* **1989**, 9 (4), 55–66.
2. Denison, D.V. Did bond fund investors anticipate the financial crisis of Orange County? *Munic. Finance J.* **2000**, 21 (3), 24–39.
3. Ewing, B.T. Economic forces and the municipal bond default risk premium. *Munic. Finance J.* **2002**, 23 (1), 17.
4. Fabozzi, F.J.; Fabozzi, T.D.; Feldstein, S.G. *Municipal Bond Portfolio Management*; Irwin Professional Publishing: Burr Ridge, 1995; 179–189.
5. Fairchild, L.M. Are federal disclosure requirements beneficial for the municipal bond market? *Munic. Finance J.* **2000**, 21 (3), 65–82.

6. Glossary of Municipal Bond Terms, <http://www.munidirect.com/glossary.php>.
7. Halstead, J.M.; Hegde, S.; Linda, S.K. Orange County bankruptcy: financial contagion in the municipal bond and bank equity markets. *Financial Rev.* **2004**, *39* (2), 293.
8. Jones, L.R. The WPPSS default: trouble in the municipal bond market. *Public Budgeting Finance* **1984**, *4* (4), 60–77.
9. Justice, J.B.; Simon, S. Municipal bond insurance: trends and prospects. *Public Budgeting Finance* **2002**, *22* (4), 114–137.
10. Kilger, D.; Sarig, O. The information value of bond ratings. *J. Finance* **2000**, *55* (6), 2879–2902.
11. Lamb, R.; Rappaport, S.P. *Municipal Bonds*, 2nd Ed.; McGraw-Hill Book Company: New York, 1987; 36–50.
12. Litvack, D.T.; McDermott, M. Special Report: Municipal Default Risk Revisited. *Fitch Ratings* **2003**.
13. Nanda, V.; Singh, R. Bond insurance: what is special about munis? *J. Finance* **2004**, *59* (5), 2253–2279.
14. Quigley, J.M.; Rubinfeld, D.L. Private guarantees for municipal bonds: evidence from the aftermarket. *Natl. Tax J.* **1991**, *44* (4), 29–39.

Democracy and Public Policy

Dale Krane

Department of Public Administration, University of Nebraska at Omaha,
Omaha, Nebraska, U.S.A.

Gary S. Marshall

Department of Social Sciences, Roskilde University, Roskilde, Denmark

INTRODUCTION

Any discussion of the definition, the attributes, or the purposes of democracy invariably entails a consideration of public policy. The two terms, although distinct and different, are closely intertwined. Democracy is a form of government and, as such, refers to a system of authority and power. Discussions of democratic theory revolve around the organization and use of political power within a society—who should govern, how they should govern, and for what ends or purposes they should govern. By contrast, public policy refers to a purposive course of action established by public officials that is binding on the residents of a community or nation. Simply put, public policy is what governments choose to do or choose not to do.^[1] Who exercises power or has authority to take actions binding on a community or a society will obviously affect what actions are selected or not selected. Likewise, the organization of authority will also shape its use as manifested in policy choice. Because of this link between form of government and performance of government, the theory of democracy, in Henry Mayo's^[2] words, "... is one answer to the question of how the political policy decisions are made and should be made."

DEFINITIONS OF DEMOCRACY

Since its emergence in Greek political thought, the literal definition of democracy, "rule of the people," has remained more or less constant. But how this simple two-element formula of *demos* (people) and *kratos* (power) is interpreted and translated into actual practice has been a major industry in political theory circles from Socrates and Solon to Rousseau and Mills to the multitude of today's theorists. Books addressing questions such as "Who are the people?," "Which people should exercise power?," and "How many people are required to make a decision binding on the community?" fill whole sections of libraries around the world. "Today the term *democracy* [italics in original]," Dahl observes,^[3] "is like an ancient kitchen midden packed

with assorted leftovers from twenty-five hundred years of nearly continuous usage." Yet, these definitional debates can be distilled down to the basic issue of identifying "... the best constitutional means of approaching the ideal, it already agreed that this ideal includes or involves a large participation of the common people in the forming of public policy"^[2] (as quoted by Mayo from Robinson, Richard: *Definition*. Oxford 1949, 166).

Much of the theorizing about democracy falls into one of two perspectives. For much of its history, democracy was conceived as a form of government applicable only to communities of relatively modest populations such as the Greek city-states or the Italian and Swiss cities of the Renaissance period. Within small-scale societies, the people, usually defined as the (male) citizens, formed the government, typically an assembly of the whole, and decisions of the assembly were binding on citizens as well as all other residents of the community. This theoretical position is often labeled as *direct democracy* and "... is principally concerned with ensuring democratic rights for the *community* as a whole." [italics in original].^[4]

Critical to direct democracy is the notion that citizens possess the capacity and the volition to govern themselves. Rule of the people requires "rule by the people," or self-government. Direct democracy rests on two core principles: 1) citizens are "sovereign"; that is, citizens make public policy and 2) each citizen is legally and politically equal to every other citizen. Also necessary to direct democracy are two important rules: 1) when unanimity does not exist among citizens, the preference of the largest number of citizens becomes public policy, and 2) freedom to express one's opinion about public policy is protected, and the majority may not "silence" the minority (though the minority must obey policy decisions until the decision is changed). Because policy decisions reflect the choice of the community, policy must be obeyed by all citizens, officials as well as non-officials. Furthermore, the processes of deciding and then administering public policies are legitimate only if the established procedures have been followed. That is to say, political power is limited by a set of rules as to how it will be exercised. Instead of

government by privilege or by force, government acts through popular consent.^[5] Direct democracy, it should be noted, is not immune to the problems commonly associated with the exercise of power such as the difficulties of 1) arriving at a community-wide consensus, 2) controlling conflict among citizens with different preferences, and 3) ensuring compliance with the collective decision.

As modern nation-states developed between the 1600s and the 1800s, rapid population growth and urbanization posed a significant challenge to democratic theorists. The sense that direct, communal democracy in a large-scale city or country was untenable became increasingly widespread. If it was impossible for each person to participate in the deliberation and choice of public policy in a large city or country, then democracy had to be redefined. It was crucial to find ways for citizens to control the government as well as to be protected from actions by the government that would lessen or eliminate “popular sovereignty.” Liberal ideas of representation were proposed as a solution, and rule by the people was redefined as the choice of one’s rulers.

The ideas motivating the revolutionary break from the medieval order strongly influenced the second perspective to democracy. Instead of democracy embodied in a community-wide consensus, “general will,” or “public interest,” democracy became associated with the revolutionary ideal that each person, to quote Thomas Jefferson, has “certain unalienable Rights.” If the people were to rule, then each person had to be guaranteed the exercise of certain rights such as belief, speech, assembly, and representation. Likewise, certain limits had to be established beyond which the community could not impose its will on the individual (NB: limits also had to be imposed on individual behavior). If every person was inherently equal, then a mechanism that allowed each person to express a preference for who would rule had to be developed. Furthermore, because most persons could not devote all of their energy and time to the occupation of ruler, then those few persons who would rule had to be held accountable to the ruled.

Elections became the set of procedures held to be most important to the operation and maintenance of a democracy. If the question is when do people exercise popular sovereignty, the answer, according to Sartori,^[6] “is easy—during elections.” Citizens would select rulers from candidates who sought votes by offering different visions of public policy, and the candidate that attracted the most support (votes) became the ruler, typically in the form of a representative to an assembly of elected officials who had the authority to make public policy.

Once elections became the mechanism through which the people ruled, then the procedural rules for

the conduct of elections became a matter of high interest for citizens, candidates, and officials. Important procedural questions included the following: who may vote, who may be a candidate, how are votes counted, may candidates form groups of like-minded candidates (i.e., political parties), how many representatives will serve in the legislative assembly, how will representatives be selected by voters (e.g., by geographic territory or by occupation), how long a term will each representative serve, and may a representative become a candidate in the next election? Because many different answers to these questions can be devised, different nations have developed different electoral procedures. Consequently, indirect or representative democracy is characterized by a variety of electoral procedures.

These two distinct perspectives on the definition of a democracy—communal vs. individual—bear directly on how one arrives at a policy decision as well as what constitutes legitimate public policy. Direct democracy holds that public policy emanates from decisions made by the whole community, or, to paraphrase Rousseau, sovereignty cannot be represented. Thus, only a collective body composed of all or as many members of the community as practical can truly determine the “... common identity, its life and its will.”^[7] The test of public policy in a direct democracy is whether it “embodies a moral imperative for people to promote common interests.”^[7] Indirect democracy, by contrast, holds that the realities of daily life make it impossible to involve all or most citizens in the continual process of policy making. Put another way, if every adult citizen devoted significant amounts of time to making public policy, there would be no one left to perform all the other tasks necessary to a functioning society. Therefore, some citizens must be selected to form a government and make public policy. Policy decisions are legitimate if they are made by the elected representatives of the people. The key to indirect democracy is the creation of one or more mechanisms by which the people exercise control over the representatives and the policy choices they make. Because each representative is presumed to express the views of the individuals who elected the representative, it is also presumed that a wide range of opinions will exist among the representatives. Consequently, policy will not reflect the “common will” or the “public interest”; instead, public policy will be, as Madison argued, the product of bargaining and negotiation among the individual representatives.

ATTRIBUTES OF DEMOCRACY

Democracy has always been a contested concept, not just in terms of what democracy means in the abstract,

but also what constitutes an actual, functioning democracy. What are the identifiable attributes or features that make possible (or better, probable) “rule by the people?” On what bases or criteria can we decide that one nation (or community) is democratic, while another is not? Democracy as an ideal ultimately has to be translated and transformed into governing institutions that resolve societal problems and produce policy decisions that reflect the consent of the governed.

Citizen influence over policy makers and policy decisions occurs in both democratic and non-democratic regimes. Benevolent dictators populate the pages of history, and even tyrants understand that it is often too costly to ignore popular opinion in every policy decision. If not popular influence, then what? The most typical single answer given by scholars is popular control of policy makers and policy decisions. But even this change of one word (control for influence) does not clearly distinguish democratic government from non-democratic ones. Obviously then, multiple mechanisms are necessary to ensure “rule by the people,” and so scholars of democracy have sought to construct lists of attributes by which they identify a government as democratic.

Although several scholars have offered particular lists, the most widely known and cited list of political institutions that characterize modern democratic government is the one proposed by Robert Dahl. Since his early masterpiece *A Preface To Democratic Theory* in 1956, Dahl has sought to identify the distinguishing marks of democratic government. His current version^[8] includes 1) elected officials, 2) free, fair, and frequent elections, 3) freedom of expression, 4) access to alternative sources of information, 5) associational autonomy, and 6) inclusive citizenship.

Of the six that Dahl identifies, other scholars typically concur with his first four or five features of democratic government: elected representatives, non-coercive elections, basic political freedoms, associational autonomy, and access to independent information. In a very real sense, the marks of a modern democracy were first specified in the U.S. Constitution’s initial 10 amendments, the “Bill of Rights.”

It should be noticed that political equality is not specifically included in Dahl’s list; rather, equality is implicit in the notion that (practically) all citizens have rights to vote, to run for public office, and to exercise a broad range of political freedoms. Also missing from Dahl’s list is the requirement for majority rule. It is omitted because majority rule is simply one of several possible decision rules by which an assembly of citizens or representatives may arrive at a decision. Representative bodies in democratic nations function with different decision rules, including plurality, simple majority, and various forms of extraordinary majorities (more than 50% + 1).

Dahl’s list of fundamental requirements must be present in any political system that is to be considered democratic, but they may be manifested in very different political institutions and policy processes. Comparative studies of democratic politics, to simplify a large body of research, recognize at least three distinct models of democratic political institutions, each of which exhibits a different style of policy making. Variations of each of the three models exist, but the basic institutional differences among the three models are critical to the particular styles of policy making and the way in which popular sovereignty is exercised.

Parliamentary institutions combined with a majoritarian political party system constitute the first model, sometimes referred to as the “Westminster” model, after the Palace of Westminster where the British Parliament meets. The key features (although not all) of Westminster majoritarian democracy are 1) fusion of executive and legislative power, 2) executive power concentrated in a ministerial cabinet, 3) asymmetric bicameralism, 4) exclusively representative government, 5) unitary, centralized government, and 6) two-party system. Majoritarian, parliamentary institutions provide for “rule by the people” through an emphasis on a responsible, strong political party model in that two parties contend for popular support by offering competing policy visions. The winning party gains essentially exclusive control over the instruments of government action and is expected to enact its policy platform. The leader of the majority party in the parliament is also the prime minister who selects the cabinet, thus there is unified control across legislative and executive institutions. Only a vote of no confidence or a loss at the next scheduled election can seriously undermine the party in power’s control over public policy. The United Kingdom is the preeminent example of this model, and many of its former colonies exhibit this form of democratic government.^[9]

Presidential government is a second model of indirect democratic government, and is characterized by a formal separation of powers between executive and legislative institutions. Instead of the fusion of legislative and executive powers that typify parliamentary government, the executive “head of state” in presidentialist governments is selected independently from the legislative branch and cannot be removed by the legislature, except for very serious reasons and through complex procedures. Presidents may propose new policy directions but cannot enact them into law because the legislature is independent of the executive. Instead, the president must build a coalition of legislators (of the president’s party or other parties) that is willing to support the executive’s policy preferences. Presidential systems may be unitary or federal, may have two or more political parties, and may have a bicameral

legislature, especially if the country is large in territory. Presidential governments provide for popular sovereignty through the following three key features: 1) the independent election of the executive from the legislative members of the government, thus reducing the concentration of power in the hands of one office or institution, 2) policy making requires bargaining within and between each legislative chamber as well as between the legislature and the executive, thus ensuring that a multiplicity of views are represented, and 3) because the executive is typically the only public official elected by the whole electorate, the president's policy positions are considered to be those of the electorate (or at least of those who voted for the incumbent).^[10] The United States of America is the preeminent example of presidential government, and it can also be found in several Latin American nations.

A third model of representative democratic government combines the parliamentary form with the presidential to create what is labeled as a "dual power," or "semipresidential" government. The key institutional features include the following: 1) an independently elected head of state [the president], 2) a legislatively selected head of government [the prime minister], 3) an executive cabinet presided over by the prime minister, and 4) a legislature elected independently from the head of state. These nations usually have unitary government, multiple political parties, and may have a unicameral or a bicameral legislature. The reputed benefits for popular sovereignty of a dual-power government: 1) the stability associated with an independent president, 2) the flexibility of a parliamentary majority, and 3) the ability to avoid potential stalemates between the president and the legislature. France is the preeminent example of a "dual-power" democracy, and other examples can be found in Portugal, Finland, the Czech Republic, Poland, Estonia, Lithuania, and Slovenia.^[10]

Each of these three models of democratic government vary by institutional features, style of leadership, nature of policy bargaining, and constitutional rules. The organization of these different types of democratic government affects differentially the ability of citizens to influence policy making. For example, in parliamentary governments, much of the bargaining and debate over policy choices occurs as part of the electoral competition, and once a party wins a parliamentary majority, it can enact its policy platform without any serious obstacles. By contrast, in presidential models, bargaining and debate over policy continue past the election of the president and the legislature, and as a consequence, it is often the case that each policy initiative requires an extensive effort to build a bipartisan majority to support the proposal. The point here is simply that the ideal of "rule by the people" must be transformed from ideas to functioning institutions

and rules by which governing institutions produce policy results that are acceptable and legitimate.

PURPOSES OF DEMOCRACY

Governments of whatever form exhibit some common purposes, among which are social order, dispute resolution, coordination of collective action, and protection from external attack. But many different forms of government have been established with the intent of achieving certain objectives. For example, a Communist government seeks to ensure a dictatorship of the proletariat and to alter the means of production to create a socialist utopia.^[11] Fascism sought to develop and maintain "the glory of the State" so that all other activity in society was subordinate to the State.^[11] But what objectives are to be achieved by establishing a democratic government? Although there is no agreement as to what purposes democracy is designed to attain, there are at least five distinct answers to the question, and each of these answers contains an expectation about the goals of public policy in a democracy.^[2]

The first answer about the purposes of a democratic government is implicit in the core notion of rule of the people. To ensure popular sovereignty, democratic government is designed to constrain the emergence of an elite or a permanent ruling class and to encourage widespread civic participation. The rules of the democratic political game, as expressed in law, create opportunities and resources so that citizens may participate in policy decisions. To put this another way, democratic government is rule by a continually changing cast of amateurs. Furthermore, there are no preordained goals for public action; instead, public policy will be the product of the continually shifting diversity of interests within the community. "All policies will be compromises, and it is unlikely that we shall find any democracy committed to one all-consuming purpose."^[2]

Second and probably the most widely accepted purpose of democracy is the protection of individual rights. Liberal democrats, in their revolutionary attacks on the absolutist State, defined freedom as independence from government control. Jefferson's (and Thomas Paine's) dictum that "the best government is the one that governs least" concisely captures this attitude, but it is important to understand that the context was one where the State controlled most spheres of human activity—commerce and trade, religion and culture, property and status. Locke, who was Jefferson's inspiration, argued that every power government exercised came at the expense of individual liberty and, therefore, the less public policy the

better. Rousseau, on the other hand, argued that individual rights could be increased by government action, e.g., where public policy restricts the rights of employers in order to eliminate unacceptable practices such as gender and racial biases in hiring or child labor.^[12] This debate over negative vs. positive conceptions of freedom does not detract from the basic point that an important purpose of democratic governments is to constrain government as well as individual action to ensure rule by free consent of the governed.

The struggles that produced modern democratic nations were motivated not only by efforts to freely exercise religion and to freely choose one's rulers, but also to protect one's property from confiscation by the State. Much of the justification for limited government rests on the protection of personal property rights, and, by extension, it is often held that an important objective of democratic government is the maintenance of a free or market economy. The fact that one can find market economies in non-democratic nations undermines part of this idea that only under democracy can capitalism flourish. However, because all democracies support capitalist markets, there is obviously a connection between these two distinct societal institutions.^[3,13,14] The link is through public policy, which creates and enforces the rules required to sustain these two institutions. Both democracy and capitalism depend on the freedom of individual choice. Just as democracy depends on a set of procedural rules that provide for competition among candidates, so also does a liberal market depend on rules that provide for competition among producers. Critical to the functioning of market economies is an extensive set of laws and regulations ensuring free choice for producers and consumers, employers and workers. The contemporary efforts to foster economic development in the nations of the former Soviet Union illustrate the necessity of an extensive body of public policy that establishes and maintains the institutions of a modern capitalist economy (e.g., property ownership, enforceable contracts).

A fourth purpose, it is argued, that democracy serves is the development of the individual. The pursuit of equality has been part of the pursuit of liberty because freedom for only some individuals leaves others unfree and unequal. The Christian ideal of the inherent equality and worth of all individuals predates the development of modern democracy, but this idea strongly shaped the earliest arguments for democracy.^[5] Without equality for all persons, the notion of government by consent of the governed is hollow, and so the campaigns to end government by hereditary rulers promised equality as well as liberty. Equality first meant equal franchise—the right to vote—but usually only for males with certain attributes—education, property, and/or race. It is only within

recent history that universal suffrage has become widely accepted.^[15]

Closely associated with equal franchise is the importance of the vote as a means for citizens to communicate their policy preferences to candidates and elected officials. If a whole class of persons is denied the right to vote (e.g., women), then matters of concern to this class of persons is likely to be ignored by public officials. Similarly, if one's vote counts less than someone else's vote (the problem of malapportionment), then one's policy interests will be disadvantaged. Furthermore, if one is not provided with the means to participate in the electoral process (e.g., education, information, finances), then one's ability to participate is diminished. Over time many democratic governments have enacted policies to foster equal participation and one can see the results in the increased diversity of today's elected (and administrative) officials, compared to previous eras.

More recent views on equality have focused on equality of opportunity—the provision of sufficient resources to individuals to permit them to pursue and “fulfill” their dreams. “The notion of democracy has always contained the notion of equality. Not arithmetical equality of income or wealth, but equality of opportunity to realize one's human capacities.”^[16] Predemocratic societies where most persons were serfs or slaves used force or tradition to prevent individuals from realizing their potential. Democratic governments were the first to proclaim the establishment of justice and the promotion of the general welfare as their principal purposes. This has led over time to a policy cornucopia of goods and services that today is labeled as the modern welfare state. From education to employment, from health to social security, the bulk of public policy in a modern democratic state aims to ensure a minimal quality of life for all and to facilitate, in the words of modern psychology, each person's “self-actualization.”^[17]

Perhaps the most unique purpose allegedly served by democracy is that it makes possible a new type of human character. “Democracy, then, both presupposes and tends to promote a particular type of character or personality; or alternatively we may say—since character is a slippery concept—that the system relies on certain attitudes or dispositions or behavior patterns and these it tends to foster because they contribute to the working of the system.”^[2] This argument goes back at least as far the writings of J. S. Mills and de Tocqueville, and has been revisited by such writers as James Bryce and Harold Lasswell. But how is it that a form of government can shape character or personality? An important answer is found in the recent writings that advocate “participatory democracy” as a remedy to the ills of liberal representative democracy. Critics of representative democracy such

as Benjamin Barber see the reliance on elections, interest groups, and political parties as producing a “thin democracy” in which “citizenship is only legal matter; people are bound together by self-interested contracts; and they are politically passive” (as quoted in Ref.^[7]). Departure from “possessive individualist ways of thought and action” (Macpherson’s phrase to describe consumerism and self-centeredness), Carole Pateman claims, “is facilitated by a change in people’s values that results from political participation itself” (as quoted in Ref.^[7]). Democratic citizens are, in a sense, forced to be free; that is, in order to solve problems in the community they must act together collectively—there is no ruler ordering them to act. In a society where all are free to hold and express different views, a democrat not only must tolerate these differences of opinion, but also must strive to find compromises to which a majority can consent. Such complex attitudes and behaviors, it is argued, depend on the development of a public-regarding spirit, and nothing does more to foster this new character than participation in public affairs; that is, action can alter attitudes.

That there are multiple purposes associated with the advocacy of democracy comes as no surprise. Democracy became the wave of the future in the 1600s because the purposes it alleged to serve “fit the minds of men,” in Burke’s phrase. Freedom and liberty, equality and individuality, were exciting ideas that pointed away from tradition and toward a new society. Democracy’s purposes, open and indeterminate, do not constitute the stuff of a dogmatic ideology,^[2] but instead offer a design within which individuals can shape public policy to fit the general interest of their community. Sen^[15] tells us that “a country does not have to be deemed fit *for* democracy; rather it has to become fit *through* democracy” [italics in original]. Democracy, as rule of the people, allows the people to enact public policies designed to pursue purposes that are beneficial to the community. This is what makes democracy, according to Sen,^[15] “a universally relevant system.”

CONCLUSION

The establishment of a democratic form of government is a fundamental public policy decision and makes popular sovereignty the primary principle of the policy-making process. Democracy depends on a set of necessary attributes, but how these critical features are built into governing institutions may vary from place to place. Parliaments and presidents are merely instruments through which popular sovereignty may be achieved; what is critical to democracy is a sufficient level of citizen participation so that the policies

selected reflect the diverse preferences and values within a community or country. Abraham Lincoln proclaimed the core ideas of democracy in his famous aphorism, “government of the people, by the people, and for the people.” The continuing challenge to theorists of democracy as well as to citizens and public officials is the adaptation of the institutional features of government to changing societal conditions without sacrificing the core ideas of popular sovereignty. Just as direct democracy was modified to accommodate population growth, so too must representative democracy be modified to ensure popular sovereignty in a world characterized by deep economic and social inequalities. Although universal suffrage is now widely accepted in most parts of the globe, public policy to provide a minimal quality of life to all citizens has not been attained. Long ago, Jefferson pointed out that certain public policies such as universal education were necessary to a functioning democratic government. The continued impoverishment of a majority of the world’s population stands as a major obstacle to the spread of democracy. But the continued existence of non-democratic forms of government in too many places also blocks action to reduce severe economic and social inequalities. Sen^[15] explains the pragmatic connection between the democratic procedures of government and the substance of public policy:

Political and civil rights give people the opportunity to draw attention forcefully to the general needs and to demand appropriate public action. The response of a government to the acute suffering of its people often depends on the pressure that is put on it. The exercise of political rights (such as voting, criticizing, protesting, and the like) can make a real difference to the political incentives that operate on a government.

Democracy offers citizens a set of instrumental procedures and rules that allow citizens to shape public policy so that “rule of, by, and for the people” is possible. Other forms of government are neither premised on this goal nor are designed to foster it.

REFERENCES

1. Dye, T. *Policy Analysis: What Governments Do, Why They Do it, and What Difference it Makes*; University of Alabama Press, 1976.
2. Mayo, H. *Introduction to Democratic Theory*; Oxford University Press, 1960; pp. 28, 31, 244–278.
3. Dahl, R. *Dilemmas of Pluralist Democracy*; Yale University Press, 1982; pp. 5, 108–115.
4. Grugel, J. *Democratization: A Critical Introduction*; Palgrave, 2002; 14 pp.
5. Muller, H. *Freedom in the Ancient World*; Bantam Books, 1961; 176–179.

6. Sartori, G. *Democratic Theory. Based on the Author's Translation of Democrazia E Definizione*, 2nd Ed.; Praeger: New York, 1965; 73 pp.
7. Cunningham, F. *Theories of Democracy: A Critical Introduction*; Routledge, 2002; pp. 124, 130.
8. Dahl, R. *A Preface to Democratic Theory*; Yale University Press, 1998.
9. Lijphart, A. *Democracies; Patterns of Majoritarian and Consensus Government in Twenty-One Countries*; Yale University Press, 1984; 1–20.
10. Keman, H.; Mallouk, M. Democratic institutions, governance and political performance. In *Comparative Democratic Politics: A Guide to Contemporary Theory and Research*; Keman, H., Ed.; Sage Publications, 2002; 257–285.
11. Morstein Marx, F. *Foreign Governments: The Dynamics of Politics Abroad*; Prentice-Hall, 1952; 417–418.
12. Muller, H. *Freedom in the Western World: From the Dark Ages to the Rise of Democracy*; Harper Colophon Books, 1963; 345 pp.
13. Lindblom, C. *Politics and Markets: The World's Political Economic Systems*; Basic Books, 1977; 161–169.
14. Macpherson, C. *The Real World of Democracy: The Massey Lectures*; Oxford University Press, 1966; 4 pp.
15. Sen, A. Democracy as a universal value. *J. Democr.* July **1999**, 10 (3), 3–17.
16. Macpherson, C. *The Real World of Democracy: The Massey Lectures*; Oxford University Press, 1966; 47 pp.
17. Bay, C. *The Structure of Freedom*; Stanford University Press, 1965.

Deontology

James R. Heichelbech

Graduate School of Public Affairs, University of Colorado at Denver,
Denver, Colorado, U.S.A.

INTRODUCTION

Deontology refers to any approach to ethics that emphasizes obligation or duty. The most notable deontologists are Immanuel Kant in the eighteenth century and John Rawls in the twentieth century.^[1,2] Deontological theories are typically contrasted with utilitarian theories. Deontology differs from utilitarianism insofar as it specifically excludes consequences as a basis for moral judgment, whereas utilitarianism claims that *only* consequences can serve to ground morality. The practical manifestations of deontology generally focus on principles and obligations, with attention to compliance and enforcement.

ADVANTAGES AND DISADVANTAGES OF DEONTOLOGY

The advantages of deontology are usually found in its intuitive appeal, as many people expect morality to concern principles and obligations. In this respect, the results of deontological thinking are perceived as theoretically well grounded and practically useful. Deontology also seems quite compatible with rights-based ethics, as one is obligated to respect the rights of others. However, this is not to say that deontology and rights always produce the same results. For example, there are obligations that might not have corresponding rights, such as keeping a generous promise.

Deontological approaches to ethics have been criticized in numerous ways. They offer a “thin” conception of morality, some argue, as they focus on rationality to the exclusion of other human qualities, such as emotion. Responses to these criticisms range from defenses of traditional theories to development of hybrid theories.

KANT

Historically, the most prominent deontologist has been Immanuel Kant, whose work remains central, if difficult reading for anyone interested in principle-oriented ethics. Although the *Groundwork of the Metaphysics of Morals* is only a small part of Kant’s work on moral theory, and not the central work at that, it is what most

academics intend when they refer to Kant’s ethics. A brief overview of the *Groundwork*, therefore, will suffice as a description of Kant’s deontological theory as it is typically characterized.

The primary goal of the *Groundwork* is to discover the supreme principle of morality. According to Kant, a principle of morality must carry with it absolute necessity, for nothing else, he argues, is adequate to the task of grounding obligation. Numerous concepts are particularly important for understanding his attempt to meet this challenge—a *good will*, *duty*, *autonomy*, and the *categorical imperative*.

Kant begins the *Groundwork* by explaining that nothing is good without qualification, except a *good will*. A good will is good through its willing alone, not through the resulting consequences. Furthermore, a good will is unqualified in this way insofar as it is determined solely by the motive of *duty*, for it is only through the motive of duty that a will is capable of obligating itself. Acting from duty requires acting from principle alone.

Autonomy is a property that a will has in virtue of its capacity for acting from duty, for being a law unto itself. An autonomous will is free in the sense that it determines the principle of its own volition apart from all natural interests or causes. Because only an autonomous will can determine itself, only such a will is capable of acting from duty. Because it is self-legislation that creates obligation, only a will that is self-legislating can fulfill an obligation. Only an autonomous will, therefore, can be good.

The *categorical imperative* is the practical expression of the principle of morality. It is a command of reason concerned only with its own form and the principle from which it follows, not with the action itself or the results. There are two specific formulations that are more frequently mentioned. First, *act only on that maxim through which you can at the same time will that it should become a universal law*. Second, *act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end*. One who is moral must necessarily will in accordance with the categorical imperative. However, even a conscious effort to act according to the categorical imperative does not guarantee moral success. One can never know whether one in fact has a good will.

RAWLS

In *A Theory of Justice*, John Rawls described what he calls “justice as fairness,” which consists of principles chosen from an “original position” of rational deliberation. He argued that principles of social justice concern the basic structure of society through the arrangement of major social institutions. Justice as fairness requires principles that “free and rational persons concerned to further their own interests would accept in an initial position of equality as defining the fundamental terms of their association.”^[2]

From this initial position, or original position, principles of justice are chosen from behind a “veil of ignorance,” without knowledge of anyone’s place in society or natural abilities. Rawls argued that two principles in particular would be chosen: 1) “each person is to have an equal right to the most extensive basic liberty compatible with a similar liberty for others” and 2) “social economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone’s advantage, and (b) attached to positions and offices open to all.”^[2]

Justice as fairness is deontological for a few reasons. First, Rawls’ account of justice fits within the social contract tradition, placing emphasis on the importance of *agreement* as a necessary condition of principles that can obligate, just as Kant claimed that moral principles must be self-legislated. A second reason is that Rawls himself labeled it as such. Because justice as fairness is not a teleological theory, such as utilitarianism, it is by definition deontological. Specifically, justice as fairness “does not interpret the right as maximizing the good.”^[2] “The original position,” Rawls said, can be viewed as “a procedural interpretation of Kant’s concept of autonomy and the categorical imperative.”^[2] Whereas utilitarianism values the satisfaction of all desires, justice as fairness accepts principles without knowledge of more particular interests.

DEONTOLOGY IN PRACTICE

Practical manifestations of deontology reflect the emphasis on obligations and principles. Although the principles are usually not explicitly self-legislated, they are typically considered principles that every rational agent would choose or should choose. It is this aspect that lends the weight of moral obligation. Deontologically grounded principles are typically directed toward basic concerns about the dignity and equality of persons. They frequently have the aim of preserving rational decision-making processes as well. Codes of conduct, regulations concerning financial conflicts of interest, and other sorts of “principle-oriented” efforts strive to clarify what is acceptable and what is not.

Deontology is a useful concept for administrators insofar as it helps to clarify the nature of administrative ethics. Traditionally, ethics in U.S. government has revolved around constitutionally based prohibitions. The deontological character of public administration consists of obligations that correspond to various rights, such as those mentioned in the U.S. Declaration of Independence—life, liberty, and the pursuit of happiness. Such “inalienable” rights obligate us in the strongest sense. However, the specific duties described in the U.S. Constitution tend to be negative. That is, government is prohibited from infringing on specific rights, such as free exercise of religion. Duties of public administration, therefore, tend to be indirect and prohibitive.

Two specific examples of this form of administrative deontology are programs directed by the U.S. Office of Government Ethics and the handling of compliance issues within human resources management. The Office of Government Ethics works to educate government employees about various obligations, most of which concern impartiality and the proper use of position. Through specific rules that prohibit or restrict such things as gifts and honoraria, these obligations help to preserve the integrity of public sector agencies, which are in turn an important piece of democratic decision making.

As important as expressions of democratic decision making are the various laws governing employment practices. The compliance issues encountered by human resource specialists are deontological, both in the sense that they concern obligations and in the sense that they concern the basic structure of the work environment, one that free and rational persons would choose.

Despite the solid grounding in negative duties, there is a growing trend within public administration toward positive duties. Whereas traditional duties have been correlates of rights, public administrators are more frequently thinking as professionals, for whom duties emerge from a sense of integrity and responsibility. Evidence of this trend is clearly visible in the American Society of Public Administration Code of Ethics, with its emphasis on personal integrity and professional excellence. There is a greater sense now than in the past that obligations must be self-imposed and principles are developed beyond external authority. In this sense, public administration is becoming more genuinely deontological.

REFERENCES

1. Kant, I. *Groundwork of the Metaphysic of Morals*; Paton, H.J., Ed.; Harper Torchbooks: New York, 1964.
2. Rawls, J. *A Theory of Justice*; Harvard University Press: Cambridge, MA, 1971; 11, 30, 60, 256.

Development Administration in Southeast Asia

Mark Turner

*School of Management and Policy, University of Canberra, Canberra,
Australian Capital Territory, Australia*

INTRODUCTION

Development administration was welcomed into newly independent countries of Southeast Asia after the Second World War as a vehicle for facilitating economic and social development. The rate and nature of its adoption varied between countries and have been mediated by factors such as resource endowment, development policies, political regimes, and history. Thus the contemporary profile of development administration in Southeast Asia varies significantly between countries in the region. The appreciation of the importance of development administration for improving human welfare and economic progress has never been stronger, with all the regions' countries currently engaged in a variety of reforms and innovations selected from a lengthening list of possible development administration initiatives.

FROM DEVELOPMENT ADMINISTRATION TO DEVELOPMENT MANAGEMENT

Development administration was a form of social engineering imported from the West in the 1950s and 1960s, which was built on a belief in the application of rational scientific principles and Keynesian welfare economics. Governments in the West and Southeast Asia perceived it as a managerial weapon in the war against communism, which would stem revolutionary appeal by delivering the benefits of modern capitalist development. Its success in this venture was circumscribed by the triumph of revolutionary forces in Vietnam, Laos, and Cambodia.

Development administration demonstrated faith in a "big government" and the leading role the state should play in development. It was also synonymous with public administration, although it was distinguished from practices in rich countries by "that inconvenient combination: extensive needs, low capacities, and severe obstacles."^[1] A tool bag of development administration interventions was created and transferred to Southeast Asia via foreign aid programs. In the 1970s, questions over the efficacy of this approach resulted in "a period of self-criticism, reflection, and uncertainty."^[2] Development administration had reached a "deadlock,"^[3] or was seen to be in "crisis."^[4]

A rethinking of development administration was urgently needed. This coincided with the rise of neo-classical economics in determining the development agenda. Big government was out, and private sector ownership and management techniques were in. Thus new public management (NPM) was soon being exported to the public sectors of developing countries in Southeast Asia, although its reception was less than welcoming in many instances. Participation also moved to the mainstream of thinking either by non-governmental organizations (NGOs), or by a renewed interest in territorial decentralization. The agenda of development administration broadened considerably beyond its early public administration focus to embrace diverse activities from privatization to community participation.

Today, we can define development administration, or development management as it is often known, as:

- An academic subfield in the social sciences
- Built around a set of problems relating to the management of development, rather than being a distinct body of theory
- Heavily but not exclusively focused on public administration
- Embracing a wide range of approaches to managing development, including actions by non-state actors
- Being concerned with power and politics as determinants of development policy making and its implementation
- Applying to almost 75% of the world's population, residing in countries displaying diverse characteristics.

CONTEMPORARY SOUTHEAST ASIA

Development administration operates in a variety of contexts in Southeast Asia (Table 1). At one extreme, there is Singapore with a GNI per capita above that of its former colonial ruler, the UK; whereas at the other extreme, there are impoverished nations such as Cambodia and Laos. Populations are as large as 212 million in Indonesia, but are as low as 350,000 in Brunei. There are countries that have functioning democracies, semidemocracies, and one-party or authoritarian states. Culture is variegated both between

Table 1 Selected environmental features of Southeast Asian countries

Country	Population of 2002 (millions)	GNI per capita PPP of 2002 (US\$)	GDP growth rate per capita (%) of 1999–2002	Adult literacy (%) age 15 years and above of 2002	Life expectancy of 2002 (years)	Political regime of 2002	Leading religion of 2002
Brunei	0.35	19,210		91.6	76.6	Sultanate	Islam
Cambodia	12	1,590	3.7	69.4	54.0	Multiparty democracy	Buddhism
Indonesia	212	2,990	1.7	87.3	66.2	Multiparty democracy	Islam
Lao PDR	6	1,610	3.5	66.4	54.5	One-party state	Buddhism
Malaysia	24	8,280	2.2	88.4	72.8	Semidemocracy	Islam
Myanmar	49	1,027	9.9	85.3	57.2	Military authoritarian	Buddhism
Philippines	80	4,280	1.8	95.4	69.8	Multiparty democracy	Roman Catholic, Islam
Singapore	4	23,090	3.4	92.8	78.4	Semidemocracy	Various
Thailand	62	6,680	3.3	95.8	69.2	Multiparty democracy	Buddhism
Vietnam	81	2,240	4.4	92.9	69.7	One-party state	Buddhism

Source: Ref.^[5].

countries and even within them. Thailand is predominantly Buddhist, the Philippines is Roman Catholic, and Indonesia is Moslem. However, there are small but significant Islamic minorities in Thailand and the Philippines, whereas Indonesia has substantial Christian and Hindu minorities. What this means is that the environment for development administration in Southeast Asia differs between countries—often quite dramatically. Thus the practices and innovations that are employed under this broad umbrella are many and vary according to circumstances.

PUBLIC ADMINISTRATION AND ITS REFORM IN SOUTHEAST ASIA

In Southeast Asia, the concept of bureaucracy as set out in Weber's ideal type has retained a strong appeal among civil service elites, who have most often been able to direct the design and implementation of public administration reform. They are attracted to what the model appears to offer—efficient administration, reliance on formal rules, hierarchy, and upward accountability—rather than by dysfunctions that have frequently characterized public administration in the region. For example, the Philippine public administration has been criticized for being under the influence of “partisan politics, geographical ties, the compadre system, familial obligation, and personal factors.”^[6] In Indonesia, dysfunctional civil service practices have included the sale of positions, extortion of money for promotions, seeking bribes for favorable decisions, and marking up project costs to secure illegal payments. Among the problems in Thailand's public service have been “patronage and corruption in appointments and promotions, lack of performance incentives, and overcentralization.”^[7] In Vietnam, there has been concern with the failure to integrate public administration reform with political strategies, the poor quality of legal documents, the overlapping and unclear definition of organizational functions, and low skill levels and budgeting geared to the number of staff rather than functional priorities or results.^[8]

The dysfunctions of public administration have not been ignored in Southeast Asia. There is a long history of reform, although there have been considerable contrasts between countries. For example, Singapore and Malaysia have almost 50 years of experience with successful incremental reforms, whereas in the Philippines, despite most presidents having a comprehensive public administration reform program, not one has ever been implemented.^[9] The governments of Vietnam and Lao PDR have, until recently, been reticent to engage in substantial public administration reforms, perhaps fearing a diminution of political control.

However, today, all countries in Southeast Asia are engaged in public administration reform, often with great enthusiasm. The reasons for the growing importance of public administration are as follows:

- A perceived link between public administration and international competitiveness
- The potential contribution good public administration can make to sustainable human development
- Democratization requiring reconsideration of the relationship between state and society
- The promotion and support of good governance, including public administration reform, by multilateral and bilateral development agencies.

The dominant theme in public administration reform across Southeast Asia has been bureaucratic modernization. Although the preferred method of reform has been incremental adjustments, there is currently some interest in systemic reforms. For example, in Vietnam, a Public Administration Master Plan has been introduced for the period 2001–2010. It builds on prior incremental changes and includes the following:

1. Program to renovate the development, issuance, and quality improvement of legal normative documents
2. Program on roles, functions, and organizational structures of the agencies in the administrative system
3. Program on staff downsizing
4. Program on improving the quality of cadres and civil servants
5. Program on salary reform
6. Program on renovation of financial management mechanisms for administrative and public service delivery agencies
7. Program to modernize the administrative system.

These are not new ideas. The novel aspect is putting them together in an integrated program on a major scale, guided by a vision of what the government hopes to achieve. Thailand has also commenced an ambitious program of systemwide public administration reforms. The civil service has, in the past, been able to resist any radical reforms due to its powerful position in the Thai state. However, the gradual erosion of that power and the impact of the Asian financial crisis in 1997 provided the policy opportunity and impetus for the introduction of far-reaching changes to the public service. These include a move toward “strategic performance-based budgeting”; the introduction of accrual accounting; a more efficient and transparent procurement process; improved procedures in revenue collection, especially through new information and communications technology; reorganization of the

structures and functions of ministries; the introduction of results-based management; and the establishment of a Public Sector Development Commission (PSDC) to monitor public sector reform and advise the cabinet on policy.^[10]

Although there has been considerable interest across Southeast Asia in public administration reforms in other countries, there has been only a limited adoption of NPM. If we view NPM as a menu of items, then we can classify the countries of Southeast Asia into three distinct groups according to the number of items selected for implementation.^[11] Singapore and Malaysia are “enthusiastic diners” having introduced a wide range of NPM-style reforms. However, some initiatives are not new for these countries, whereas other items from the NPM menu have not been introduced. Borrowing reform initiatives and modifying them to suit local conditions have long histories in Malaysia and Singapore. The Philippines, Thailand, and Indonesia can be described as “cautious diners” because they are familiar with the menu but so far have experimented with a few reforms from the NPM menu. However, Thailand’s new reform program may lift it into the “enthusiastic diners” category. Finally, there are those countries that are “unfamiliar with the menu,” such as Lao PDR, Cambodia, and Vietnam. The governments appreciate the need for public sector reform and are, to some degree, engaged in it. However, they lack institutions and other environmental conditions necessary for NPM reforms.

There are several important variables that explain the differential adoption of NPM and the widespread persistence of bureaucratic modernization as the overriding reform theme across Southeast Asia. First, the most NPM-friendly countries are also the wealthiest, and, in the cases of Singapore and Malaysia, have many institutions derived from Britain—the heartland of NPM. In poor countries such as Lao PDR or Cambodia, NPM-style reforms may be unaffordable, impractical, and irrelevant. Second, state–society relations in Southeast Asia involve greater responsibility for national development being claimed by the state than in Organization for Economic Cooperation and Development (OECD) countries. Third, there are Southeast Asian values that may clash with NPM notions of public administration reform. These values could include group reference rather than individual reference; conflict avoidance; importance of “face”; respect for authority and seniority; paternalism; respect for academic credentials; undervaluation of the professional role of women; belief in cosmology and superstition; and the importance of family support.^[12,13] Although these values may also be found in Western countries, their relative importance and particular combination in Southeast Asia produce different behavioral outcomes. Fourth, the political

regimes of Southeast Asia show considerable diversity—ranging from authoritarianism in Burma, through one-party regimes in Vietnam and Lao PDR, to semi-democracies in Malaysia and Singapore, to “elite democracy” in the Philippines. Some of these regimes are not receptive to particular NPM initiatives, especially those that could be seen to diminish regime power and legitimacy. Finally, some items on the NPM menu fit well with the longstanding Southeast Asian theme of bureaucratic modernization focusing on efficiency gains.

DECENTRALIZATION

One of the most significant trends in development administration in Southeast Asia has been decentralization. Power and authority have been taken from centralized bureaucracies and have devolved to elected local councils or deconcentrated offices of central agencies. The driving forces behind decentralization have been democratization and the belief that service delivery will be improved. The Philippines was the first country to decentralize through the Local Government Code of 1991. This legislation gave local government units responsibility for basic services in health, agriculture, public works, social welfare, and environment and natural resources. Funds to pay for these new functions were also given to the elected councils at provincial, city, municipal, and *barangay* (community) levels. Participation by NGOs and the private sector in local governance was also decreed in the Local Government Code of 1991. The general opinion is that the decentralization in the Philippines has been “fairly successful” in bringing devolved democratic governance to people across the Philippines.^[14,15] Whether it has led to improved service delivery is more difficult to determine.

Indonesia was the next country to engage in decentralization. Under the authoritarian New Order regime of President Suharto, Indonesia had been a highly centralized state. Hasty legislation (Laws 22 and 25 of 1999) after Suharto’s removal from office resulted in Southeast Asia’s most radical decentralization program. The regional hierarchy was abolished, leading to a considerable reduction in the authority and importance of the province, and a great increase in the authority of the districts. A long list of functions was placed under the control of the autonomous districts. The list includes public works, health, education and culture, agriculture, communication, industry and trade, capital investment, environment, land, cooperatives, “manpower affairs,” and the management of national resources. Locally elected assemblies have become very important as they now set regional policy and manage vastly increased resources. Implementation

has been problematic. The schedule was too short; the legislation had some fundamental weaknesses and is to be revised; the regulatory program necessitated by Laws 22 and 25 has not been completed; there are regional inequities; accountability provisions are suspect; and “money politics” has arrived in the regions.^[16]

Further experiments in decentralization have taken place in Thailand where the current government is enhancing the powers of the provincial governors—now referred to as chief executive officers (CEOs)—at the expense of central ministries, and is also trying to enhance the role of the grassroots Tambon Councils. In Cambodia, steps have been taken to both devolve and deconcentrate.^[17] Devolution is to the elected Commune Councils while deconcentration is to the field offices of ministries such as health, education, and agriculture. Progress has been slow, with very limited resources and functions devolved to the Commune Councils, whereas deconcentration moves at a pace and in a manner determined by cautious central ministries.

CONCLUSION

All governments in Southeast Asia appreciate the importance of development administration for national development, and all are engaged in reform initiatives to improve the performance of their public service and local government institutions. The longstanding theme of bureaucratic modernization is still evident in public administration reforms to increase the efficiency of the government by initiatives such as reducing red tape, introducing ICT, and downsizing and strengthening budgetary procedures. This has been accompanied by decentralization experiments in the Philippines, Indonesia, Cambodia, and Thailand. The pace and content of reforms have varied between countries according to a range of environmental features. Interest in reform and continued activity in the field will undoubtedly persist as Southeast Asian countries seek to improve their international competitiveness, extend participation and democracy, and seek to improve the amount and quality of services delivered to their citizens.

REFERENCES

- Schaffer, B.H. The deadlock in development administration. In *Politics and Change in Developing Countries*; Leys, C., Ed.; Cambridge University Press: Cambridge, 1969; 177–211.
- Turner, M.; Hulme, D. *Governance, Administration and Development: Making the State Work*; Macmillan: Basingstoke, 1997.
- Hirschmann, D. Development administration? A further deadlock. *Dev. Change* **1981**, *12* (3), 459–479.
- Dwivedi, O.P.; Nef, J. Crises and continuities in development theory and administration. *Public Adm. Dev.* **1982**, *2* (1), 59–77.
- Statistics are from the websites of the World Bank (www.worldbank.org), United Nations Development Programme (www.undp.org) and Asian Development Bank (www.adb.org), accessed on March 10, 2004.
- De Guzman, R.P.; Reforma, M.A. Administrative reform in the asian pacific region: issues and prospects. In *Administrative Reform Towards Promoting Productivity in Bureaucratic Performance*; De Guzman, R.P., Reforma, M.A., Eds.; EROPA: Manila, 1993; 2–14.
- RIAP (Research Institute for Asia and the Pacific). *Building Institutional Capacity in Asia: Public Sector Challenges and Government Reforms in South East Asia*; RIAP, University of Sydney: Sydney, 2001.
- GSC-PAR (Government Steering Committee for Public Administration Reform). *The Overall Report: Review of Public Administration Reform*; GSC-PAR: Hanoi, 2000.
- Turner, M. Choosing items from the menu: new public management in southeast asia. *Int. J. Public Adm.* **2002**, *25* (12), 1493–1513.
- World Bank. Thailand country development partnership: governance and public sector reform, monitoring workshop, Bangkok, June 19, 2003; OCSC (Office of the civil service commission). In *The Next Steps of Public Service Reform*; OCSC: Bangkok, 2000.
- Choosing items from the menu: new public management in southeast asia. *Int. J. Public Adm.* **2002**, *25* (12), 1493–1513.
- Mauzy, D.K. The human rights and ‘Asian values’ debate in southeast asia: trying to clarify the key issues. *Pac. Rev.* **1997**, *10* (2), 210–236.
- Mahbubani, K. *Can Asians Think?* Times Books International: Singapore, 1997.
- Tapales, P.D. The Philippines. In *The Changing Nature of Local Government in Developing Countries*; McCarney, P.L., Ed.; Centre for Urban and Community Studies, University of Toronto: Toronto, 1996; 197–219.
- Rood, S. *An Assessment of the State of Knowledge Concerning Decentralized Governance Under the 1991 Local Government Code. Paper Presented at the Third European Conference on Philippine Studies*; April 27–29, 1997.
- Turner, M.; Podger, O. *Decentralisation in Indonesia: Redesigning the State*; Asia Pacific Press: Canberra.
- Turner, M. Whatever happened to deconcentration? Recent initiatives in Cambodia. *Public Adm. Dev.* **2002**, *22* (3), 353–364.

Development Policy

Lawrence S. Graham

Department of Government, University of Texas, Austin, Texas, U.S.A.

INTRODUCTION

Development policy is the general term used to capture the dynamics and the processes of purposive change focused on creating the conditions for economic growth, technological change, and human development. There is growing consensus that although economic growth centered in markets and measured in terms of increases in gross domestic product (GDP) is indispensable, it is not a sufficient condition to bring about improvements in the human condition in such a way that human development can occur. Although the human development concept embraces numerous different indices, it is generally defined as the process of enlarging people's choices in terms of securing human rights, guaranteeing their basic necessities, and enabling them to improve the quality of their own lives. Linked to this shift toward thinking of development in more humane terms are initiatives involving empowerment, participation, equity, and sustainability. In this debate over how to both engender economic growth and increase the capacity of people in all nations and communities to improve their conditions of life, the tasks remain daunting.

Although problems of poverty, ethnic and cultural conflict, and inequities in income distribution and in opportunities to improve conditions of life continue to present obstacles, advances have been made in setting priorities for the development programs and projects that comprise development policy. In the midst of thousands of reports and publications related to development, the annual publication of the World Bank's World Development Report (WDP) and the United Nations Development Programme's Human Development Report (HDP) are primary sources of information and valuable documents to consult in understanding the changing agendas that dominate the discourse on development. As a consequence, in exploring development issues, reference to the Human Development Index and awareness of the range of indicators used for this index are as important as discussions of GDP, tables surveying economic growth, and understanding the economic and social indicators used to measure development on a worldwide basis. Especially useful documents are the WDP and HDP reports for 2000, issued at the turn of the century, which reflect the changing concepts and increased understanding of issues involved in development policy.

LINKING PUBLIC ADMINISTRATION AND POLICY WITH DEVELOPMENT

There is a relevant public administration and policy literature that reflects changing perceptions of the issues linked to development over the last half century. Although there was an earlier literature on civil service systems in Western Europe published in the 1930s, generally comparative concerns in public administration and policy date back to the 1960s, to the formation of the Comparative Administration Group (CAG), under the leadership of Fred W. Riggs, with funding from the Ford Foundation. Despite criticism of the CAG's publications and Riggs' work on the ecology of public administration in developing countries on the grounds that they were excessively theoretical, they generated a new body of scholarship, centered largely in South and East Asian experience, which called attention to the very different ethos surrounding public administration in the developing countries. Names that stand out in this literature are William J. Siffin, John Montgomery, Ralph Braibanti, Ferrel Heady, and Milton J. Esman. Many of these scholars responded to the critique of excessive theorizing in the CAG by emphasizing that the endeavor they were embarked on was development administration, explaining the conditions in the national bureaucracies of the developing countries that worked against development initiatives. What was needed, they emphasized, was a more relevant public administration, capable of responding to the issues they had confronted in programs and projects linked to economic development, social change, and improvement in such diverse areas as agriculture, public health, and education. Sources summing up this debate and the evolution of this field into what has become known as comparative and development administration are works authored or edited by Heady, Esman, and Farazmand.

Although the public administration literature that is international in its perspective and identified with a professional school orientation has continued to focus on the Third World across the years, an important parallel is to be found in the comparative policy work of political scientists such as Peters; Heidenheimer, Heelo, and Adams; Aberbach, Putnam, and Rockman; and Freeman. Essentially European in its concerns, as this literature has evolved, it has focused more generally on the industrialized countries and the transformation of

these societies through changing technologies, immigration policy, and social welfare.

Both perspectives on public administration and policy in the developed and the developing countries, however, have undergone considerable evolution, as the issues of development and change in the modern world have become more complex. In the public administration field, in addition to evolving perspectives on comparative and development administration, the impact of business administration can be traced in the literature on public management and the emergence of a distinctive development management literature (with its own particular concerns with public management in developing countries and the techniques and skills public managers need for their work in government and civil society, in postindustrial and industrializing societies). More recent concerns, in the shift toward policy and process in development work, are to be found in the policy reform literature. The best way to capture these developments is to consult the work of Brinkerhoff and Crosby.

Yet another path taken in the public administration and policy field, which falls within the purview of development policy, is that centered on institutional analysis and development. Derived from the rational choice institutionalist literature, the evolution of this line of thinking about how to better confront development dilemmas is best reflected in Ostrom's work. The themes dominating this writing concern the need for precise, operational concepts and applications in policy analysis that focus on thinking of institutions as rules that over time structure politics, policy, and action, and have a direct impact on political and administrative behavior in various national settings, be they Asian or Western European. Ostrom's work in development policy is explicit in how it lays out the various incentives and disincentives that influence policy in such areas as common property resource management. Issues concerning communal grazing, irrigation systems, fisheries, and forestry services, and how to retain a common set of rules and expectations center on the use of these resources in such a way that they can be sustained over time without destroying them. Another area of activity involves policy issues concerning questions of infrastructure development and maintenance—for example, the use of simple and complex market arrangements, centralized management and control initiatives, decentralized policy incentives, and polycentric management schemes. These latter, more complex schemes aim at providing more dynamic mixes of centralization (to provide more effective coordination and control of development initiatives) and decentralized programmatic initiatives (to encourage wider citizen participation and greater involvement of the users of public services) in maintaining roads, sufficient water supply, and sanitation and waste systems.

CONCLUSION

The most important point to be made about development issues in public administration and policy is the convergence that has been occurring in the contemporary world in its concern with issues related to economic growth, the maximization of social capital, and sustainable conditions of life. Although there are obviously enormous differences between the most technologically advanced countries and the least developed societies and economies, we can no longer comfortably focus on the shift to knowledge-based industries in the new economies of the more advanced countries, to the exclusion of the enormous problems in public health, poverty, and deficiencies in education that confront those societies, which remain apart from the market economies and levels of living that have accompanied the affluence generated in the First World, as opposed to the poverty and hardships characteristic of the Third World. The confluence of both developments is foremost among the challenges to be faced in responding to the enormous changes occurring in science and technology, in the economy, and in society, and in confronting the conflicts and tensions present when these two worlds intersect. No document sums up more effectively this convergence in development policy (with its emphasis on the need to reconcile security concerns with those of development in a world that no longer can separate issues of development from those of underdevelopment) than the European Commission's May 2001 report, *The European Union's Role in Promoting Human Rights and Democratisation in Third World Countries*. This report adopts the language of development policy and focuses on human development as the basis for promoting human rights globally, and creating more open societies and economies.

BIBLIOGRAPHY

1. Aberbach, J.D.; Putnam, R.D.; Rockman, B.A. *Bureaucrats and Politicians in Western Democracies*; Harvard University Press: Cambridge, 1981.
2. Brinkerhoff, D.; Crosby, B. *Managing Policy Reform: Concepts and Tools for Decision-Makers in Developing and Transitioning Countries*; Kumarian Press: West Hartford, CT, 2001.
3. Communication from the Commission to the Council and the European Parliament. *The European Union's Role in Promoting Human Rights and Democratisation in Third Countries*; European Commission: Brussels (COM [2001] 252 final, 8 May 2001).
4. Esman, M.J. *Management Dimensions of Development: Perspectives and Strategies*; Kumarian Press: West Hartford, CT, 1991.
5. Farazmand A., Ed.; *Handbook of Comparative and Development Administration*; Marcel Dekker: New York, 2001.

6. Freeman, G., Jupp, J., Eds.; *Nations of Immigrants: Australia, the United States, and International Migration*; Oxford University Press: New York, 1992.
7. Heady, F. *Public Administration: A Comparative Perspective*; Prentice-Hall: Englewood Cliffs, NJ, 1966 (with numerous revisions and expansions up through the 1980s).
8. Heidenheimer, A.J.; Heclo, H.; Adams, C.T. *Comparative Public Policy: The Politics of Social Choice in Europe and America*; St. Martin's Press: New York, 1975.
9. Ostrom, E. *Governing the Commons: The Evolution of Institutions for Collective Action*; Cambridge University Press: New York, 1990.
10. Ostrom, E. Institutional rational choice: an assessment of the institutional analysis and development framework. In *Theories of the Policy Process*; Sabatier, P.A., Ed.; Westview Press: Boulder, CO, 1999.
11. Ostrom, E.; Schroeder, L.; Wynne, S. *Institutional Incentives and Sustainable Development: Infrastructure Policies in Perspective*; Westview Press: Boulder, CO, 1993.
12. Peters, B.G. *Comparing Public Bureaucracies: Problems of Theory and Method*; University of Alabama Press: Tuscaloosa, AL, 1988.
13. United Nations Development Programme. *Human Development Report 2000*; Oxford University Press: Austin, 2000.
14. World Bank World Development Report 1999/2000. *Entering the 21st Century (2000)*; Oxford University Press: New York, 2007.

Discipline and Dismissal

Steven W. Hays

*Department of Government and International Studies, University of South Carolina,
Columbia, South Carolina, U.S.A.*

INTRODUCTION

Discipline and dismissal comprise the human resource management processes by which organizations seek to correct deviations from acceptable conduct, and to rid themselves of individuals who are unable or unwilling to comply with reasonable standards of performance. Discipline is an ongoing responsibility of every manager, all of whom are expected to identify problems arising from employee behavior and introduce corrective measures. The application of discipline is not necessarily a negative process, in that the ultimate objective of a disciplinary procedure should be to provide employees with proactive guidance on how to improve their deficiencies. Dismissal, in contrast, should occur only when a worker has committed an offense that is so serious that it demands immediate action, or has proven to be unable to perform at an acceptable level after being provided with ample opportunity to address performance shortcomings.

THE GOVERNMENTAL CONTEXT OF DISCIPLINE AND DISMISSAL

Unlike conditions that prevail in much of the private sector, public managers confront numerous additional constraints when imposing discipline or undertaking dismissal actions against workers. Non-unionized private employees have few recognized rights in their employment. They serve “at will,” meaning that they can be hired or fired with or without a substantive cause. In contrast, civil servants enjoy a considerable number of protections from any “adverse action” that is pursued by a supervisor. Because of the fear that political factors might contribute to their mistreatment or termination, public employees typically cannot be sanctioned without careful attention to numerous procedural, legal, and constitutional guidelines. In legalistic terms, this means that public workers retain certain property and liberty interests in their employment that require extreme caution on the part of supervisors seeking their punishment or termination.

In brief, civil servants are entitled to due process whenever faced with a disciplinary action that substantially threatens their job status, such as a demotion,

suspension, or termination. Due process potentially includes many factors, but the most basic requirements are as follows:^[1]

1. Managers must abide by the human resource management procedures that are in place; any requirements included in a policy manual, employee handbook, or other document might be interpreted in court as an “implied contract.”
2. Any adverse action must be based on a job-related reason or cause; employees cannot be sanctioned for actions or behaviors that do not influence their job performance in some negative fashion.
3. Employees are protected by all relevant anti-discrimination legislation; disciplinary or termination actions are forbidden if they are motivated by discriminatory intent involving a person’s race, religion, ethnicity, age, sex, disability, or other protected categories.
4. Supervisors cannot ignore constitutional protections during disciplinary proceedings; employees retain all constitutional rights to free speech, assembly, and other liberties, unless the exercise of those rights severely impinges on the agency’s ability to deliver public services effectively.
5. Any adverse action that is conducted in an “outrageous” manner or that is contrary to “public policy” is illegal. Outrageous actions include those that are conducted in a harmful, crass, and embarrassing manner (e.g., a highly public tongue-lashing), whereas violations of public policy occur when punishment is based on malice, bad faith, or is contrary to statutory requirements. For instance, employees cannot be fired in retaliation for filing sexual harassment charges or worker’s compensation claims, nor can they be punished for complying with a legal obligation (e.g., testifying under subpoena against the employer).

Despite these restrictions on discipline and dismissal, the widespread perception that civil servants “can’t be fired without an act of God” is seriously

exaggerated. In most jurisdictions, the necessary procedures are in place for effective employee discipline, provided that supervisors play by the rules. Another important development in recent years has been the impact that the reinvention movement has exerted on employee protections. The “merit systems” of many states have been abolished or weakened, a phenomenon that generally reduces employees’ ability to challenge the personnel decisions of their superiors. Many of the traditional systems of employee protection—such as multiple avenues of appeal and unreasonable restrictions on managerial discretion—have been dismantled. For these reasons, there is a clear and distinct trend toward exposing civil servants to greater risks of being disciplined and/or terminated for misconduct or poor performance.^[2]

PROGRESSIVE DISCIPLINE PROCEDURES

Whatever their legal status in challenging supervisory actions, public servants are like any other group of workers. Most are well intentioned and industrious, yet there are always some (hopefully only a few) who are “problem employees.” The presence of such individuals in the organization has multiple repercussions. It has been estimated that a poorly performing worker will consume at least 10 times more of the manager’s time than any (or all) of the competent employees. Clearly, the best way to address this crisis is to avoid it in the first place. That is, a properly functioning human resource management system should weed out most of the misfits before they are hired. Likewise, mentoring systems, careful background checks, and the effective use of probationary periods are helpful ways of solving potential problems before they become acute. Only after an employee has successfully completed a probationary period—usually 1 calendar year—does that individual acquire legal rights to challenge adverse personnel actions. But once that threshold has been successfully crossed, public managers need to exercise care and tact in the handling of employees whose performance is subpar.

A Non-Punitive Focus

The use of a progressive discipline procedure is generally considered to be an effective means of responding to problem employees in contemporary public agencies. These procedures are based on the premise that discipline ought to have a positive rather than a punitive focus. The underlying ethos is to provide the offending worker with notice that a problem exists. Once the performance dilemma has been identified, the requisite steps to resolve the situation should then

be articulated. Thus, supervisors have the obligation to confront workers with their complaints and to delineate the specific reasons why they are displeased. Especially in this era of worker empowerment, managers also are expected to involve the subordinate in the problem-solving process. This might be accomplished by discussing the problem with the worker, allowing them to suggest appropriate responses. A poorly performing individual may need additional training or simply the reality check that comes with a confrontation from the boss. In other cases, a detailed work improvement plan may need to be established. The basic point is that no worker can be expected to remedy a deficiency until she has been notified that the problem exists and then provided with feedback on what constitutes acceptable behavior.

Non-punitive approaches to employee discipline are believed to create fewer tensions between managers and workers, and therefore have a greater likelihood of use and success. Research has also demonstrated that managers need to be trained in the technical and behavioral aspects of disciplinary actions.^[3] In addition to increasing the manager’s self-confidence, training improves their appreciation for the legal and emotional pitfalls that surround employee discipline. Another noteworthy characteristic of the non-punitive approach is that, if it is followed carefully, it will withstand most legal challenges by disgruntled employees. Perhaps the most oft-cited explanation for worker grievances against their supervisors is that they were “surprised” by the discipline that was invoked. A regularized system of meeting and discussing performance problems reduces the frequency of such complaints.

Consistency Through Procedural Guidance

Another possible pitfall in the application of discipline is the potential for treating employees arbitrarily or differentially. That is, one worker might be punished for a minor infraction, whereas other employees committing the same offense are ignored. If this phenomenon occurs (which it often does), the sanctioned employee can successfully challenge the disciplinary action on the grounds that they have been singled out for punishment. The essential issue is that employees must be treated in a consistent fashion or else the entire disciplinary scheme is threatened. Although certainly not a panacea, the prescribed response to this dilemma is to provide employees and supervisors with a comprehensive list of infractions and the sanctions that will be imposed for each incident. These progressive discipline procedures are usually inserted in employee handbooks, and consist of lengthy lists of violations accompanied by the punishment that will be invoked for the first, second, and third offenses.

For instance, tardiness might be countered with a verbal warning in the first instance, a written warning on the second occurrence, and a formal reprimand on the third occurrence. More serious offenses, such as stealing or sleeping on the job, might call for immediate termination for the first or second offense.

Benefits of a Written Procedure

The existence of a written procedure fulfills three important purposes. First, it constitutes a form of prior notice to employees concerning the norms of acceptable and unacceptable conduct. This makes it more difficult for a worker to allege that they “didn’t know” that a particular behavior (e.g., making personal telephone calls on agency time) was forbidden. Second, the procedure contributes to the organization’s effort to ensure that workers are treated consistently under the same set of circumstances. A clear delineation of appropriate sanctions essentially constitutes a disciplinary template that ought to be applied uniformly. (No one would be naïve enough to assume that such procedures always result in consistent treatment, but at least their presence provides supervisors with guidance in pursuit of that elusive goal.) Finally, the creation and dissemination of a list of prescribed conduct strengthens the agency’s hand in complying with a number of legal expectations concerning the handling of its workforce. One notable example is sexual harassment. Prior to the last decade or so, concise agency standards concerning the punishment for various types of sexual misconduct were rare. Now, with the inclusion of expansive sexual harassment provisions in the agencies’ progressive discipline codes, the organizations are able to demonstrate a proactive posture toward eliminating such conduct. This does not exonerate them from liability if they fail to act, but it does demonstrate to judges and juries that they have taken steps to avoid problems and to punish transgressors.

DISMISSAL: THE ULTIMATE SANCTION

When positive efforts to correct employee misconduct fail, the agency is confronted with a narrow range of unpleasant choices. The honorable, but often most difficult, response is to initiate formal termination proceedings. This requires that the supervisor be able to document the performance or behavioral deficiencies, and be prepared to spend considerable time justifying and defending their actions in the event that the employee files a grievance. This occurs with the prior knowledge that the supervisor’s conduct will be exposed to as much scrutiny as that of the offending worker. Supervisors must be able to demonstrate adequate

reasons (job related) for their actions, and to show that they followed all applicable procedures and treated the employee fairly. Dismissal also requires, of course, the necessity for a face-to-face session in which the decision is explained to the employee. Because this is a messy and time-consuming process, terminations in government are far less common than in the private sector. On average, private employers terminate about 8% of their workforces annually, whereas dismissal rates in government hover below 1%.^[4]

There is little mystery about what happens to public employees who deserve to be terminated but who avoid formal dismissal actions. Too often, the agency’s response is either to ignore the inappropriate conduct, or to deal with it in a more proactive yet underhanded fashion. Common responses include promoting the bad employee “up and out of the way,” providing a glowing letter of recommendation so as to pawn the worker off on another agency, or to transfer them to a position that carries no responsibility (so-called “turkey farm” jobs). These are age-old “solutions” to the problem employee puzzle, but clearly represent inadequate (even dishonest) responses that are unfair to everyone involved.

Fortunately, the number of successful terminations in public agencies is likely to increase over time in response to the reduction in employee protections that is occurring as part of government reinvention. A related phenomenon is the increased use of alternative dispute resolution techniques, such as mandatory mediation and perhaps even binding arbitration. By reducing the number of appeals and expediting the review process, reformers hope to increase government’s ability to rid itself of inept and corrupt employees. As such, there is a discernable move away from highly formalized litigation that is time consuming, expensive, and intimidating to all concerned. The only downside of this trend is the possibility that some innocent and conscientious civil servants might be sacrificed as procedural protections are reduced. In the eyes of most reformers, however, this is a risk that they are willing to take.

REFERENCES

1. Tidwell, R. Employment at will: limitations in the public sector. *Public Pers. Manage.* **1983**, 12 (3), 293–304.
2. Walters, J. Who needs civil service? *Governing Magazine* **1997**, 10 (8), 17–21.
3. Belohlav, J.; Popp, P. Making employee discipline work. *Pers. Adm.* **1983**, 23 (3), 22–24.
4. Hays, S. Employee discipline and removal: coping with job security. In *Public Personnel Administration: Problems and Prospects*, 3rd Ed.; Hays, S., Kearney, R., Eds.; Prentice Hall, 1995; 145–161.

Diversity

Sonia M. Ospina

James F. O'Sullivan

*Robert F. Wagner Graduate School of Public Service, New York University,
New York, New York, U.S.A.*

INTRODUCTION

Public agencies have the mandate to consider the plurality of values, concerns, and voices of the larger population in their work, as well as to include a wide variety of citizens in their workforce. When diversity is pursued as an organizational objective, more efficient management and the democratic values of responsiveness and representation in public administration are both said to be better achieved.

Workforce diversity emphasizes functioning with employees with differences in personality, work style, and the visible manifestations of social diversity, such as gender and ethnicity, and using those differences to better fulfill the organization's mission. According to the literature, increased workforce diversity in public organizations provides many benefits. These benefits include increased production capacity and organizational flexibility, contributions to fairness in the workplace, greater compliance with personnel regulations, increased representation and responsiveness among members of the bureaucracy, and more grassroots support for agency programs and policies. Also, the organization may enhance its reputation and, as a result, be able to attract and keep the best employees.

Unresolved diversity problems, it has been shown, often deflect employee energy and attention from performance, with a consequent loss of productivity, and can cause an organization to lose valued employees. Critics of diversity management sometimes argue that achieving diversity requires an organization to compromise on the quality of its employees or services, but merit should not be an issue if job attributes are well defined and staffing practices are not limited by stereotypes.

Diversity management in the public sector, as in any organization, is a multidimensional challenge that requires permanent and focused managerial attention. Proponents of diversity management suggest that organizations must be able to realistically diagnose the status of their organization to determine the right strategies for diversity management, and several schema have been developed for that purpose.

WORKFORCE DIVERSITY IN PUBLIC ORGANIZATIONS: A NEW IMPERATIVE

The shifts from an industrial to an information-based society and to a service economy have radically challenged assumptions about how to best organize tasks and people. They have also raised demands for organizations to innovate their production functions and to use more diverse talents. Different work arrangements, flatter organizational structures designed around teams and networks, and more permeable boundaries for intra- and interorganizational cooperation have produced a new complexity in work operations. A diverse workforce—in both functional and social terms—becomes a prerequisite for large organizations in both the private and public sectors to manage their turbulent environment.

Workforce diversity has, indeed, become recognized as an imperative for organizational competitiveness and effectiveness,^[1,2] and diversity management is increasingly becoming a principle of human resources (HR) management.^[3] Today, state-of-the-art management literature in both public and private sectors views workforce diversity not as a problem to be managed away but as a requirement of organizational effectiveness.^[2] Traditional affirmative action and equal employment opportunity efforts are no longer viewed as sufficient in themselves to manage today's workforce effectively, even though they may still be an important component of the agenda.

Increased workforce diversity in public organizations, according to the literature, results in ethical, legal, HR management and organizational benefits.^[4] Ethically, diversity contributes to fairness in the workplace, and helps to create economic opportunity and reduce social inequality. Legal and policy benefits include greater compliance with personnel regulations, increased representation and responsiveness among members of the bureaucracy, and increased grassroots support for agency programs and policies. Human resource benefits are an enhanced agency reputation and the resulting ability to attract and keep the best

employees, while promoting creative approaches to work. Organizational benefits are substantial, including increased internal capabilities, greater organizational flexibility and ability to address change and fluidity in organizational design, and decreased discrimination litigation. These claims appear repeatedly in normative and empirically based literatures.

Diversity challenges are compounded in public agencies by specific pressures they face as organs of government. The government environment in the United States is characterized by a changing workforce, declining confidence in government, reduced budgets, downsizing at the federal level, higher demands for productivity, and proposals for alternative approaches to public service, including contracting out and privatization.^[5] In this context, two broad diversity demands affect contemporary public personnel systems. One is the demand for increased performance, whereby diversity becomes a performance requirement; the other is a legitimization demand, in which workforce diversity is a tool to address the political and ethical mandates for a bureaucracy that is representative of the citizenry. Combined, these two demands produce a strong incentive to pursue and manage workforce diversity effectively.

The published literature on diversity is large but limited. In a literature review of diversity, Wise and Tschirhart^[6] found 106 empirical studies of the outcomes of diversity, published between 1961 and 1998 in peer-reviewed academic journals. Thousands of other papers related to diversity and diversity management were published in the same 37 years, but they were mainly descriptive, reviewed ideological or policy perspectives on diversity, or presented survey findings on attitudes toward aspects of diversity. The authors also found that the generalizability of published empirical studies was limited because experimental populations or simulated work situations were most often used in the research. Moreover, published papers tended to be authored by business and psychology professors, with the result that public management issues around diversity were not well addressed. Starting in the 1990s, however, the authors found that scholarly research on diversity began to appear regularly in highly ranked public policy and public personnel journals.

CLARIFYING THE NATURE OF DIVERSITY

There is no single common usage of the concept of diversity in the organizational literature.^[7,8] Simple definitions of diversity most often equate it with variety. Diversity can be studied at a societal level or within particular institutions. Applied to organizational life, there are at least three categories of diversity

related to performance and strategy. These are structural or functional diversity (differences based on organizational functions and tasks such as administrative vs. operational), business diversity (differences in markets, products, and services), and workforce diversity (different types of employees).^[9] At the most general level, Gentile^[2] suggested that in the United States, diversity has become a code word to discuss issues related to how various types of difference impact individuals' and groups' life experiences.

Most of the diversity management literature focuses on workforce diversity.^[8,10] Workforce characteristics can range from attributes directly related to work or tasks such as differences in skills, to those that are social in nature and—in theory—only indirectly related to work, such as gender. There are several types of workforce diversity. Occupational diversity (e.g., locksmiths vs. teachers), professional diversity (e.g., nurses aides vs. registered nurses), and social diversity. The latter refers to variations in the characteristics that identify a person with a social category (e.g., male vs. female). The interdependencies among types of workforce diversity create situations that must be managed (e.g., the experience of a male doctor is different from that of either a female doctor or a male nurse).

The organizational challenges of diversity are based on the existence of documented patterns of exclusion from society's resources along social identity lines. The existence of clusters of jobs occupied by individuals with similar social traits often unrelated to the job, which experts call job segregation, or clusters of persons in occupations and professions who belong to the same identity group, which is called occupational segregation, are well documented.^[11,12] These realities produce differential opportunity in processes and outcomes for members of different identity groups, and thus reproduce inequality.^[13]

SOCIAL DIVERSITY

Research suggests that the salience of an attribute is often socially and historically constructed.^[7,13,14] Social markers include gender, race, ethnicity, religion, sexual orientation, physical ability, and age, as well as family, economic, educational, and geographic backgrounds and status. Some scholars distinguish visible from invisible social dimensions;^[15] others also consider behavioral attributes associated with functional diversity, such as differences in learning, communication, and work styles. Combined, these social attributes produce the "social types" that make up an organization's workforce,^[4] and determine its particular diversity challenges.

For several reasons, social diversity offers the hardest challenges to managers. Social identity

involves both self-definition and attributes perceived by others.^[1] Also, categorizing individuals on the base of a single identity attribute is usually considered detrimental because the parts of peoples' identities occur simultaneously (e.g., a "typical" Latino employee may be a member of certain religion, could be gay or straight, or may have a disability not readily apparent to the naked eye). Tough diversity dilemmas emerge from these social dynamics, and from the interactions between social and functionally related identities (e.g., occupation or position in the organization).

The empirical literature to date does not offer much assistance on dealing with multiple identities. Wise and Tschirhart^[6] concluded that findings about the impact or management of diversity within one group of employees are not generalizable to other groups. They also find that diversity related to educational attainment, work experiences, and tenure with an organization were less studied, and disability was largely absent, as compared with studies of sex, ethnicity/race, and age.

BROAD AND NARROW DIVERSITY MANAGEMENT STRATEGIES

Two distinct approaches have characterized the theory and practice of managing workforce diversity. In the world of practice, a broader definition of workforce diversity emphasizes differences in personality, work style, and visible manifestations of social diversity. In contrast, a narrower definition focuses on social exclusion of what the United States law calls "protected classes," that is, women, racial and ethnic minorities, persons with disabilities (and in some jurisdictions, but not all, gay, lesbian, bisexual, and transgender individuals).

The broader approach places diversity within a context that expands beyond representation and compliance of the law. The narrower approach focuses on efforts to redress existing patterns of inequality and to reduce disparities in the future. Thomas (1991) argued that both approaches are necessary: without rigorous "narrow" efforts to recruit and promote persons not part of the dominant social group, no "new faces" would enter the organization or managerial ranks. However, without a "broad" organizational focus to make the organization a welcoming place to all, the original effort would go to waste. In this sense, the broad and narrow approaches are complementary.

The distinction also appears in the diversity management literature. Work focusing on the practice of managing diversity is normative and emphasizes the broad definition. Empirical research on issues of diversity in the workplace tends to concentrate on specific

dimensions that suggest a narrow approach, such as examining the representation of specific populations or the impact of diverse membership on performance. The distinction stems from the historical development of diversity efforts, from ideological battles fought in that process, and from the way research has followed practice in this area.

In the United States, following the Civil Rights Act of 1964 and the Equal Employment Opportunity Act of 1972, diversity in the workplace was originally measured by counting the number of employees that fit into each of the desired categories. As a result, diversity management is frequently confused with affirmative action. However, since the mid-1960s, three distinct phases of diversity management have evolved. First, affirmative action and equal employment opportunity (EEO) employment policies sought to increase the representation of minorities and women in many areas of employment and to reduce discriminatory practices. Then, as persons from different backgrounds entered, organizations started to try to change their workplace culture via awareness courses and celebrations of diverse characteristics. These activities, known as "valuing diversity," were often isolated from organizational strategy or job needs, however. Later, a new type of effort, known as "managing diversity," started to link changes in work practices to workforce diversity. These new approaches combine solving organizational problems and addressing justice concerns. This points to the need to consider both the ethical and practical drives underlying diversity management. In theory, the phases occur sequentially within the life cycles of organizations, but in practice most organizations will use parts of all three phases simultaneously.

In a final iteration of this development, "diversity management" has become the preferred term, mixing aspects of the narrow and broad approaches. Gilbert et al.^[16] defined diversity management as a managerial principle used to make HR decisions and implement practices that create greater inclusion of all individuals into both formal programs and informal social networks. Meyerson and Fletcher^[17] referred to diversity management as a "persistent campaign of incremental changes that discover and destroy the deeply embedded roots of discrimination" (see Ref.^[17], p. 131).

ETHICAL AND PRACTICAL MOTIVATIONS FOR DIVERSITY MANAGEMENT

The diversity literature highlights ways in which diversity makes good business sense. Public sector scholars promoting the value of diversity have followed this lead under the assumption that managers take pragmatic arguments more seriously than ethical arguments.

It is true that the costs of not attending to workforce diversity can be high. Unresolved diversity problems can deflect employee energy and attention from performance toward justice and fairness issues with a consequent loss of productivity. Poor diversity management can cause the organization to lose good employees and therefore incur the expenses of recruiting and training their replacements, as well as potentially diverting financial and human resources to deal with litigation or to pay punitive damages.^[4,18] But there are also compelling reasons to stress the ethical underpinnings of diversity.

Diversity management is based on the normative assumption that diversity is a value, and that its management requires a commitment to social inclusion and therefore equality. In this sense, diversity refers to the degree to which there is a variety of attributes (including social, cultural, functional) within a particular structure of governance (be it organizational, sectorial, societal), and the extent to which this variety is distributed at all levels of the structure. Equality, in turn, refers to the extent to which opportunities, resources, information, and power are distributed across diverse individuals and groups within a particular structure of governance. Lack of social diversity and patterns of organizational inequality usually go hand in hand.

Gilbert et al.^[16] highlighted the relevance of ethics in diversity by reviewing three ethical principles that support successful diversity initiatives. The first principle is the *Golden Rule*: if you want to be treated fairly, treat others fairly. The second is the *Disclosure Rule*: you must be comfortable with decisions after asking whether you would mind if others became aware of them. The third is the *Rights Approach*, which assumes that people should have the ability to freely choose what they will do with their lives.

Even when the motivation to address the diversity challenge is not ethical, these ethical principles contribute to its successful implementation. Abiding by the first rule, successful diversity programs are inclusive and provide fair treatment to all employees. The openness needed to administer diversity programs responds to the second rule. Diversity management addresses the third rule by allowing people to choose opportunities according to their interests and abilities. If these diversity principles are removed, Gilbert et al.^[16] argued, diversity initiatives will collapse. Critics of diversity management may argue that achieving diversity requires a trade-off between equity, merit, and diversity. Such trade-offs, however, contradict theoretical definitions of diversity. Merit is not at question if attributes of the job and staffing practices are not limited by societal expectations or stereotypes. Equality means that every one qualified to do the job is allowed to do it.

Managers implementing diversity initiatives need to expect some friction even under the most favorable circumstances. The performance and representation requirements for increased diversity in the public workplace generate interpersonal, organizational, and political challenges that require direct managerial attention. Williams and O'Reilly^[7] argued that although diversity increases the opportunity for creativity and the quality of group work, it also increases the likelihood of group conflict, member dissatisfaction, turnover, and failure in the implementation of ideas. Hence, as a “double-edge sword” diversity “requires careful and sustained attention to be a positive force in enhancing performance” (see Ref.^[17], p. 120). It requires careful diagnosis and designing an agenda that is part of a broader framework to enhance organizational performance.

DIVERSITY MANAGEMENT FRAMEWORKS

Proponents of diversity management suggest that organizations must be able to realistically diagnose the status of their organization to determine the right strategies for diversity management. Cox^[1] developed a helpful typology to assess an organization's distance from achieving the goal of diversity. The “monolithic organization” comprises mostly similar employees, and the culture rewards only those who conform to the norms of the dominant group. In such an organization, a person with a language accent may not be promoted as fast as others providing similar input and ideas into the organization's processes. Next, “the plural organization” has a mixed group of employees, but the systems and culture remain dominated by one group. Women who have advanced through the bureaucracy but receive fewer promotions at the higher levels may have hit a “glass ceiling” because predominantly masculine features such as assertiveness continue to be seen as the necessary values for success. In all monolithic and some plural organizations, informal communications, networks, and key decision-making bodies are closed to non-dominant employees. Cox's third organizational type is the “multicultural” organization, in which differences are viewed as potential organizational assets. In these organizations, inclusion and fairness are important values, and this is reflected in the demographic composition of the workforce at all levels.

Locating the organization along the “diversity continuum” (between discriminatory and non-discriminatory, from monolithic, to plural, to multicultural) can help managers to identify methods to pursue their diversity strategy. Ospina^[4] proposed that organizations must differentiate among four consecutive managerial tasks: *considering, pursuing, managing,*

and *maximizing* diversity, depending on the organization's place in the diversity continuum. To overcome the monolithic organization, managers first start by *considering* diversity, that is becoming aware of the costs and benefits of maintaining a monolithic workforce. In *pursuing* diversity, they look for strategies to attract and recruit a diverse workforce, usually starting with EEO initiatives. Once the organization opens up to a more diverse workforce, becoming more plural in nature, leaders find ways to ensure that policies and practices support all employees, thereby *managing* diversity. Finally, managers use diversity strategically to add value to the organization strategic goals by supporting the unique contributions each organizational member of their multicultural workforce brings, independent of their background, thus *maximizing* diversity. As an organization achieves multiculturalism, diversity management also requires looping constantly among these tasks over time, so that the diversity challenge never stops.

In their diversity management framework, Thomas and Ely^[19] differentiated existing approaches to diversity according to the underlying philosophy of the managers involved. Managers pursuing efforts based exclusively in a diversity philosophy seeking proportional representation expect all employees to assimilate to the dominant culture (Thomas and Ely call this the *discrimination and fairness* paradigm). Other managers pursue efforts that acknowledge the strategic function of diversity to target minority consumer groups. In this approach, diverse employees are not allowed to integrate their unique features to the larger organization, but instead are valued only because of their potential to interact effectively with clients of similar backgrounds (Thomas and Ely call this the *access and legitimacy* paradigm). Finally, in organizations where the *learning and efficiency* diversity paradigm is espoused, employees are not pegged to market niches, but are encouraged to use their diverse backgrounds to enhance productivity and to create new opportunities for the organization. These organizations are said to value innovation, creativity, and diversity in problem solving and decision making, independent of each employee's social identity.

Cox's organizational typology, Ospina's developmental tasks, and Thomas and Ely's philosophies represent valuable frameworks to provide guidance to managers in implementing a diversity agenda. The underlying assumption is that diversity is an organizational imperative, not just a managerial choice. For this reason, the responsibility for diversity management must, in a large organization, be shared. Program managers are responsible for developing and implementing the goals of the organization. Personnel managers, in turn, are responsible for the design and maintenance of HR systems that support both the

successful execution of the organization's mission and programs that sustain a diversity philosophy. In partnership, personnel and program managers ensure that the diversity challenge is woven into the organizational goals, its systems and functions, and overall managerial decisions, so that eventually diversity becomes embedded in all practices and routines of the organization.^[18]

DIVERSITY, PUBLIC PERSONNEL MANAGEMENT, AND CIVIL SERVICE

The functional and political requirements for diversity in public agencies take place in a climate where many are questioning the traditional institution of civil service. Originally implemented to increase rationality in public employment, today civil service is often viewed as an obstacle to attaining the flexibility required for organizational adaptation. In addition, some argue that many systems have failed to address problems of social exclusion, political favoritism, and lack of social representativeness in public service.^[13] Although there is no consensus around these claims, there is a generalized call for reforming public employment institutions to ensure they accomplish their role in a democratic society.^[20] Diversity management must be included in any discussion of civil service reform.

Although civil service systems can provide general guidelines to ensure diversity, the specifics of diversity management fall within the purview of each agency because every organization is unique. Ultimately, diversity management in the public sector, as in any organization, is a complex and multidimensional challenge that requires permanent and focused managerial attention. Some civil services do not function well (e.g., where patronage and corruption still predominate in spite of the system), whereas in others the system is functional but has produced exclusionary practices and outcomes. Moreover, in civil service contexts, diversity issues may complicate the managerial requirements to balance contradictory pressures for employee protection and stability required for professionalizing public employment, and demands for accountability and transparency to make public service more efficient.

In this context, a thorough diagnosis to assess workforce diversity problems includes an analysis of the system as a whole, as well as audits of each agency the system regulates. Relevant questions include the extent to which a civil service system promotes both effectiveness and multiculturalism, diversity of the workforce in the entire jurisdiction and its component agencies, and the trade-offs associated with introducing system-wide strategies to promote multiculturalism versus incremental changes to address agency-level diversity problems. Answering these questions and designing a

reform strategy are tasks that require collaboration among actors of the several jurisdictions regulating public employment and managers who need to make decisions to satisfy performance requirements at the agency level.

Ultimately, all stakeholders must strive to design a system of public employment where:

- Opportunities in hiring, compensation, promotion, and personal development are available to all employees across jobs and levels.
- Access to information and networks in an organization are available to all employees across jobs and levels.
- Every employee believes they are treated as a unique individual whose multiple identities and abilities are respected and appreciated for their potential contributions to the organization. In turn, employees see each other and themselves as valuable members of a working community, while still comfortable with and proud of their identities.
- The right policies, systems, and processes exist to ensure the agency's ability to attract, retain, and develop employees with diverse backgrounds.
- Members of the community can see themselves reflected and represented in the workforce that serves them.

In this context, the envisioned workplace is characterized by an organizational climate and an HR management system in which employee diversity becomes a "normal" condition of organizational life.

DIVERSITY AS A WIDER ISSUE FOR PUBLIC ADMINISTRATION AND PUBLIC POLICY

The list of topics to be included in a full discussion of diversity in public administration and public policy is constantly expanding and cannot be limited to workforce considerations alone. The requirement for diversity in public organizations is rooted in both efficiency and the democratic values of responsiveness and representation in public administration. Public agencies have the mandate to represent in their workforce a wide variety of citizens, as well as consider the plurality of values, concerns, and voices of the larger population. Social movements have challenged traditional patterns of exclusion in the most important institutions of society, and it is no surprise that the same happens in work institutions and public employment.

Although the diversity literature has primarily focused on workforce diversity and diversity management, these topics are not isolated from broader discussions about topics such as multiculturalism and participatory democracy. Scholars of deliberative

democracy seek to acknowledge and incorporate different viewpoints in public policy formation. The need for broad representation in democratic decision making and policy implementation makes diverse participation in the policy process a priority for these authors.^[21] Stafford^[22] went even further, arguing that the assumptions and analytical tools used in policy analysis are the product of historical and ongoing processes that exclude certain groups of society, and are therefore not representative enough to serve as the foundation for truly democratic policy decisions. Participatory policy making therefore implies the need for managing diversity among the citizenry and not just the workforce.

If public organizations are viewed as part of the implementation stage of public policy, both workforce variety and diversity in decision making are of concern to all interested is a democratic society. Writers from all points of the political spectrum invoke diversity considerations in most contexts of governance. No current scholar presently forecasts the end of this discussion.

REFERENCES

1. Cox, T., Jr. *Cultural Diversity in Organizations: Theory, Research and Practice*; Berrett-Koehler: San Francisco, CA, 1993.
2. Gentile, M. *Managerial Excellence Through Diversity: Text and Cases*; Waveland Press: Prospect Hills, IL, 1996.
3. Mathews, A. Diversity: A principle of human resource management. *Public Pers. Manage.* **1998**, 27 (2), 175–185.
4. Ospina, S. Realizing the promise of diversity. In *Handbook of Public Administration*; Perry, J., Ed.; Jossey Bass: San Francisco, CA, 1996; 441–459.
5. Berman, E.M.; Bowman, J.S.; West, J.P.; Van Wart, M. *Human Resource Management in Public Service: Paradoxes, Processes and Problems*; Sage Publications, Inc.: Thousand Oaks, CA, 2001.
6. Wise, L.R.; Tschirhart, M. Examining empirical evidence on diversity effects: How useful is diversity research for public sector managers. *Public Adm. Rev.* **2000**, 60 (5), 386–394.
7. Williams, K.Y.; O'Reilly, C.A., III. Demography and diversity in organizations: A review of 40 years of research. In *Research in Organizational Behavior*; JAI Press Inc.: Greenwich, CT, 1998; Vol.2, 77–140.
8. Schneider, S.K.; Northcraft, G.B. Three social dilemmas of workplace diversity in organizations: A social identity perspective. *Human Relat.* **1999**, 52 (11), 1445–1467.
9. DeLuca, J.; McDowell, R. Managing diversity: A strategic "grass roots" approach. In *Diversity in the Workplace: Human Resources Initiatives*; Jackson, S.E., et al., Eds.; Guilford Press: New York, 1992; Vol.1, 227–247.
10. Jackson, S.; Stone, V.; Alvarez, E. Socialization amidst diversity: impact of demographics on work team old-timers and newcomers. In *Research in Organizational*

- Behavior*; Cummings, L., Straw, B., Eds.; JAI Press: Greenwich, CT, 1999; Vol.15.
11. Tomaskovic-Devey, D. *Gender and Racial Inequality at Work: The Sources and Consequences of Job Segregation*; ILR Press: Ithaca, NY, 1993.
 12. Guy, M. *Women and Men of the States: Public Administrators at the State Level*; M.E. Sharpe, Inc.: New York, 1992.
 13. Ospina, S. *Illusions of Opportunity: Employee Expectations and Workplace Inequality*; Cornell University Press: Ithaca, NY, 1996.
 14. Williams, R.E. *Hierarchical Structures and Social Value: The Creation of Black and Irish Identities in the United States*; Cambridge University Press: New York, 1990.
 15. Mor Barak, M.E. Beyond affirmative action: Toward a model of diversity and organizational inclusion. *Adm. Soc. Work* **2000**, *23* (3–4), 47–68.
 16. Gilbert, J.A.; Stead, B.A.; Ivancevich, J.M. Diversity management: A new organizational paradigm. *J. Bus. Ethics* **1999**, *21* (1), 61–76.
 17. Meyerson, D.E.; Fletcher, J.K. A modest manifesto for shattering the glass ceiling. *Harvard Bus. Rev.* **2000**, *78* (1), 126–136.
 18. Ospina, S.; O'Sullivan, J.F. Working together: meeting the challenges of workforce diversity. In *Public Personnel Administration: Problems and Prospects*, 4th Ed.; Hays, S., Kearney, R., Eds.; Prentice Hall: Englewood Cliffs, NJ, 2003; 238–255.
 19. Thomas, D.A.; Ely, R.J. Making differences matter: A new paradigm for managing diversity. *Harvard Bus. Rev.* **1996**, *74* (5), 79–90.
 20. Klinger, D.; Lynn, D. Beyond civil service: The changing face of public personnel management. *Public Pers. Manage.* **1997**, *26* (2), 157–173.
 21. Weeks, E. The practice of deliberative democracy: Results from four large-scale trials. *Public Adm. Rev.* **2000**, *60* (4), 360–372.
 22. Stafford, W. In *Multiculturalism and Public Administration*, Presentation at the National Association of Schools of Public Affairs and Administration Annual Conference, Raleigh, NC, October 17, 1997.

Economic Development and Organization

C. Fred Baughman

*Department of Public Administration, University of Central Florida–Daytona Beach,
Daytona Beach, Florida, U.S.A.*

INTRODUCTION

A new trend within the economic development profession is the reconsideration of the organizational form of local economic development agencies. Recently, the Fort Wayne–Allen County Economic Development Alliance was established as a partnership between the City of Fort Wayne, Allen County, and the Greater Fort Wayne Chamber of Commerce. Before its creation, each of the three entities had its own economic development staff.^[1] This is an example, therefore, of governmental and private sector agencies combining to form a public/private partnership.

However, not all recent reorganizations have been in that direction. The local elected officials in Volusia County, Florida, withdrew their financial support from the existing public/private partnership (Enterprise Volusia, Inc.). Instead, Volusia County is going to create an in-house economic development department. Because of concern by the private sector in regards to the public sector's ability to accomplish effective business recruitment, three of the local chambers of commerce within Volusia County have also formed an alliance to perform business attraction activities. This is an example, consequently, of a public/private partnership being split into separate governmental and private sector economic development entities.

HISTORICAL PERSPECTIVE

Historically, the source of funding has played an important role in the type of economic development activities within a community. Due to federal funding, many local planning offices were involved in the preparation of economic development planning documents in the late 1970s. Economic development committees were established to prepare Overall Economic Development Programs (OEDP) funded by the U.S. Economic Development Administration. In addition, again because of the availability of federal funding, many local governments were involved in urban renewal efforts that improved the real estate of communities.

However, the actual implementation of business recruitment efforts, in order to create new local jobs, was normally the role of the chamber of commerce

(private sector). Other non-profit organizations were formed to take advantage of programs from the U.S. Small Business Administration. Communities began, therefore, to form revolving loan programs as incentives for development, oftentimes targeted to minority- and women-owned businesses.

As federal funds became scarce for economic development activities, the trend was toward the creation of public/private partnerships, where both entities have a stake in local economic development efforts. The public and private sectors would provide funding for the economic development activities, however, the percentage of the contribution from the sectors was different in various organizations.

Currently, local economic development organizations can be divided into three categories: public sector agencies, private sector organizations, and public/private partnerships.^[2] No matter which form of economic development entity is established, each will have its own advantages and disadvantages. The local community must decide which form fits best with local needs.

PUBLIC SECTOR AGENCIES

Some communities believe that economic development is most effective when accomplished by an agency that is an integral part of local government (city or county). The public sector agency model of economic development organization is unique in that it is funded and directed entirely by the public sector. The local elected politicians provide the policy direction for economic development. In addition, through the annual budget process, they appropriate the available funds for economic development efforts. The economic development staff that carries out the required activities is composed of government employees that report directly to the local elected officials.

Advantages of the Public Sector Agency

Secure source of funding

There is a more secure source of funding for economic development efforts. This source might come from the

general fund of the local government. Other communities have established impact fees to support economic development. The State of Indiana allows its counties to create an Economic Development Income Tax (EDIT). No matter what the source of the funds, economic development staffs in governmental agencies are not expected to be involved in fund-raising activities.

Proximity to decision makers

Most financial incentives in economic development projects come from the public sector. If the economic development organization is a governmental agency, then it is much closer to the decision makers that will commit those incentive dollars.

Interdepartmental cooperation

Similar to being closer to incentive decision makers, the staffs of public sector economic development agencies work closely with other governmental departments. Many times, an economic development deal will require the cooperation of other departments for permit approvals or infrastructure improvements.

Citizen participation

Another advantage of public sector economic development agencies is the legal requirement for citizen participation. This requirement ensures that the economic development staff is not acting in a manner contrary to the overall vision of the community.

Disadvantages of the Public Sector Agency

Timing of economic development projects

Many local officials are elected to two-year terms. Consequently, those local elected officials focus on programs with quick results. However, economic development is not an activity that normally produces benefits in a fast time frame. From the time that an economic development staff begins talking to a business prospect until that business makes a locational decision can typically be from two to four years. Therefore, local elected officials creating an economic development agency, and making annual budget appropriations, might not see a successful economic development project during their term in office. It is commonly believed that local elected officials often-times support those activities that get them reelected. David Osborne and Peter Plastrik, in their book, *Banishing Bureaucracy*, indicate that leaders of the future are going to have to be committed, “willing to stick it out...willing to invest the time it takes to

communicate their vision; and if they can prove themselves...they can succeed.”^[3] There is no guarantee that economic development efforts can produce desired results in the short run—elected leaders are going to have to take a long-term perspective.

Political infighting

Another disadvantage of public sector economic development agencies is the local infighting between political organizations. For example, if the leadership of a local city is politically different than the leadership of the county, then there might not be the cooperation required for a successful economic development program by either entity. In addition, the infighting within a political organization can have just as serious consequences. The lack of political leadership can be a large disadvantage for a public sector economic development agency.

Inefficient government structure

An inefficient government structure can also be a disadvantage. As stated earlier, an advantage of the public sector agency was its closeness to decision makers. However, if the organizational structure of the government does not allow for this proximity, then the advantage is no longer realized. Therefore, economic development staffs should not be buried in an umbrella department. In addition, it should be recognized that the motivating factors of planning staffs and economic development staffs are very different. While both groups should pursue smart and managed growth, the economic development staff will normally be promoting growth, and the planning staff will normally be constraining growth.

Governmental bureaucracy can be a limiting factor to the success of economic development efforts. The saying “time is money” is never more true than in the area of economic development. Business prospects expect, and require, the immediate turnaround of information and decisions. Any delay in the opening of a new business is an opportunity cost to that company.

Public disclosure requirements

A further disadvantage of a public sector economic development agency is public disclosure requirements. These are commonly known as “Sunshine Laws.” The overall concept of sunshine laws is positive in its belief that the activities of government should be conducted in the public view. However, most business prospects demand confidentiality during the early stages of the economic development process. For labor and market reasons, it is important for them that information of their potential relocation does not become public.

The sunshine laws place a public sector economic development agency in an uncomfortable position.

Ability to attract professional staff

The final disadvantage of the public sector agency is its ability to attract and retain competent economic development professionals. This constraint might be because of the pay level of other city staff members or just of the legal structure of the human resources procedure required of governmental entities.

PRIVATE SECTOR ENTITIES

As stated above, early in the formation of the economic development profession, most business recruitment activities were done by the local chamber of commerce. In some communities, the chamber of commerce is still the focal point for economic development. The private sector entities obtain their leadership, agenda, and funding entirely from private enterprise. Oftentimes, this leadership is composed of real estate brokers, attorneys, bankers, accountants, and small business owners. Not surprisingly, these individuals see a direct potential business benefit from successful economic development efforts.

Advantages of the Private Sector Entities

Freedom from governmental legal constraints

The first advantage of the private sector organization is its freedom and the fact that it is not constrained by the many legal structures of local government. Those constraints might be procurement procedures, personnel regulations, sunshine laws, and others. Another related advantage is the private sector's ability to recruit and compensate talented economic development professionals.

Not part of the government organization

By being outside the governmental structure, the private sector organization can often act as an ombudsman during the economic development process. If a project gets held up in government "red-tape," then the private sector organization can use steps to remedy the problem without worrying about hurting the feelings of government employees. This type of organizational entity does not require the government to "police" itself.

Able to respond quickly

As stated above, timing is often critical in economic development activities. In many cases, the private

sector will respond more quickly and is more efficient than the public sector. As Osborne and Plastrik state, "The problem with government monopolies is that they have no real reason to improve their performance."^[4]

Disadvantages of the Private Sector Entities

Responsiveness to community sentiment

One of the disadvantages of the private sector organization is that it can be less responsive to the community sentiment for economic development.

Conflicts of Interest

Another disadvantage is that the private sector leadership might have a conflict of interest when it comes to economic development projects. Even though a project might be a benefit and create many well-paid jobs for the community, some private sector leaders might not support the project because of the competition it creates for their own labor force needs.

Secure funding sources

The final disadvantage of the private sector economic development organization is that the funding stream is less secure. Rather than a dedicated source of funds, most private sector organizations rely upon fundraising for their budgetary needs. Consequently, a staff hired for economic development expertise can spend a considerable amount of time instead on organizational fund-raising activities.

PUBLIC/PRIVATE PARTNERSHIP

The public/private partnership is the form of organizational structure most recently preferred by many communities for economic development activities, because it can combine the advantages of the public sector agency and the private sector organization. As the name implies, the public/private partnership is a collaborative effort of the public and private sectors. It is estimated that over 80% of current organizations have a board of directors comprised of members from the public (government) and private sectors.^[5]

Advantages of the Public/Private Partnership

Vehicle for communication

One of the main advantages of the public/private partnership is that it provides a vehicle for the interaction

between local government officials and leaders in the private sector. This may be one of the few opportunities within the community for this type of interaction to occur. Consequently, the public/private partnership has the opportunity to maximize the powers of the public and private sectors.

Broader base of funding

Economic development activities are becoming more expensive. Funding of the public/private partnership normally comes from both sectors but can be contributed at different percentages. The specific strengths and weaknesses of the local community will assist in establishing that ratio for the organization. However, less time will be spent on fund-raising than in a fully private sector organization.

Need for confidentiality

As stated earlier, the economic development prospect often requires total confidentiality in regards to its interest in a potential relocation. Because of "sunshine laws," the government entities are not able to provide this needed service. As a result, the prospect will many times communicate with a public/private partnership much earlier in the process than it would with a strictly public sector organization.

Disadvantages of the Public/Private Partnership

Multiple sources of responsibility

The primary disadvantage of the public/private partnership is that the economic development staff may be forced into a position of reporting to multiple "bosses."

Project coordination

Another disadvantage is that unless all the entities involved in the partnership trust the process, there may be charges of leads being steered to certain locations. The professional staff will need to establish a lead-sharing procedure and rely heavily upon it.

CONCLUSION

As stated earlier, every type of organizational structure will have advantages and disadvantages. Again, the individual strengths and weaknesses of every community will be important in determining the economic development organizational structure that is most effective. A recent article comparing the St. Paul model

(public sector agency) to the Philadelphia model (public/private partnership) in conclusion stated, "We think the quasi-public corporation, with its legal flexibility and combination of public and private resources, offers the broadest range of powers and the strongest opportunity for coordination between the public and private sectors. Although independent to some degree, it also recognizes its responsibilities to both sides—to the private sector in creating a favorable business climate and to the public sector in creating job opportunities and improving the city's tax base."^[6]

In general, therefore, the trend is toward the public/private partnership because of its ability to contribute the advantages of the public sector agency and the private sector organization while minimizing the disadvantages of both. Dr. David Kolzow recently concluded, "the advantages of teamwork in the local economic development process are becoming increasingly apparent in the face of changing economic trends, budget constraints, and the growing need for the combined resources of the private and public sectors. Communities in which the public and private sector band together to form economic development teams through an organization structured as a public-private partnership are more likely to experience effective actions and successful results."^[7]

REFERENCES

1. <http://theallianceonline.com/index2.htm> (accessed August 2001).
2. Whitehead, W.; Ady, R. *Organizational Models for Economic Development, Practicing Economic Development*, 3rd Ed.; Koepke, R., Ed.; American Economic Development Council Education Foundation: Rosemont, IL, 1996; 41.
3. Osborne, D.; Plastrik, P. The culture strategy. In *Banishing Bureaucracy: The Five Strategies for Reinventing Government*; A Plume Book, a member of Penguin Putnam, Inc.: New York, NY, 1998; 296 pp.
4. Osborne, D.; Plastrik, P. The consequences strategy. In *Banishing Bureaucracy: The Five Strategies for Reinventing Government*; A Plume Book, a member of Penguin Putnam, Inc.: New York, NY, 1998; 130 pp.
5. Whitehead, W.; Ady, R. *Organizational Models for Economic Development. Practicing Economic Development*, 3rd Ed.; Koepke, R., Ed.; American Economic Development Council Education Foundation: Rosemont, IL, 1996; 41–42.
6. Knack, R.; Bellus, J.; Adell, P. Setting up shop for economic development. In *Economic Development in Local Government*; Kemp, R., Ed.; McFarland & Company, Inc.: Jefferson, NC, 1995; 44 pp.
7. Kolzow, D. *Public/Private Partnership: The Economic Development Organization of the 90s. Practicing Economic Development*, 3rd Ed.; Koepke, R., Ed.; American Economic Development Council Education Foundation: Rosemont, IL, 1996; 47 pp.

Economic Development, Citizen Participation and

Terry F. Buss

Center for International Affairs, National Academy of Public Administration,
Washington, District of Columbia, U.S.A.

F. Stevens Redburn

U.S. Office of Management and Budget, Executive Office of the President,
Washington, District of Columbia, U.S.A.

Marcela Tribble

School of Policy and Management, Florida International University, Miami, Florida, U.S.A.

INTRODUCTION

The 1990s witnessed an increased interest among policy makers, scholars, and advocates in expanding and deepening citizen participation processes, particularly in community and economic development activities. The Bush Administration, early in its tenure, philosophized about the idea of a “citizen-centric” government. The Clinton Administration, under the leadership of Vice President Al Gore, worked on numerous citizen participation initiatives as part of its “reinventing government” program. Scholars, most notably, Robert Putnam, in books such as *Bowling Alone*,^[1] called attention to the decline in civil society. Redburn and Buss, in their monograph, *Modernizing Democracy*, called attention to the power of new information technology and the Internet to engage citizens in public life in more sophisticated ways, and outlined a program to accomplish this goal. Advocates, such as the Orton Family Foundation, have invested heavily in the development and marketing of software—*CommunityViz*—to improve the quality of citizen deliberations on community and economic development policy and programming (www.communityviz.com). Representatives from neighborhood groups, the planning profession foundations, think tanks, and universities met in Tampa in January 2002 to form a national association to raise the visibility of and expand opportunities for citizen participation in building communities (www.PlaceMatters.com). Hundreds of web sites on citizen participation now dot the Internet landscape (e.g., www.democracyinnovations.org).

This article first examines reasons why many want to deepen and expand citizen participation processes. Next, it looks at mandates requiring citizen participation processes at the local level, especially those that are part of public programs for economic and community development. Then the article discusses different modes of citizen participation, evaluating

the advantages and disadvantages of each. The entry concludes with an overview of the issues in the design and administration of citizen participation processes intended to support community and economic development.

WHY DEEPEN AND EXPAND CITIZEN PARTICIPATION?

Influential Movements

Thomas Jefferson once pointed out that if the people appeared not enlightened enough to exercise their control of government, the solution was not to take away the control but to “inform their discretion by education.” The cooperative processes that are springing up around the country are doing just that, giving to large numbers of citizens a new comprehension of the complexity involved in government decisions, out of which has got to come a heightened appreciation of, and tolerance for, the necessary work of government. If these processes work, if they spread, if they become an indispensable part of government at all levels, we may take it as a sign that we, as a people, have moved up a grade in democracy’s school. It holds out the hope that, eventually, the United States will be ready for self-government.

—William Ruckelshaus,
Restoring Public Trust in Government,
National Academy of Public Administration,
November 15, 1996

Political theorists—Robert Dahl, Charles Lindblom, Mancur Olson, William Riker, Anthony Downs, and Alexis D’Tocqville, to name a few—have long been interested in citizen participation as a requisite for democracy. There is no single source for the explosive rise in interest in citizen participation, especially in

community and economic development. But we suspect that five independent, although interrelated, forces are at work: a renewed interest in democracy, likely arising out of efforts to assist former Soviet Bloc countries in their transition away from communism; a concerted effort, now worldwide, to “devolve” government as close to the people as possible; the “reinventing government” movement, almost single-handedly launched by David Osborne in his books, *Laboratories of Democracy*^[2] and *Reinventing Government*;^[3] anti-corruption and civil society movements promoted by multilateral and bilateral aid organizations in the Third World; and interest in collaborative shared decision-making models of government inspired separately in business management and public administration leadership literatures (e.g., Refs. ^[4] and ^[5]).

American experts, helping to rebuild societies in Russia, its former republics, and Eastern Europe, not only engaged in serious thinking about how to build democracy and economic reconstruction in these countries, but they also called into question how democratic countries in the West really were. The Open Society Institute, based in Budapest, Hungary, and funded by the Soros Foundation, is a think tank devoted to issues of democracy, particularly citizen involvement in redirecting local economies Eastern Europe and Russia (see www.osi.hu). The World Bank, United Nations, and other multilaterals, not to mention efforts across the United States, have extensive programs to decentralize government, devolving power to local people. Micro-credit programs targeted at poor neighborhoods and villages as a community/economic development tool is but one example.^[6] Anticorruption is a major thrust of World Bank development policy. The World Bank Institute offers a technical assistance and training program—specializing in fiscal decentralization and economic development—to empower local decision making on public expenditures (see www.worldbank.org). The reinventing government movement, a descendent of past “good government” and reform movements, is a focus of a National Academy of Public Administration’s Alliance for Redesigning Government in its efforts to build capacity for community development (see www.alliance.napawash.org). The Pew Trust has in place a nationwide project, the Civic Entrepreneur Initiative, to promote local leadership (see www.pew-partnership.org).

Goals

The rationales offered by these five sources of new interest in citizen participation illustrate the diversity of motives and perspectives from which it can be argued that citizen participation processes ought to be expanded and deepened.

Making democracy more democratic

In democratic theory, popular sovereignty refers to the capacity of a people for independence and self-government. Modern American democracy boasts, according to Schattschneider, only a “semisovereign” people, dependent on an expansive, paternalistic political elite intent on reducing deliberative forms of participation.^[7] Expanding and deepening citizen participation is a way to make democracy more democratic—that is, direct, deliberative, and participatory. As Thomas Jefferson so aptly stated, “[there is] no safer depository of the ultimate power of society but the people themselves.”

Redefining power structures

For some, increasing citizen participation among selected underrepresented groups is a way to redistribute power.^[8] The Administrative Procedure Act of 1946 required federal agencies to submit their proposed regulations to public review for the first time. The Housing Act of 1954 expanded participation for the general public by requiring agencies to appoint citizen advisory committees as a prelude to undertaking community development activities. In the 1960s, civil rights legislation enfranchised many Black voters and struck down legal barriers to participation. The Johnson Administration’s “War on Poverty” programs under the Equal Opportunity Act of 1964, at the same time, briefly challenged many existing power structures, supporting emergent local organizations with federal funds in an effort to create “maximum feasible participation.” Community action agencies, still functioning today, gave many poor people “voice” in local development efforts. The decades following saw the empowerment of groups concerned about local issues, including for our purposes, environmental, planning, zoning, and land use. Citizen participation occurred first through litigation, then through legislation and executive order, requiring nearly all federal agencies to facilitate citizens’ input. Executive Order 12898, *Federal Actions to Address Environmental Justice in Minority Populations and Low-Income Populations*, for example, mandates meaningful participation and access for minorities to resolve issues such as toxic waste dump location decisions.

Enhancing credibility

Many believe that governments that reach out to citizens will likely gain credibility from greater openness, transparency, and responsiveness. When government is credible, citizens will support government programs, projects, or initiatives. Today, much social science research finds a public increasingly alienated and

disconnected from government at all levels. Less than one-half of registered voters have bothered to show up at the polls since the 1960s. According to monthly “tracking polls” conducted by Harvard’s Shorenstein Center on the Press, Politics, and Public Policy, less than one-half of voters, when asked in the months preceding the 2000 campaign, thought about or discussed the campaign. Reengaging alienated citizens is a challenge to those who seek to restore government’s credibility.

Managing conflict and building consensus

If everyone agreed on public issues, there might not be a need for government. Because people do not agree, citizen participation processes may be useful in either building consensus or resolving conflict.^[9–12] The 1980s brought an increased focus on “stakeholders,” and precipitated efforts to get those most affected by public policy to at least be consulted and at most to have a say on outcomes. Analysts soon discovered that involving stakeholders in public decisions was perhaps a necessary condition, but not a sufficient one for developing consensus. Involving stakeholders who disagreed with one another often caused more rather than less problems for public officials. For example, in the U.S. Northwest, farmers wanting to divert water for irrigation, Native Americans desiring to fish depleted waters, environmentalists seeking a primeval environment, and developers promoting development all came into conflict over future economic development of a river region. Where disagreements tended to be intense, as say with environmental issues, new methods for resolving disputes and mediating disagreement were tried. The Environmental Protection Agency (EPA) recognized the need to create an office of conflict resolution.

Feedback

The most sought-after product of citizen participation and the least controversial is the opportunity it affords to allow citizens to provide valuable input on how programs or policies should be crafted, how well they work, and how they might be improved. The notion is that citizens as consumers are best positioned to answer these questions. A major development in this area is the citizen satisfaction survey in which government asks a representative group of citizens formally for feedback on services, in the same way that businesses do. Many community development agencies, for example, periodically survey citizens to gain insight into their needs and expectations.

Accountability

Accountability of public officials for the decisions they make and the services they provide is a major goal of

citizen participation processes. If citizens have a say in government, they can influence decisions more frequently and more specifically than simply voting in elections. Public officials who know that citizens are engaged will likely make better decisions.

MANDATED CITIZEN PARTICIPATION PROCESSES

At the federal, state, and local level, few mandates compel local public officials to engage extensively in promoting and using citizen participation processes. Those mandates that do have the force of law are limited, discretionary, and flexible. Transportation planning in Oregon State and economic and community development planning at the Department of Housing and Urban Development (HUD) are illustrative.

Oregon has a long history of citizen participation mandates.^[13] In the 1970s, Oregon created the Land Conservation and Development Commission (LCDC) specifically to create the “opportunity for citizens to be involved in all phases of the [transportation] planning process” (p. 7). In 1990, LCDC and the Oregon Department of Transportation developed the *Transportation Planning Rule*, which put even more responsibility on agencies to elicit citizen input. Over time, citizens in Oregon have developed high expectations of government at all levels. Many professionals believe that Oregon leads the nation in citizen participation applications.

Department of Housing and Urban Development’s Community Development Block Grant (CDBG) program, a \$4.4-billion flexible funding source for community and economic development projects in 50 states and more than 1000 local “entitlement communities,” administered by the Office of Community Development and Planning (www.hud.gov). The *Code of Federal Regulations* [24 CFR 91.105] requires a strategic plan—“the Consolidated Plan”—that includes a citizen participation component allowing citizens to help in developing, amending and evaluating CDBG projects. People with low and moderate incomes, particularly those living in slums or blighted areas, represent a special focus. Entitlement communities must take “whatever actions are appropriate to encourage the participation of all their citizens, including minorities and non-English-speaking persons, as well as persons with disabilities.” Entitlement communities must hold at least two public hearings annually, one hearing must be scheduled to allow citizens to participate in plan development. Entitlement communities meet publication requirements by publishing a plan summary in a general circulation newspaper and by making copies of the plan available at libraries,

government agencies, and public places. Entitlement communities must follow their citizen participation plans, but amendments are permitted as long as public comment is invited. Citizen participation requirements cannot restrict the responsibility or authority of an entitlement community for the development and execution of its plan.

At the local level, few communities mandate specific citizen participation processes, except for hearings on budgets, land use planning, and the like. Most citizen participation activity is mandated by either federal or state law or regulation.

MODES OF CITIZEN PARTICIPATION

A panoply of options exists to satisfy citizen participation requirements under law or because they are considered desirable in themselves (see, e.g., Ref.^[14]).

Referendum, Initiative, Recall

At the state and local level, the oldest and perhaps the most basic form of citizen participation is a vote in a general or special election. Citizens typically vote on the performance or expected performance of candidates for public office. In rarer cases, citizens may be asked to pass or defeat referendums or initiatives on issues—citizens may unite to put their own issues on ballots—that have policy implications. Only 21 states have both a referendum and initiative option, 2 have initiatives only, 3 referendums only, and 18 neither option (www.iandrinstitute.org). Or they may vote elected officials out of office—recall—before their terms are up. Although elections are at the heart of democracy, they do not allow citizens to participate except by giving thumbs up or down. Some argue that the referendum process has been largely co-opted by monied interests able to influence electoral outcomes through sheer spending.^[15] In any case, elections may be a necessary condition of citizen participation in democracy, but are not sufficient in any sense. They also are not terribly relevant in community and economic development locally where most activities are project based and not the proper object of elections.

Hearings

Public hearings are the most common form of citizen participation in practice, at least locally. Virtually every unit of local government must hold hearings when plans are developed, projects are approved or amended, or budgets are proposed or revised. In many cases, federal, state, and local units of government may mandate hearings. Department of Housing and Urban

Development requires hearings on its Consolidated Plan, states require a hearing on matching funds tied to federal funding, and local governments may require hearings on federal and state funding allocated to the local budget, for example. Although the most frequently used form, hearings could be the least effective vehicle for citizen input. Typically, hearings are poorly attended and citizen comment is likely not representative, even when millions of proposed expenditures on economic and community development funding are at stake. Hearings are also risky for public officials because they are impossible to control—anyone can say just about anything, all before the mass media. Most officials can recount horror stories about how an angry citizen unexpectedly derailed a “safe” project in a hearing. The setting encourages posturing and discourages dialog and new thinking. For most hearings on economic and community development projects, public officials do not want fully developed projects changed or rejected in final stages after a great deal of effort has been expended. Perhaps this is why public officials view hearings as a pro forma requirement to get out of the way.

Advocacy Group Influence

Since the 1960s scholars (e.g., David Truman’s, *The Governmental Process*) have pointed out that interest groups have come to dominate citizen participation processes, usually around a narrow set of issues or programs that directly concern their members. For instance, directors of state-level community development agencies, such as the Council for State Community and Economic Development Agencies, lobby HUD on community and economic development issues, on everything from lead abatement in housing projects to management information systems used in reporting block grant expenditures. Although advocacy group input is important—these groups often have the expertise and information necessary to make good policy decisions—they may dominate citizen participation processes so much so that the process becomes lopsided. At HUD, for example, only two groups are mandated to be included through outreach in citizen participation processes in HUD’s CDBG grant—labor unions and the disabled community. All other groups must rely on their own devices to be heard.

Surveys and Focus Groups

Increasingly, communities are turning to surveys and focus groups as a preferred mode of citizen input (e.g., Ref.^[16]). Surveys, by telephone, mail, e-mail, group distribution, or in person, afford policy makers the opportunity to elicit large amounts of information

from representative groups of people at relatively low cost. Their disadvantage is that they are highly structured and not especially rich in the information they produce. Focus groups—small groups of people whose interaction yields rich information—often supplement survey information as a mode of input. An issue with surveys and focus groups is that they have become a kind of barometer used by public officials in decision making. In other words, decision makers decide based on trends in polls and focus groups. However, polls and focus group data may tap fickle opinions that can vary greatly in a short period of time. This is the case in Boston's \$15-billion "Big Dig" transportation/economic development project since the 1990s (www.bigdig.com). Public opinion support ebbed and flowed as bridges, highways, and tunnels appeared, while neighborhoods were disrupted; billions in cost overruns were reported; and dust, noise, and water pollution become commonplace.

Meetings/Workshops/Retreats

Meetings, workshops, and retreats are increasingly popular as ways to involve citizens in governmental decision making. These modes are most often used to gather information from citizens or provide them with it, greatly limiting their utility. Even so, these methods are especially useful in economic development where maps, models, and other visual aids are necessary.

Neighborhood/Community/Citizen Advisory Boards

Some communities formalize citizen input into programs or policy areas by creating special boards that offer input into planning processes. Boards do not have decision-making authority (these should not be confused with zoning or planning boards, which do have authority to make decisions). Boards typically strive to be representative of communities or neighborhoods by including one or more "stakeholders" from each group likely to be affected. The problem with

boards is that they cannot really be representative of neighborhoods, except in some nominal way. Board members also tend to have different values than either elected officials or the general public.^[17]

Group Facilitation

Group facilitation modes have been around for decades, but seem to be enjoying a renaissance in the last few years.^[18] Most methods have in common that citizens are chosen to participate in a process over time in which they reach consensus on problems and solutions (see Table 1 for a list of large group process methods). Public officials may participate in groups as members. The latest innovations in group methods have to do with increasing dramatically the number of participants to give their deliberations more credibility and to increase diversity of opinion. A shortcoming of group methodologies is that on the whole they have never undergone widespread scientific evaluation: What works and why? Are group methods superior to other methods? It's unclear whether the numerous techniques available are really different or just boast different "brand names."

Complex Voting/Computers/Internet

E-government is a high priority at the federal level, as well as at state and local agencies. At present, e-government focuses on the development of web sites to disseminate information to the public and e-mail links to elicit citizen feedback.^[19] Fourteen counties in northwestern Georgia, for example, linked together in the Growth Management Initiative web site to provide a forum on growth/antigrowth issues in a project funded by the Appalachian Regional Commission (www.arc.gov). Although common, these methodologies do not yet take full advantage of new technology. High-speed commuters, sophisticated decision-making software, and Internet access could revolutionize citizen participation processes. Citizens working at terminals, for instance, could participate in inexpensive sophisticated simulations and decision-making games and activities.^[20] However, even this method of citizen participation can be problematic. No matter how simple or user friendly, some individuals may be excluded as participants.

Table 1 Examples of large group process methods

Method	Web site
Citizen juries	www.jefferson-center.org
Collaborative input	www.ci.eugene.or.us/news/cisp
Shared decisions	www.ci.eugene.or.us/news/cisp
Open space technology	www.marwww.marand.si/business
Future search	www.futuresearch.net/fsis.htm
Appreciative inquiry	www.pancultural.com

CITIZEN PARTICIPATION PROCESSES IN ACTION

Despite the varied options available to public officials to improve access to government, there is wide

variation among communities in the extent to which they promote citizen participation opportunities. We believe that most communities offer minimal options; consider HUD's programs. As required by law, all entitlement communities must have at least two hearings related to the Consolidated Planning process. Many communities create citizen advisory boards, composed either of neighborhood residents or community representatives, including advocacy groups, citizen groups, nonprofit agencies, and the like. Many communities conduct surveys or focus groups to elicit citizen input. These can take many forms, ranging from narrow attempts to identify specific needs to more global looks at community development issues. Judging from HUD's Best Practice Awards program in the recent past, the latter activities are uncommon enough to enjoy the status of "best practices" (see HUD web site, "Blue Ribbon Practices in Community Development").

Richmond, California, for example, received a citation for best practice in citizen participation, stated as:

The City undertook a number of activities to expand and enhance citizen participation in the Consolidated Planning process. The City formed a Consolidated Plan Advisory Task Force that was made up of citizens, including Public Housing Authority residents and members of neighborhood councils. The Task Force reviewed drafts of the Plan and provided suggestions for improving the document. In addition, the staff of the Task Force brought together an inter-departmental working group and a variety of other agencies.

Columbus, Ohio's efforts under CDBG, by contrast, are exemplary. They distributed 10,000 surveys, door hangers, and fliers. They visited dozens of nonprofit, social service agencies, and public places (e.g., shopping malls) to elicit input. They set up community information hotlines; held numerous community forums, inviting citizens to attend by letter invitation; sent out direct mailings; and provided information to the mass media.

ISSUES

Because citizen participation is not straightforward in any sense, the following issues will always be of concern.

Defining Stakeholders

Most modes of citizen participation depend on targeting certain kinds of stakeholders who will participate.

But how should "stakeholders" be defined? An obvious way is to determine who is most affected by development activity, then bring them to the table. Who is most affected is always relative. Is it persons directly or indirectly affected? Are those who are slightly affected but intense considered, and consequently some will be excluded or included "inappropriately" in the process? Are taxpayers who will be asked to pay the bill stakeholders? Is impact enough or should it take into account severity of impact? The list of qualifications is endless. Once defined, how should stakeholders be identified? Should stakeholders those in certain demographic groups (e.g., age, race, language, ethnicity, income), residents of an area, or both. Definitions lead to more or less inclusion or exclusion. For example, HUD encourages communities to establish neighborhood strategy areas to address the needs of the poor. However, there are few areas with relatively large homogenous populations. Rather there is usually a mix of low- and moderate-income people, and even high-income people occasionally, all with a "stake" in development.

Decision-Making Authority

Many advocates want to see citizen participation processes have more "teeth" when it comes to implementation,^[21] but this is a double-edged sword. Why should citizens participate if their deliberations are not taken seriously? The other side of the coin is that elected and appointed officials are in office to make decisions. If they do badly, then citizens might vote them out. Laws and customs requiring citizen participation processes greatly favor the latter. An untenable situation exists when citizens devote inordinate amounts of time to providing input only to find it rejected. Some public officials desire to change this. In Seattle, Mayor Paul Schell inaugurated his "city of choices" initiative in which citizens provide early input on economic and community development issues that are then implemented by the city. Public officials gave citizens in Eugene, Oregon, control over small grants for community/economic development as a start in citizen empowerment.

Bad Advice

Public officials often eschew citizen participation processes because they want to avoid bad advice, especially publicly given. Although this seems reasonable on its face, it denies citizens their right to be heard—often and vociferously. It also is highly paternalistic, assuming that elites are much more likely to be "right." An emerging trend in this field is the notion that public officials are unreasonably afraid of

the public mostly because these officials have not received the training necessary to manage citizen participation processes effectively or to their advantage. Bramson advocates training public officials in the art and science of engaging citizens in a dialog, under the assumption that public officials have unfounded fears about engagement that might be overcome through learning and training.^[22]

Necessity to Triangulate Methods

Importantly, although there are numerous ways to engage citizens in the workings of government, employing only a handful of methods is not sufficient to fulfill the requirements of democracy. Different methods reach only selected people, but equally important, different methods when applied to the same people sometimes yield differing results. A survey and focus group methodology applied to residents of a neighborhood can easily produce different, perhaps conflicting, results. The dilemma for public officials is what methods to choose and how to interpret results. The best approach is to engage in as many methods as possible then reconcile disparate results where appropriate by bringing in additional information or consulting experts, or gathering additional input. However, this can be expensive and time consuming, not to mention complicated.

E-government

E-government, even in its most sophisticated and technologically advanced aspects, for many tends to be viewed as another way to facilitate citizen input. Although this is the case, e-government, according to columnist Neal Peirce, should not be considered a convenience, but rather a way to personalize government and make it more responsive, even for people of little wealth or influence.^[23] Viewed in this way, it could become essential as an expression of a properly functioning government. However, it will take considerable effort to move from convenience to responsiveness.

Planning v. Participation

Some argue that citizen participation and planning processes are in continual tension, so much so that the former is the Achilles heel of the latter.^[24] Citizen participation improves democratic processes, while planning mostly involves negotiation; relies on formal processes, while planning is much more informal; and is inclusive and routinized, while planning is selective, changing, and flexible. Citizen participation is, by

definition, transparent, while planning may occur, for legitimate reasons, behind closed doors. The implication is that effective citizen participation may lead to ineffective planning.

Evaluation

In spite of mandates for and the prevalence of citizen participation activities across the United States, few have ever been formally evaluated (e.g., Ref.^[25]). In part, public officials do not see the need to evaluate citizen participation processes, believing that investing in evaluation reduces funding available for programs, and may not want to know what works and what does not. These arguments against evaluation are unfounded. Evaluation can help to determine what works and what does not so that funding from all sources can be better spent. If performance-based management, grounded in goals, objectives, and outcomes is meaningful for program implementation, then evaluation of citizen participation processes would be critical in determining all phases of the policy process—design, implementation, evaluation, and feedback.^[26,27] Most federal funding and a great deal of state funding allows for and encourages expenditures for administrative costs, including evaluation generally and citizen participation specifically. Communities spend a lot on administration: why not channel it into evaluation. Public officials may not want to be held accountable, but if they are not, change will likely not occur. Citizen participation is the engine of change and must be done effectively—good or bad.

Inappropriate Citizen Input

Some argue that a few policy issues defy any rational use of citizen participation processes. These issues tend to be those that are complex, time sensitive, secret, or proprietary. The EPA exempts many of its policy issues from citizen input because administrators do not believe that people are sophisticated enough to understand the science and that people in any case would not have sufficient time to review and absorb documents necessary to acquire the scientific knowledge. The Defense Department, State Department, Federal Bureau of Investigation, Central Intelligence Agency, and numerous other agencies may exclude citizen involvement because of the sensitive nature of the policies (e.g., war and peace) being deliberated. All agencies may face time pressures on some policies (e.g., natural disasters) that preclude public comment. Aside from national security issues, exempting citizens for expediency's sake from input is likely overstated.

Promoting the Idea of Citizen Participation

Assuming that expanding and deepening citizen participation opportunities is desirable, who should take the lead in making it reality? Foundations and government agencies can fund demonstration projects to draw attention to the issue (www.pew-partnership.org). Universities, foundations, advocacy and professional groups, and government can publish best practice information for wide dissemination (e.g., www.hud.gov). Federal agencies can provide grantees, say under CDBG, with incentives to undertake citizen participation activities. Federal agencies or Congress could expand regulations to require greater participation opportunities. Perhaps what is also needed is a national, federally funded organization charged with promoting citizen participation nationwide. It is ironic that no such agency now exists, considering that America is a beacon for democratic practices, yet itself has a considerable way to go. The U.S. funds such an agency—Institute of Peace—for countries in transition, but not for our own country (www.usip.org).

CONCLUSION

Citizen participation is essential in any democracy, but must continue to be nurtured. Renewed interest in citizen input makes this a propitious time to experiment with methods to expand and deepen opportunities for citizens to interact with their government on economic and community development issues, and also other agendas. With computers, software, and the Internet the options are limitless, but a concerted effort in government and the public must occur.

REFERENCES

1. Putnam, R. *Bowling Alone: The Collapse and Revival of American Community*; Simon and Schuster: New York, 2001.
2. Osborne, D. *Laboratories of Democracy*; Harvard Business School Press: Cambridge, MA, 1990.
3. Osborne, D. *Reinventing Government*; A Plume Book: New York, 1993.
4. Seifter Economy and Hackman. 1997.
5. Chrislip, D.; Larson, C. *Collaborative Leadership*; Jossey-Bass: San Francisco, 1994.
6. Snow, D.; Woller, G.; Buss, T.F. *Microfinance and Economic Development Policy*; Nova: New York, 2001.
7. Schattschneider, J. *Semi-Sovereign People*; Thompson Publishing: New York, 1997.
8. Cortner, H. *Public Involvement*; Institute for Water Resources, US Corps of Army Engineers: Alexander, VA, 1993.
9. Madigan, D., et al. *New Approaches to Resolving Public Disputes*; National Institute for Dispute Resolution: Washington, DC, 1990.
10. Susskind, L.; Field, P. *Dealing with an Angry Public*; The Free Press: New York, 1996.
11. Susskind, L.; Cruikshank, J. *Breaking the Impasse*; Basic Books: New York, 1987.
12. Susskind, L., et al. *Consensus Building Handbook*; Sage: Thousand Oaks, CA, 1999.
13. Federal Highway Administration. *Public Involvement at Oregon Department of Transportation*; US Department of Transportation: Washington, DC, 1997.
14. Office of Intergovernmental and Public Accountability. *How to Design a Public Participation Program*; US Department of Energy: Washington, DC (undated).
15. Gerber, E. *The Populist Paradox*; Princeton University Press: Princeton, NJ, 1999.
16. The Leadership Initiative. *Strategic Planning For Community Development*; Department of Economic Development and Finance, State of North Dakota: Bismark, ND, 2001.
17. Redburn, F.S.; Buss, T.F.; Foster, S.; Binning, W. How representative are mandated citizen participation processes? *Urban Aff. Q.* **1980**, 15 (3), 345–352.
18. Group methods for whole systems change. *Public Organization Review*; Bramson, R.A., Buss, T.F., Eds.; Springer: The Netherlands, 2002; Vol. 2, 211–221.
19. Buss, T.F.; Redburn, F.S. Information Technology and Governance. In *Sound Governance*; Farazmand, A., Ed.; Marcel Dekker: New York, 2002.
20. Redburn, F.S.; Buss, T.F. *Modernizing Democracy*; Brookings Institution: Washington, DC, 2002.
21. Advisory Council on Intergovernmental Relations (ACIR). *Citizen Participation*; ACIR: Washington, DC, 1979.
22. Bramson, R.A. *Public Dialogue*; Union Institute: Cincinnati, 2000.
23. Peirce, N. *E-government: Not Just Convenience*; Alliance for Redesigning Government (www.alliance.napawash.org): Washington, DC, 2000.
24. Benveniste, G. *Mastering the Politics of Planning*; Jossey-Bass: San Francisco, 1989.
25. Office of Policy. *Stakeholder Involvement and Public Participation at the U.S. EPA*; Environmental Protection Agency: Washington, DC, 2001.
26. Council for Urban Economic Development. *Performance Monitoring: Achieving Performance Excellence in Economic Development*; Author: Washington, DC, 1998.
27. Department of Public Instruction. *Community Education Tool Kit*; Author, State of Wisconsin: Madison, WI (undated).

Economic Development, Community Organizations and

Herbert J. Rubin

Department of Sociology, Northern Illinois University, DeKalb, Illinois, U.S.A.

INTRODUCTION

In conventional economic development programs, locales rely upon a variety of incentives to compete with one another to attract upscale office, research, and commercial properties to downtowns and to suburban commercial, industrial, and research parks. As a consequence of these economic development efforts, public wealth is transferred to established firms, while the economic needs of older, industrial, inner-city neighborhoods already harmed by structural changes in the economy are neglected.^[1,2] Scholars and activists argue that an alternative approach toward economic development—one in which concerns with social equity, of helping the poor and poor neighborhoods—predominates.^[3,4]

Unfortunately, the administrative complexities of implementing equity economic development within disadvantaged neighborhoods can be daunting. Projects tend to be small, provide little publicity for elected officials, and are undertaken in the neighborhoods with which most public officials are least familiar. In addition, in inner-city neighborhoods, individuals face numerous social problems that complicate the efforts to bring about economic repair.

To bring about equity in economic development, activist scholars suggest following a “community option in urban policy” through which “local services and redistributive subsidies, along with decision-making authority and accountability, might be decentralized...from public to community-based organizations.”^[5] With the community option, the public sector is not directly involved in programs of neighborhood repair, but instead it facilitates those economic development efforts chosen by community members and implemented by community-based development organizations (CBDOs). Such an approach can succeed. The National Congress for Community Economic Development, the trade association for CBDOs, notes that community organizations have already developed 71 million square feet of commercial and industrial space, lent \$1.9 billion to business enterprises, and created 247,000 full-time jobs mostly benefiting poor communities and their residents.^[6]

TYPES OF CBDOS

Community-based development organizations are organizations responsible to a neighborhood or constituency group of disadvantaged individuals, usually but not always incorporated as non-profits, whose stated mission is to economically improve neighborhoods of deprivation in ways that benefit current residents and individuals in need. A wide array of community-based development organizations undertake equity economic development work.

Many *neighborhood organizations* and *block associations* include economic development efforts as part of their ongoing community organizing efforts. These non-profit and volunteer organizations encourage city hall to provide infrastructure repairs or to increase services (e.g., garbage pickup, frequency of public transit) needed to retain existing businesses in the neighborhoods, join coalitions to advocate for neighborhood projects, and encourage the formation of specialized organizations that focus on economic development work. In several progressive cities, government have worked with neighborhood planning organizations to obtain citizen input into economic development efforts. In Minneapolis, for instance, neighborhood organizations are responsible for the local plans that city line agencies follow in their renewal efforts. In Boston, the Dudley Street Neighborhood Initiative has prepared for its own neighborhood and city redevelopment plan. In several dozen other locales, comprehensive community initiatives (CCIS) set up by foundation-city-neighborhood partnerships to coordinate social programs, housing efforts, and infrastructure renewal to improve the community environment in ways that will create jobs or attract businesses.^[7]

In contrast to neighborhood organizations, planning groups, CCIs, community development corporations (CDCs), and community development financial institutions (CDFIs) focus their entire efforts on economic development and housing projects. Community development corporations are community-based organizations that undertake capital-intensive, physical development projects, such as building affordable housing, setting up commercial properties, and

promoting industrial renewal. Although most of their work focuses on housing renewal, many CDCs also act as neighborhood economic developers. Community Development Corporations are structured as non-profits that can receive grant money, but are also chartered as quasicapitalistic organizations that can participate in commercial, housing, and industrial development work. Community development corporations can become stock holders, partners, or owners in for-profit ventures; acquire debt; and manage firms. In addition, CDCs act as a neighborhood business advocates, run training programs that help people to improve their job skills or learn how to set up a business. To complement their efforts in housing and economic development, CDCs collaborate with social service agencies to help individuals combat personal, family, or economic problems that limit their economic activities.^[6] Another cluster of community-based-organizations, CDFIs, focus their efforts on providing capital needed for economic renewal within poor neighborhoods. The array of CDFIs include community credit unions, community venture capital development funds, community development banks, and microenterprise funds that collectively act as financial intermediaries within poor neighborhoods.^[8]

PHILOSOPHIES OF COMMUNITY-BASED DEVELOPMENT

How do the efforts of community-based development organizations differ from those of conventional economic developers? To begin, the work of CBDOs is guided by a shared organic theory^[9] that promotes the idea that economic development is about empowering the poor, not merely about doing projects. Redevelopment projects initiated by CBDOs explicitly attempt to increase social equity in ways that benefit those most in need. Promoting dead end, unskilled jobs does not bring about social equity; neither do high-tech projects that displace the poor and gentrify the neighborhoods. Instead, in their efforts to make sure that the poor benefit from economic development efforts, a CBDO would set up a training program to prepare community members for the skilled positions offered by firms that are helped by the CBDOs. For instance, in Kansas City, a CBDO helped a telephone company to set up a back office facility in an inner-city neighborhood that employed community members who received the training necessary to upgrade their skills to the level needed by the company.

As a second component of the organic theory, CBDO-sponsored economic development projects are intended to increase capital flow within poor neighborhoods. When firms that expand in the neighborhood hire community members, wages earned are recycled

as employees purchase goods in locally owned stores. Community-based development organizations have helped to develop supermarkets in neighborhoods that lacked convenient and affordable shopping, insisting that the supermarkets both train and employ community members. Or, a CBDO might assist a local business that provides insulation to homes, enabling community members to save money on utility money, money that is exported from the community, and to spend it in neighborhood stores.

Third, economic development is not done in isolation but instead becomes part of broader, programs of holistic social repair. Efforts to improve schools or provide day care are incorporated within the broader economic development agenda. Without better schools, employees are not prepared for jobs; for single mothers to be able to leave home and work, day care is needed. Women are more likely to be victims of poverty than are men, so many CBDOs establish programs to aid female entrepreneurs obtain capital or learn how to convert a skills or talents into a microenterprise business. Because businesses fear crime; crime watch and antidrug programs become part of the CBDO's broader economic development agenda. The housing build by CDCs creates a stable residential environment for the poor, while luring back the working class. In turn, the residents of the new, affordable housing form a market that encourages commercial ventures to return to the neighborhood. In regional efforts, CBDOs join with environmentalists to promote smart-growth programs that restore inner-city communities, helping community residents and reducing urban sprawl.^[10]

TECHNIQUES OF COMMUNITY-BASED ECONOMIC RENEWAL

In guiding the economic renewal of poor neighborhoods, CBDOs follow five overlapping approaches that vary from directly sponsoring entrepreneurial, capitalistic projects to community-organizing efforts meant to pressure government not to abandon poor communities.

As entrepreneurial capitalistic investors, CBDOs themselves establish profit-making enterprises, usually those supplying goods or services lacking within poor communities. In undertaking such efforts, the CBDO will first obtain public or foundation support to pay for the social costs of remediation—acquiring and repairing an abandoned property, cleaning up a brownfield, removing abandoned vehicles. Then, with the land improvements accomplished, the CBDO will by itself or in partnership with conventional firms develop a commercial project on the restored site. For instance, the Kansas City CDC rehabilitates abandoned inner-city sites, one so derelict that it was

used as a set for a postatomic war movie, and then builds inner-city shopping malls often owned and managed by the CDC itself. In a disadvantaged Newark neighborhood, New Communities Corporation (in partnership with a grocery chain) built a full-service super market on a previously derelict lot; other CBDOs have started firms to process and recycle waste materials found in the older industrial neighborhoods into usable products. People in poor communities are consumers of social and medical services for which they often have to leave the neighborhood to receive. Community-based development organizations, such as Bethel in Chicago, will set up social service agencies—medical clinics, for instance, that provide needed services within the community while also training community members to work in these agencies. As social service providers, such CBDOs themselves become a major community employer. As quasicapitalists investor, CDFIs endeavor to expand the capital available in poor communities, setting up development banks that directly invest in neighborhood projects or forming community development credit unions that allow community members both to deposit their earnings and to receive loans.

With catalytic economic development projects, CBDOs encourage government to provide the incentives that enable for-profit firms to expand their efforts within neighborhoods of need. For instance, as neighborhood shopping strips disappeared, poor communities found themselves devoid of economic centers that provided needed goods and services and symbolized economic vitality. Acting as chambers of commerce, CBDOs form business investment districts that obtain public resources to restore such shopping strips. Community-based development organizations have initiated transportation-oriented development projects (TODs) in which government is pressured to improve a transit node and, in doing so, rebuild in ways that encourage stores to locate near this high traffic generator. By setting up industrial or commercial incubators, CBDOs catalyze economic renewal. In incubator projects, a CBDO obtains a subsidy to convert an older building, often an abandoned industrial plant, into affordable space for start-up firms, both from the neighborhood itself and those that will employ community members. At times, funds are even available to hire a successful business person to act as a mentor for start-up companies. In other catalytic efforts, CBDOs help to form worker cooperatives in which individuals learn skills, market products, and collectively own their firms. In Alaska, for instance, indigenous people have set up a Musk Ox Cooperative to process and sell the exotic wool of that Arctic animal.^[11] Community development financial institutions provide loans (that mainstream financial institutions refuse to make available) to

conventional entrepreneurs willing to invest in neighborhoods of need, while revolving loan funds are established to provide the small amount of seed capital required to set up microenterprises. Community-based development organizations will teach neighborhood capitalists how to prepare business plans and pro formas that can then be used in obtaining investment capital, while introducing them to loan officers at banks.

Most CBDOs recognize that to repair the economic damage in poor communities, efforts are needed to expand social and human capital. Poor people often lack both training and capital to start a business or to obtain the specialized education needed for better employment. To encourage this expansion in human capital, CBDOs link training and education efforts to individual development account programs (IDAs). While community members further their education or undertake apprenticeship training, money is put aside for them in an escrowed account. On completion of the training, people receive the lump sums to be used as investment capital either in a business or in further education. Individuals from poor communities lose wealth by being victimized by exploitative predatory lenders. Community development financial institutions establish financial literacy programs, teaching people how to negotiate the increasingly complicated financial system and to avoid losing their limited capital to unscrupulous lenders.

Community-based development organizations expand social capital by linking people to each other. Appalachian Center for Economic Networks (ACENET) in Appalachian Ohio set up flexible manufacturing networks that linked small manufacturers, designers, and marketers to promote products, while Esperanza Unida in Milwaukee brought together older experienced community members to train unemployed youth as mechanics. In Chicago, the Jane Addams Resources Corporation responds to requests by neighborhood companies to help workers the companies already employ to update the skills needed to enable the companies to survive in a changing economy.^[12] After the major change in the mid-1990s in welfare laws, CBDOs expanded their efforts to provide training and job access to the welfare poor. For instance, in Texas, Project Quest run by a CBDO provides potential employees with the specific training requested by firms, virtually guaranteeing future work.^[13] Social service agencies can expand needed social capital. For instance, day care, especially services that are open during the irregular shifts that the poor are forced to work, create jobs within the community for those who provide the service, while allowing others to take jobs away from home. Increased neighborhood pride also expands social capital. Several CBDOs have initiated projects that celebrate a community's culture,

such as the Jazz Museum and Black Baseball Museum established by a CBDO in Kansas City.

As advocates for neighborhood economic development, CBDOs provide an assertive voice that speaks for the economic needs of poor communities. Community-based development organizations set up “early warning systems” to shout out to the public sector when established businesses threaten to leave inner-city neighborhoods, often because of the lack of city services. Community-based development organizations join with local labor alliances, such as in the Calumet Project in Indiana, to pressure companies not to abandon poor neighborhoods.^[11] Community-based organizations such as BREAD in Columbus, Ohio, pressure city officials to set up transit centers that enable those from the inner city to commute to suburban jobs.^[10]

Working in citywide coalitions, CBDOs advocate for first-source hiring programs that mandate that companies that receive public subsidies must hire community members, often people trained by the CBDO. In living wage campaigns, CBDOs join with other community activists to insist that no public dollars be provided to firms unless the companies pay wages that at least match the family poverty level. Advocacy coalitions such as Chicago’s Neighborhood Capital Budget Group have brought together CBDOs and neighborhood activist organizations to campaign to ensure that neighborhoods, not just downtown, receive funds for infra-structure repairs.

GOVERNMENT POLICIES IN SUPPORT OF CBDOs

How can public policies and the actions of public officials be redirected to help the community-based development in this movement? To begin, public officials must recognize that heavily subsidized, grandiose, downtown projects do little to help the poor. Instead, they must accept that economic development is about creating social redress and that CBDOs play a crucial part in bringing about economic equity.

Public subsidies to establish economic development trust funds can help to pay for the seed money for projects in poor communities, as does capital to set up revolving loan funds. In poor communities, land ends up in public ownership because of tax forfeitures. The public sector can donate this land to CBDOs for economic development work, but before doing so paying for the remediation that many of these brownfields require. In places such as in the Dudley neighborhood in Boston^[12] or in Chicago during the progressive Mayor Washington administration, government delegated authority to community groups for planning, land use, and zoning, and seed money for start-up

firms,^[14] enabling communities to better control their own development. Local government salaries must recognize that CBDOs rarely make sufficient profit to pay for their core operating expenses (e.g., salaries of ongoing employees, office space) and should use community development block grant, or tax increment financing funds, to provide a substantial proportion of the core operating expense of CBDOs.

National policies, too, could be introduced or enhanced to help the community economic development movement. The Office of Community Services in the Department of Health and Human Services administers a small grant program that provides capital investments to neighborhood economic development projects that have leveraged a significant private sector investments. This program must be continued and expanded. Public funding helps to pay the administrative costs of CDFIs, and enables these financial intermediaries to work on the smaller, overhead-intensive, investment projects needed to stimulate neighborhood business growth. Federal programs such as the New Markets Tax Credit that provide tax credits that encourage both for-profits and CBDO-for-profit partnerships to invest in poor communities need to be expanded. Most important, federal regulators must enforce the Community Reinvestment Act (CRA) that mandates banks to reinvest in poor communities, both for housing and economic development projects. New federal regulations are needed that require the increasing array of non-bank financial institutions to provide capital to communities of need.

Recognizing that they need help, activists within the community redevelopment movement, working with foundations, have set up support organizations that provide technical assistance and political advocacy for community developers. These organizations include trade associations, technical assistance providers, financial intermediaries, training academies, specialized consultants, and accounting and legal firms that concentrate on community renewal programs, and increasingly include academic institutions that offer programs in community redevelopment. Expanded public funding of these support organizations would dramatically increase the knowledge available on community economic development.

REFERENCES

1. Imbroscio, D. *Reconstructing City Politics: Alternative Economic Development and Urban Regimes*; Sage Publications: Thousand Oaks, CA, 1997.
2. Bright, E. *Reviving America's Forgotten Neighborhoods: An Investigation of Inner City Revitalization Efforts*; Garland Publishing: New York, NY, 2000.
3. Rubin, H. Partnering with the poor: why local government should work with community-based development

- organizations to promote economic development. *Int. J. Public Adm.* **2000**, 23 (9), 1679–1709.
4. Krumholz, N. Equitable approaches to local economic development. *Policy Stud. J.* **1999**, 27 (1), 83–95.
 5. Clavel, P.; Pitt, J.; Yin, J. The community option in urban policy. *Urban Aff. Rev.* **1997**, 32 (4), 435–458.
 6. Rubin, H.; Rubin, I. *Community Organizing and Development*, 3rd Ed.; Allyn and Bacon: Boston, MA, 2001.
 7. Meyer, D.; Blake, J.; Caine, H.; Pryor, B. *On the Ground with Comprehensive Community Initiatives*; The Enterprise Foundation: Columbia, MD, 2000.
 8. www.cdfi.org/what.cdfi.html (accessed July 30, 2001).
 9. Rubin, H. *Renewing Hope Within Neighborhoods of Despair: The Community-Based Development Model*; State University of New York Press: Albany, NY, 2000.
 10. Kalinosky, L.; Desmond, K. *Smart Growth, Better Neighborhoods: Communities Leading the Way*; National Neighborhood Coalition: Washington, DC, 2000.
 11. www.grass-roots.org (accessed 3/14/2000).
 12. Pitcoff, W. Developing workers: community-based job training brings families out of poverty. *Shelterforce* **1998**, xx (6)10–13, 28.
 13. Harrison, B.; Weiss, M. *Workforce Development Networks: Community-Based Organizations and Regional Alliances*; Sage Publications: Thousand Oaks, CA, 1998.
 14. Mier, R. *Social Justice and Local Development Policy*; Sage Publications: Newbury Park, CA, 1993.

Economic Development, Financing

William L. Waugh Jr.

*Department of Public Administration and Urban Studies, Georgia State University,
Atlanta, Georgia, U.S.A.*

INTRODUCTION

Economic development is a policy priority for most local, regional, and central governments. Economic development is a means to reduce unemployment, increase tax bases, leverage other resources from the community, and curry public favor and support. When economic development slows, stagnates, or declines, it has an adverse effect on the community and political and economic leaders are held responsible. Therefore, the competition for new and expanding businesses can be intense and economic development programs may include a broad range of incentives to lure firms from other communities and to cultivate new firms locally. Financing economic development is a crucial concern and one that raises fundamental issues concerning who should pay for government programs, and how much should and can be invested.

DIRECT AND INDIRECT FINANCING

The financing of economic development programs may be direct in terms of investments of public monies in specific firms or sites, or indirect in terms of encouraging private sector investment. To some extent, there is an expectation that economic development will be guided and funded by those who reap the most direct benefit, private firms and individuals. However, governments guide choices and invest public resources to ensure that economic development serves the best interests of the community. In fact, many governments have pursued development with little consideration given to the social and economic impact on their communities, but such unfocused approaches are much less common today. To increase the likelihood of appropriate or desired kinds of development, communities typically offer incentives to locate, relocate, or expand in their communities. Incentive programs may include economic or financial incentives, along with other kinds of incentives. Financial incentives provide direct or indirect support. For example, governments may invest public monies in specific businesses as low-interest loans, or they may invest monies in training and infrastructure thereby permitting firms to use their money for other purposes.

The unavailability of adequate financing has been a major impediment to economic development in many jurisdictions. As a private sector issue, financing has largely focused on the cultivation of venture capital when traditional financing (i.e., bank loans) is unavailable. As a public sector issue, the financing of economic development has been through direct and indirect means. Governments have invested public monies directly in economic development efforts (e.g., providing loans to new and expanding businesses and providing access to pools of money such as public pension funds). They have also invested indirectly—offering tax abatements to reduce the financial burden of property and other taxes; issuing bonds to finance infrastructure and facilities; providing grants usually for specific purposes, such as training; assisting with debt financing, such as providing loan guarantees to help reduce interest rates; and offering more direct financial assistance in exchange for royalty payments.^[1]

Revolving Loan Funds

One of the most direct means of financing economic development with public money is to create a revolving loan fund. Loans are made available for start-ups, expansions, and other needs. As loans are repaid, the money is returned to the fund and made available for new businesses.

Loan Guarantee Programs

Rather than making money available through a public loan program, governments can provide guarantees to banks and other financial institutions that loans made to businesses locating in an industrial park or enterprise zone will be repaid even if the firm fails. Loan guarantee programs are useful when traditional lenders consider the risk of business failure too high for them.

TAX INCENTIVES

Tax incentives have been one of the more popular means of encouraging economic development, and

they are widely assumed to have a significant impact on the site selection and/or expansion decisions of businesses. The popularity of tax incentives has characterized local and state development programs for several decades, despite the number of studies that demonstrate their marginal impact. Tax incentives generally take three forms: 1) broad-based incentives; 2) targeted incentives; and 3) selectively applied incentives. The broad-based incentives are offered to any and all businesses that choose sites within a jurisdiction, make particular investments, or provide jobs of a particular kind or number. The government offers tax credits, exemptions, and/or deductions in exchange for investment, new jobs, research and development, and/or depreciation of equipment and/or other assets. Targeted incentives are offered for specific kinds of firms, such as firms choosing to locate in areas of high unemployment or in free enterprise zones or so-called high-tech or manufacturing firms. Communities may attempt to lure “clean” industries, such as movie making, hospitality or tourism, research, and education.^[2-4]

Property Tax Abatements

For local governments, in particular, property tax abatements can provide substantial financial relief for new businesses. Local governments generally do not have the financial wherewithal to create revolving funds and to make large investments in infrastructure and human resources, but they do have authority to levy property taxes. Reducing or eliminating property taxes for a period of time, sometimes for 10 or more years, can be a significant incentive for new businesses.^[5] State laws may also permit exemptions from inventory taxes through the establishment of “free ports.”

Also, economic development authorities, such as community development corporations, can purchase sites and build facilities and lease them to firms without the firms being liable for property taxes. They can also assist in securing loans for start-up and other needs, provide technical assistance, and act as an intermediary in dealing the regulatory agencies.

Tax Increment Financing

Tax increment financing has been a popular means of funding economic development because it is essentially a product of that development. Additional property taxes that result from the development in an enterprise zone or development district are paid to the general-purpose (town or city or county) government or special-purpose (school district) government and placed into a fund to finance improvements to the zone or district.^[6,7]

Industrial Revenue Bonds

The broadly focused approach has been characterized by the use of incentives, such as industrial revenue bonds, subsidies for compliance with pollution control regulations, property tax abatements, loan guarantees, accelerated depreciation of business assets, and corporate income tax exemptions.^[8]

Free Enterprise Zones

In the 1980s, “free enterprise zones” were the fashion and many still exist in the United States. “Free enterprise zone” programs are characterized by the designation of an area for development and the provision of various incentives for businesses locating within the zone. Property tax abatements, public investments in infrastructure improvements and job training, increased police protection, loan guarantees, tax credits for hiring residents of the zone, and other benefits are offered to lure businesses to the zone and to reward those that remain.

INDIRECT PUBLIC INVESTMENT

Infrastructure Improvements

To some extent, public investment in infrastructure, including the creation of industrial and research parks, provide indirect financing for economic development. The investments can reduce start-up and operating costs to business and can facilitate growth.

Technical Assistance Programs

Programs to provide technical assistance with marketing, export regulations, and other crucial capacities, and management training to improve internal operations, may also be considered indirect financing. Such programs increase the likelihood of success and, thereby, increase the likelihood of funding by banks and other venture capital sources.

Land Banks

Some cities have created land banks to assist firms in acquiring properties for development. Abandoned, tax delinquent, donated, and transferred properties are put into the land bank. The bank assures that the titles are clarified, waives tax liability, and sells the property to developers. The land bank can be used to put together parcels for large development projects, such as manufacturing plants and condominium complexes.

The bank can reduce the cost of property to developers and facilitate the process of land acquisition by assuring that large parcels can be acquired without having to deal with many different property owners.^[9]

LEVERAGING PRIVATE FINANCING

Perhaps the most obvious and less risky approach to economic development is through leveraging private investment. Economic development can be financed through the encouragement of traditional bank financing and through foundation grants. Community commitment to development can increase the likelihood of success among participating firms. Banks and foundations may judge that the community commitment will increase the likelihood of success. Similarly, government officials and private entrepreneurs can cultivate venture capital from private investors.

Community Improvement Districts

Private investment is also encouraged through the creation of community improvement districts. Development within the districts is facilitated by public-private arrangements and funded directly by a system of self-taxation. Public investment is generally in the form of enhanced public services, including public safety, and infrastructure improvements.^[10]

CONCLUSION

The trend is toward broad incentive packages that are responsive to the needs of both the community and the participating firms. In the state of Georgia, for example, the state offers job tax credits, investment tax credits, retraining tax credits, child care tax credits, sales tax exemptions, and freeport (property) tax exemptions.^[11] The stakes are high in economic development efforts. There is much risk in direct investments of public money because of the high failure rate of new firms and at least some risk in indirect investments in such things as infrastructure improvements that may be firm- or industry-specific because firms may choose to leave for reasons beyond the control of the local government. In short, communities may be left with millions of dollars in bond debt for improvements that the next firm may not find useful or desirable. To address that problem, Ledebur and Woodward suggested that governments evaluate development incentives carefully and assure that they are cost effective. When firms do not live up to their responsibilities, governments should be

prepared to reduce or eliminate public subsidies for non-performance.^[12]

The trend has also been toward more creative forms of financing to overcome constitutional and statutory limits on taxing, spending, and borrowing (debt) authority. Local governments have few sources of revenue and limited political autonomy, except in some larger cities and counties. State governments, too, have limits on how they can raise and spend revenues. The use of general obligation bonds may be severely limited and there is tremendous political opposition to new taxes regardless of their purpose. These are the constraints within which economic development authorities operate.^[13]

REFERENCES

1. Hill, E.W.; Shelly, N.A. An overview of economic development. In *Financing Economic Development: An Institutional Response*; Bingham, R.D., Hill, E.W., White, S.B., Eds.; Sage Publications: Newbury Park, CA, 1990; 13–28.
2. Waugh, W.L., Jr.; Waugh, D.M. Baiting the hook: targeting economic development monies more effectively. *Public Adm. Q.* **Summer 1988**, 12, 216–235.
3. Dewar, M.E. Tax incentives, public loans, and subsidies: what difference do they make in non-metropolitan economic development. In *Financing Economic Development*; Bingham, R.D., Hill, E.W., White, S.B., Eds.; Sage Publications: Newbury Park, CA, 1990; 40–54.
4. Ihlanfeldt, K.R. Ten principles for state tax incentives. In *Approaches to Economic Development*; Blair, J.P., Reese, L.A., Eds.; Sage Publications, 1999; 68–84.
5. Hy, R.J.; Waugh, W.L., Jr. *State and Local Tax Policies*; Greenwood Press: Westport, CT, 1995; 219–221.
6. Bland, R.L. *A Revenue Guide for Local Government*; International City Management Association: Washington, DC, 1989.
7. Paetsch, J.R.; Dahlstrom, R.K. Tax increment financing: what it is and how it works. In *Financing Economic Development*; 1990; 82–98.
8. Rasmussen, D.W.; Benedick, M., Jr.; Ledebur, L.C. Evaluating state economic development incentives from a firm's perspective. *Bus. Econ.* **May 1982**, 17, 23–29.
9. Hy, R.J.; Waugh, W.L. *State and Local Tax Policies*; Greenwood Press, 1995; 226–227.
10. Hy, R.J.; Waugh, W.L. *State and Local Tax Policies*; 1995; 225–226.
11. Ihlanfeldt, K.R. *Ten Principles for State Tax Incentives*; 1999; 77–79.
12. Ledebur, L.C.; Woodward, D.P. Adding a stick to the carrot: location incentives with clawbacks, rescissions, and recalibrations. In *Approaches to Economic Development*; Blair, J.P., Reese, L.A., Eds.; Sage Publications: Newbury Park, CA, 1999; 61–66.
13. Schoettle, F.P. What public finance do state constitutions allow. In *Financing Economic Development*; 1990; 57–74.

Economic Development, Immigration, Self-Employment and

Marcela Tribble

*School of Policy and Management, Florida International University,
Miami, Florida, U.S.A.*

Terry F. Buss

*Center for International Affairs, National Academy of Public Administration,
Washington, District of Columbia, U.S.A.*

INTRODUCTION

Immigration for decades has interested economists, sociologists, and historians, at least in part because America is a nation of immigrants. Only recently has immigration attracted much attention from economic development researchers, policy makers, and practitioners. Even though immigration is a huge economic policy issue in America, little solid empirical evidence exists to inform debate, and the information that does exist tends to be contradictory. Yet there is enough information available to cobble together insights around the following questions:

- Immigrants arrive in the United States from around the world: which groups are likely to become entrepreneurial and why?
- Immigrants affect local economies in various ways—good and bad: what are these effects, and why do they occur?
- What are the policy implications facing local economic development decision makers?

We focus mostly on self-employment among immigrant groups because this is central to economic development. We do not address the importance of immigrants to the labor force—immigrants are critical in filling jobs in lesser skilled and, in some cases, higher skilled occupations—because this is another policy domain, better addressed elsewhere. Before discussing these questions, we provide an overview of trends in immigration.

IMMIGRATION TO AMERICA

Commentators have christened the decade of the 1990s as the *Second Great Migration*, the first occurring at the turn of the twentieth century.^[1] In the past decade, about 9 million legal immigrants arrived, a number

unprecedented since the early 1900s when 8.8 million migrants came to America. In fiscal year 2000 alone, the Immigration and Naturalization Service (INS) legally admitted 849,807 immigrants into the United States. The number of foreign born, according to the 2000 Census, is now 31,107,889, or 11.1% of the population. If children of immigrants born in the United States are included, the immigrant population becomes 20.4%. Immigrants account for 37% of U.S. population growth, 50% if calculations include their American-born children.

States receiving most immigrants in 2000 were California (217,753), New York (106,061), Florida (98,391), Texas (63,840), New Jersey (40,013), Illinois (36,180), and Massachusetts (23,483). Immigrant groups tend to migrate to places having large populations from similar cultures. Chinese cluster in California and New York; Cubans and Haitians in Miami; Indians, Filipinos, and Salvadorians in California; and Mexicans in California, Texas, Arizona, and Florida. U.S. Census data show that 95% of immigrants live in urban areas, and that 65% live in just 10 cities, probably because immigrants receive a better reception in cities, can access more services, and have better opportunities.

Immigrants come from all over the world, but some countries contribute a much greater share. In 2000, Mexico (173,919), China (45,652), Philippines (42,474), India (42,046), Vietnam (26,747), Nicaragua (24,029), El Salvador (22,578), and Cuba (20,831) were major contributors. This represents a substantial shift away from European and Canadian origins, where economies are in good shape, toward Latin American and Asian countries, where prosperity is limited.

Of those admitted to the United States in 2000, few (189,357) reported having an occupation. But 94% of working-age immigrants were in the labor force, as compared with 91% of natives. Some 28%, or 237,841 immigrants, were 19 years of age or younger, making them on average much younger than the native population. One-half (54.4%) of all immigrants were married. About one-third arrived with less than a

high school education, but 29% had a college degree. Those with no occupation, no family, and little education will have difficulty becoming self-employed.

Anywhere from 7.1 to 8.5 million immigrants are illegal aliens. Two-fifths entered the United States legally, lost their legal status, and remained here as illegals. About one-half of illegals are Mexicans. Illegal aliens will find it much more difficult to find work, let alone successful self-employment.

WHO BECOMES SELF-EMPLOYED?

Self-Employment Theory

From an economic development perspective, one would want to know why some groups become productive members of communities through self-employment, whereas others arriving in the United States do not. There are three widely invoked models to explain immigrant new business formation—*middlemen minority*,^[2] *ethnic enclave*,^[3] and *reactive cultural*^[4] theory—all of which focus on *supply-side*, meaning those who produce goods and services, and, comparisons of intergroup differences.^[4]

Middlemen minority theory holds that some immigrants enter majority society only for commercial reasons where they serve as middlemen between well-off producers with goods to sell and poorer consumers desiring to buy them. They remain isolated or detached from majority society because they eventually intend to return home. Korean grocers in Los Angeles, for example, are middlemen between white producers and distributors of produce and poor African Americans in distressed neighborhoods. Ethnic enclave theory asserts that successful immigrant entrepreneurs tend to locate in clearly bounded geographic areas where their culture, language, networks, access to labor and capital, and vertically and horizontally integrated markets give them a comparative economic advantage over those in the larger economy. Cubans, once densely concentrated in Miami's "Little Havana," served as a model of this characterization. Reactive cultural theory (and its predecessor, cultural theory) emphasizes trust, solidarity, and mutuality among immigrant groups who pool capital for investment in small businesses, and coordinate social and economic activities. Japanese immigrants to America in the years prior to World War II used ethnic solidarity and ethnic resources to establish and manage small businesses.^[5]

Although these theories persist in professional and popular literature, they have been roundly criticized, and even debunked (e.g., Refs.^[5]). Middlemen theory overlooks the fact that many immigrants do not intend to return home. Ethnic enclaves cannot explain how immigrants initially started businesses, seemingly only

explaining what happened much later in the business formation process when resources had already become available. In addition, researchers have found few, if any, ethnic enclaves, because economies are much more diverse, less spatially isolated, and less integrated than this narrow model suggests. Cultural theory overemphasizes cultural solidarity, failing to take into account class differences among immigrants. Upper-class immigrants have access to scarce resources, especially capital, not available to everyone in that culture. Cuban refugees fleeing Castro's regime in 1959–1960 were well off, whereas Mariel Boatlift refugees, composed of many criminals and institutionalized populations, of the 1980s were much less prosperous. The former groups took over Miami's economy; the latter were left out.

The most convincing explanation for successful entrepreneurship is that immigrants to the United States, or anywhere else for that matter, must possess the capacity either to create wealth themselves (or to obtain work that allows them to be independent or to eventually become so), and they must locate at or migrate to a place that provides economic opportunities. Successful immigrants are those whose capacity matches opportunity—*interactionism* theory.^[4,6] Taiwanese, having math, computer, and engineering skills, came to America to participate in the "dot.com" revolution.

Other theories try to explain why immigrants become self-employed. The two leading contenders have yet to gather convincing empirical support, although both persist in the literature.^[7] One is the *blocked mobility* theory—that immigrants facing discrimination or barriers in labor markets are forced into self-employment as the only alternative to unemployment or welfare. Discrimination takes many forms—racial, linguistic, cultural, and religious. But this cannot explain why larger numbers (9 in 10) succeed as workers, and why some immigrants facing barriers never break out. The other is the *predisposition* theory—that some groups are predisposed to self-employment because they were highly entrepreneurial in their home countries and desire to continue the tradition. But this theory cannot account for such phenomenon as the success of recent Eastern European and Russian immigrants who had no business start-up experience.

Ethnic Businesses

A study by Razin and Light^[4] shed light on ethnic self-employment in 16 of the largest metropolitan areas. The rate of self-employment was highest in Sunbelt and western metropolitan regions characterized by economic and cultural diversity: Miami, Los Angeles, San Francisco, San Diego, and Seattle. Lowest rates were in declining manufacturing regions: Detroit,

Cleveland, Chicago, and Philadelphia, and in the nation's capital, Washington, DC. Immigrant self-employed, as a percentage of the immigrant population, are prevalent in central cities, 10.3%; suburbs, 12.5%; and nonmetro areas, 7.9%.

The higher the rate of self-employment among immigrants, the higher the rate of self-employment overall in the labor force; that is, entrepreneurial places breed entrepreneurial behavior across a wide spectrum. Similarly, the higher the level of immigrant self-employment, the higher the level of income within immigrant groups.

Greeks and Koreans were consistently self-employed in cities across the country. Other "conspicuously" self-employed groups include Middle Eastern and North Africans, among them Iranians and Israelis. Dutch, Hungarian, Czech, and Polish self-employed are dispersed nationwide. Some immigrant groups dominate selected cities. Cubans monopolize Miami, probably more than any other immigrant group controls an American city.

The least entrepreneurial seem to be people from the Caribbean and Philippines. Mexican and Puerto Rican immigrants have lower levels of self-employment, but some argue that they are undercounted because they work in the underground economy as street vendors, handy men, gardeners, babysitters, and the like.^[8] But many analysts would reject these activities as they do not really represent self-employment, but causal, part-time labor.

Capacity for Self-Employment

Factors affecting self-employment capacity are generally well-known and apply to immigrant populations. Immigrants having business experience, appropriate skills or education, access to networks (e.g., professional, family, or groups) able to provide support, values or cultural attributes contributing to wealth creation, and access to capital tend to have a greater propensity for successful self-employment.

English

The ability to speak English and its relationship to self-employment has not been studied. Findings for the labor force may generalize to the self-employed, but they are mixed; some finding English language to be a major factor, others not so. Generally, those who speak English improve their earnings over non-English speakers, while those in jobs not requiring English or those living in areas not dominated by English speakers are somewhat less affected.^[9] It is likely that speaking a language that fits requirements of the market is an important factor and that being bilingual cannot

hurt business prospects. Generally, as time passes, immigrants master some English and their children tend to have much less difficulty. There is, of course, a continuing debate in this country about how and when to immerse immigrant children in an English-only curriculum.

Education

Education is the foundation for success in most societies.^[9] Those with an American education or its equivalent tend to do better than those not so fortunate, at least in the American market. For example, it is no accident that a high proportion of Indians immigrating to the United States are physicians or have PhDs; their skills are in demand.

Family ties

For immigrants, family is a major factor in the success of small- or microbusiness development. Family-owned businesses with no paid employees account for more than three-fourths of all immigrant enterprises.^[10]

The family can be viewed as a network of obligations that embodies the social, economic and cultural investments made prior to immigration, and that immigrants draw on and continue to invest in during the process of adaptation. The family endows each of its members with the backing of collectively-owned capital, a credential that entitles members to credit Thus, the family comprises a social network that can be effectively harnessed to achieve collective goals.

It is also the case that married people are much more likely to operate businesses than single people. But as immigrant businesses continue to develop or advance, family resources (e.g., financial, information, and counsel) tend to decline precipitously over time.^[5]

Social capital

After the family, ties with the broader ethnic community become important. These networks, referred to as *social capital*, are sources of business information, including markets, supplies, and opportunity. These networks, often discussed in the context of enclaves, are powerful resources. Networks, for example, so dominate business relationships that they even affect *intracommunal transfers* of businesses. One study in Los Angeles found that Koreans tended to sell their businesses to Koreans, and Hispanics to Hispanics, even though other native buyers were available.

Start-up capital

Obtaining start-up capital is an issue for most businesses.^[11] Immigrants largely obtain start-up or

expansion capital from friends, relatives, or acquaintances. Capital obtained is generally in small amounts. Once businesses begin to grow, immigrants seek funding from formal sources (e.g., banks, investors), often their countrymen. Capital may originate overseas or locally. In a study of Taiwanese entrepreneurs, 57% reported obtaining start-up capital from their homelands.^[12] In a study of Korean entrepreneurs, researchers found that Korean suppliers extended credit terms, lower prices, and easy access to information.^[5] Successive waves of Cubans fleeing Castro's government were offered *character loans*—no collateral required—by wealthy Cuban immigrant bankers who were willing to invest in refugees.^[13] Indian entrepreneurs in high-tech centers (e.g., Silicon Valley in California, Route 128 in Boston) are highly organized into self-employed networks that can raise millions for new business startups.^[11]

Market Opportunity

To be successful, ethnic entrepreneurs must be able to take advantage of market opportunities—*demand-side* factors. This translates into concentrated economic activity, economic linkages, and globalization.

Economic sector

Ethnic self-employed, regardless of where they originate from, are likely to start restaurants or drinking places. These places serve fellow immigrants and natives with a penchant for “foreign” cuisine. Also, construction companies are popular in most areas. Many worked construction in their home country or acquired skills as laborers, often unskilled, in this country. Many are nonunion shops, making them less costly and more competitive. But some immigrant groups specialize, often quite successfully. Taiwanese and Indians dominate, given their relatively small numbers, the high-tech, info-tech, hardware and software, and communications industries, nationally and internationally (see sections below). These cultures produce people with math, computer, and engineering skills. They come to the United States, receive even more specialized education, then start up or work for high-tech firms.

Linkages

Just as some immigrants exploit networks to enhance their capacity domestically, *transnational networks* or *communities*—cross-border business relationships—have opened up markets considerably over the past few decades.^[14] Saxenian studied immigrant entrepreneurs in California's high-tech region, Silicon Valley. She found that one-half of Chinese and Indian immigrants set up subsidiaries, joint ventures, subcontracts,

and/or other business operations in their home country. Half of those starting new businesses did so with two or more cofounders from their home countries. So prevalent are Chinese entrepreneurs in Monterey Park, California, who travel incessantly on business that they have been dubbed “astronauts.” Similarly, Chinese immigrants in Queens are widely known to have raised capital for businesses from contacts in Hong Kong and Taiwan.

Globalization

Globalization's effects on local economic development and immigration are only just now beginning to be understood.^[15] One factor facilitating international migration is the *liberalization*—removing barriers—of trade. Labor is free to move in search of job opportunities as economies open up. In advanced economies, governments have an interest in opening up their markets to imports from less advanced economies. This allows poorer countries to create wealth through exports that in turn allows them to accept imports and keeps workers at home, gainfully employed. The North American Free Trade Agreement (NAFTA) of 1993 between Mexico, Canada, and the United States was an attempt to remove most trade barriers and in the process stem the tide of immigration.

WHAT ARE THE IMPACTS?

Local policy makers want to know how immigrant economic activity affects economic development and how social factors affect economies. In spite of the increasing prevalence of immigrant populations in America, both positive and negative consequences of immigrant economic activity are little supported by solid research, greatly misunderstood or misrepresented, and highly controversial.

Economic Development Impacts

Immigrant entrepreneurs create jobs, build wealth (both private and public), pay taxes, build communities and local economies, and help needy relatives back home. But some also displace native-born Americans; exploit workers from their own culture; participate in the hidden, illegal, or underground economy; and engage in criminal behavior and corruption, in the process retarding community building. It is important not to stereotype people. No group is entirely positive or negative, and we suspect that most people contribute positively whether they are immigrants or natives. It is also important not to exaggerate contributions, either positive or negative.

Growth

At the macro level, recent immigrants increased the value of U.S. gross domestic product (GDP) from \$6 billion to \$20 billion, depending on the study.

Even though relatively modest as a proportion of a multitrillion dollar economy, the amount that immigrants have contributed to economic growth is a matter of controversy in the literature, in part having to do with who should be considered an immigrant? Many studies consider foreign-born people living in the United States to be immigrants, but their American-born children to be natives. Other studies include, or at least do not explicitly exclude, second and even third generations of immigrant parents in the growth calculation, arguing that subsequent generations would not have been entrepreneurial but for their parents.

In 1997, using narrow definitions, self-employment was equally prevalent among natives, 11.8%, and recent immigrants, 11.3%. Immigrants are neither more nor less likely to be self-employed than natives. Using somewhat broader definitions, according to the U.S. Small Business Administration (SBA), between 1988 and 1998 numbers of self-employed Asians rose by 56.5%, Hispanics by 30.1%, and Whites by 1.1%. Whites account for 90.4% of self-employed people. The growth in self-employment of 3.9% nationwide, then, is based on immigrant, as broadly defined, small-business activity.

Data, over time and across generations, show that immigrants contribute significantly to the economy. The SBA estimates that in 1997, Asian self-employed people generated \$275 billion in revenues and employed 1,917,244 workers, whereas Hispanics produced \$184 billion and employed 1,492,773 workers. Like most small businesses, the lion's share, 87%, of immigrant businesses employ 10 workers or less. Some immigrant groups contribute economically despite their small numbers. In a study of Silicon Valley high-tech firms, Saxenian^[14] found that 24% had either Chinese or Indian executives. About 2000 firms were managed by Chinese entrepreneurs and another 774 firms by Indians. These companies had \$16.8 billion in sales and employed 58,000 workers. Companies averaged between \$200,000 and \$300,000 in sales per employee.

In spite of their contributions to economies, immigrant economic activity has been criticized by some observers.

Remittances

Remittances are transfer payments made by immigrants living in the United States to relatives living in their country of origin. Some critics worry over money being sent overseas. A report on President Bush's visit to Miami in May 2002 by National Public Radio

estimated that Cuban exiles in South Florida send about \$500 million to Cuba annually. Castro's government receives more money from remittances than from exports of sugar and nickel, and net payments for tourism. The Inter-American Development Bank estimates that immigrants will send about \$300 billion home over the next 10 years. The World Bank estimates that Latin America and the Caribbean receive about \$15 billion in remittances annually.

According to the World Bank, immigrants earning less than \$20,000 annually send as much as \$3000 home each year. Remittances slow economic progress for those compelled to send money home. But the longer immigrants remain in the United States, the less likely they are to continue sending money home.

Some consider the decline in U.S. foreign aid to developing countries in recent years to be more than compensated for by remittance transfer payments. Ironically, according to studies by the World Bank, remittances do not reach very poor people; those better off tend to receive transfer payments. As a consequence, remittances and foreign aid may not be a comparable trade-off.

Self-employment activity in transnational communities yields important economic development benefits for developing countries that may in turn benefit America. Residents of Ticuani, Mexico, for example, needed a new town water system, a facility their local government could not provide.^[16] Residents petitioned relatives in Brooklyn who in turn formed a committee, referred to as a *home town association*, to raise funding, referred to as a *community remittance*, for the water plant. Provision of the new water plant only occurred because of the continued linkage between Mexican-American entrepreneurs and villagers. Potential positive benefits of transfer payments are that future business development ties may be forged, but such linkages have not been studied.

NAFTA

Critics of NAFTA argue that immigration has not abated and that jobs have been lost in America. Although accurate figures are unavailable, as many as 1 million U.S. jobs, actual and potential, may have been eliminated to NAFTA since the 1990s.^[15] Job loss went relatively unnoticed because of the unprecedented growth in the American economy over the decade. Nonetheless, some communities and states felt NAFTA job loss much more than other.

Intergroup competition

Self-employment activity by immigrants might take away opportunities for natives, African Americans being most at risk because they may compete directly with immigrants in some communities. Immigrants working

together to erect barriers prevent natives from firmly competing. Fairley and Meyer,^[17] in a study of 132 largest metro areas, found that immigrant self-employed appear to displace natives, keeping them from starting new businesses. They found that immigrant self-employed did not displace African Americans overall. But a study by the National Academy of Sciences found some displacement for African Americans living in areas with high concentrations of immigrants. Presence of immigrant self-employed did not affect earnings of natives.

Immigrants might displace or impede self-employment for other immigrants. Having a common language, for example, does not guarantee that economic benefits will accrue. In Miami, although Spanish-speaking Cubans dominate, Spanish-speaking Mexicans and other Latinos have been excluded from the larger Cuban economy for cultural reasons.^[18]

Low wages

Since the 1990s, immigrant earnings have steadily declined relative to natives, with natives earning one-third more than immigrants. This is largely because recently arrived immigrants are unskilled and poorly educated compared with their predecessors. Being unskilled, they are forced to work for less. This depresses the wages of native unskilled labor that must compete with immigrants willing to work for less, but drives up wages for skilled workers who become scarcer. It takes about 16 to 20 years for an immigrant to reach wage parity with a native, after controlling on education and English fluency.^[1]

Studies are mixed as to whether immigrant self-employed exploit fellow immigrants by paying them low wages. Some argue that immigrants take advantage of recent arrivals because they lack education, language skills, experience, and capital to either make it on their own or to work for higher wages. Where immigrant populations are concentrated, workers become vulnerable to exploitation. Others argue that immigrant workers eventually do as well as natives, and in many cases better (e.g., Taiwanese). In the worse case, highly concentrated areas act as incubators, helping to launch people into better jobs as they acquire skills, and in many cases into self-employment as they accumulate experience and capital. Likely there is some truth in both observations. But it is certain that migrant farm workers, illegal aliens, and unpaid family members may not always prosper in America.

Hidden economy

Many people believe that some immigrant groups are prone to illegal activity, either because they come from a culture that condones such behavior or because

they adopt this activity out of necessity. Immigrants, especially illegals, from poor countries may be forced into the *hidden* or *underground* economy just to get a start in America. Miami is awash with “handymen” from Latin America who operate home repair businesses, some quite large, all outside the view of the tax collector and business regulators. Many of these businesses produce low-quality work. When the supply of cheap labor dried up after the Mariel Boatlift, Cuban entrepreneurs in the garment industry “contracted” with Cuban immigrants to do “homework” for low wages, no benefits, and employment instability.^[13]

The hidden economy is large, but there are no good, reliable estimates. And, it is not possible to parcel out underground activity among natives, foreigners and immigrants. One study of the informal economy in Hialeah, Florida, a city with a high concentration of Cuban immigrants, estimated that there was a \$300 million off-the-books economy in operation.^[13]

Criminal Activity

Immigrant populations are “popularly” associated with crime and corruption.^[13] *Proposition 187* in California, for example, asserts that the people of California have suffered personal injury and damage caused by the criminal conduct of illegal aliens in the state. In 1999, cities spent about \$1.5 billion incarcerating illegal aliens for criminal activity. In Miami, many immigrant businesses are deeply tied to the illegal drug trade, either as fronts for money laundering or as investments by “drug lords” in legitimate businesses. Many consider Miami the drug capital of the world, taking in \$50 billion annually. Eventually, illegal drug money ends up in off-shore banks where it cannot be accounted for. In 1979, the Federal Reserve Bank in South Florida had a cash surplus of \$3.2 billion, an amount representing 77% of cash surplus nationally. Immigrants from former Soviet block countries have fallen easily into criminal activity. The Russian mafia is now prominent in the United States and indeed worldwide, where they exploit networks once dominated or created by the KGB.

Even though crime among immigrants may be perceived as prevalent, research shows that recent inflows of immigrants have had no effect on crime rates in cities where they are heavily concentrated.^[19] Youth born abroad were significantly less likely to be involved in crime than native youth.

Social Impacts

Immigrant economic activity also creates positive and negative social impacts in communities that in turn affect economics.

Ethnic contributions

No one would disagree with the proposition that immigrants have enriched American culture. Indeed, if there is an American culture, then it is an amalgamation of the contributions of disparate ethnic groups. This is a positive legacy, but there are possible negative consequences.

Ethnic conflict

Immigrant populations either dispersed across the country and assimilated, isolated themselves in enclave neighborhoods or communities, or took over communities once dominated by native-born Americans. Cubans in Miami and Taiwanese in Los Angeles are illustrative of the latter. Cuban refugees descended on Miami in 1959, after Castro took over Cuba in a Communist coup. This takeover was followed by the 1980s Mariel Boatlift. Now, in 2002, Cubans make up 30% of the population. Many Cubans chose not to assimilate, maintaining Spanish as the language of business and private life, preserving Cuban culture including preservation of the Santería religion, and remaining isolated in large enclaves.^[13] Taiwanese took over large portions of Los Angeles not by gradually moving from the central cities to better neighborhoods, then to the suburbs, but by leap frogging to the suburbs directly.^[12] Their wealth allowed them to bypass the usual migration patterns prevalent among other immigrant groups.

When immigrant populations take over communities, they tend to preserve their own culture sometimes at the expense of those already there. In Miami, Cuban cultural change fostered an *English First* movement whose goal was to quell the rise in Spanish speaking. When that and other counterattacks on Cuban culture failed, Anglos began moving north in large numbers. This in turn caused many Cubans to further eschew assimilation.

Taiwanese immigrants formed “Little Taipei” in Monterey Park, California. Because of their wealth, home and commercial property prices began to skyrocket. Taiwanese were criticized for speculating in land—“The Taiwan Syndrome”—they intend to sell later. This practice has so driven up land prices that high-tech workers cannot afford to live there, creating a labor shortage. Anglos also blame Taiwanese for unplanned development, traffic congestion, and loss of community. Anglo businesses have gradually fled, as have many of its original residents.

It should be noted that the history of the United States is one in which one immigrant group successfully displaces another, either from abroad or internally. In Miami, Jews were replaced by Cubans who are now being replaced by Central Americans. Who knows what group will emerge next?

Ghettos

More often than not, poor immigrants, especially illegals, are forced either into distressed neighborhoods where housing is inexpensive or into more prosperous neighborhoods where multiple families share single-family dwellings. “Little Haiti” in Miami is a poor neighborhood dominated by Haitians and Dominicans. By contrast, the City of Miami Beach, the playground of the idle rich, also houses poor families in single rooms who survive by pooling resources. When the City of Miami faced a fiscal crisis in 1996, officials set out to identify illegal residences and collect taxes.^[13] About 4500 illegal residences were found and \$2.1 million in property taxes collected. High concentrations of poor people make neighborhoods depressed until they acquire resources either to move on to better accommodations or to revitalize it themselves, a rarer occurrence. But poor neighborhoods are the price societies pay to have inexpensive labor who produce low-cost goods and services.

Costs and Benefits

The problem with cost-benefit studies of immigration is that they are “questimates” with dubious assumptions and scant data yielding conflicting conclusions. Both pro- and anti-immigration advocates cite the same “definitive” National Academy of Sciences (NAS) report as evidence in favor of their conclusions, for example. Overall, NAS estimated that immigrant contributed between \$1 billion and \$10 billion in benefits to society, while suggesting that they cost \$15 billion to \$20 billion to taxpayers. In another study by the National Research Council, in 1997, immigrants made tax payments in excess of \$80,000 over the cost of services received. When only the first-generation immigrants are taken into account, recent immigrants consume more in services than they pay. Those who are poorly educated and unskilled tended to consume \$13,000 more in services than they paid in taxes. About 22% of immigrant households are on welfare, as compared with 15% for natives. Another study looking at gross domestic product (GDP) found that it cost native workers about \$72 billion annually or 1% of GDP.^[20] The only certainty about costs and benefits is their uncertainty. We agree with the following conclusion:^[21] “There may be costs for native workers in America, yet the gains to immigrant workers far outweigh them.”

LOCAL POLICY ISSUES

For economic development professionals, like it or not, immigrant populations necessitate a greater role

in, or at least an awareness of, national and even international factors (e.g., NAFTA), that now affect local economies. Likewise, greater attention must be devoted to cultural issues, social services, housing, and the like that accompany immigration; areas that have not traditionally been the province of economic development.

Within this broad context, what are the policy implications of immigration, self-employment, and economic development? Economic development professionals basically face the same issues as they would if only natives were involved, but with some twists. Should self-employed people be left to fend for themselves, or should the public sector assist them through technical assistance, training, or lending? If the public sector assists them, should immigrant businesses be targeted, or should they compete for services directly with natives? If immigrants are targeted, should immigrants control the services they receive?

Other articles in this encyclopedia address options for and benefits of facilitating economic development, so they need not be reviewed here. The real issue is targeting. If immigrant businesses represent culturally separate economic activity, then assistance to them must likewise be customized. Training Spanish speakers in English, for example, may not be effective. The best way to customize assistance is to provide funding and resources to ethnic communities and then allow them to meet their own needs. But, self-determination should occur within a broader city- or regionwide economic development strategy, of which an ethnic community is a part.

REFERENCES

1. Orrenius, P.M.; Viard, A.D. The second great migration. *Southwest Econ.* **2000**, 7 (3), 1–8, May/June.
2. Bonacich, E. A theory of middleman minorities. *Am. Sociol. Rev.* **1973**, 38, 583–594.
3. Portes, A.; Bach, R. *Latin Journey*; University of California Press: Berkeley, 1985.
4. Razin, E.; Light, I. Ethnic entrepreneurs in America's largest metropolitan areas. *Urban Aff. Q.* **1998**, 33, 332–360, January.
5. Yoon, I.-J. The changing significance of ethnic and class resources in immigrant businesses. *Int. Migr. Rev.* **1991**, 25, 303–331.

6. Waldinger, R.; Aldrich, H.; Ward, R. *Ethnic Entrepreneurs*; Sage: Newbury Park, CA, 1990.
7. Raijman, R.; Tienda, M. Immigrants' socioeconomic progress post 1965. In *The Handbook of International Migration*; Hirschman, C., Ed.; Russell Sage Foundation: New York, 1999; chapter 13.
8. Raijman, R. Mexican immigrants and informal self-employment in Chicago. *Human Organ.* **2001**, 60 (1), 47–55.
9. Bureau of International Labor Affairs. *The Triennial Comprehensive Report on Immigration*; US Department of Labor: Washington, DC, 1999.
10. Light, I.; Har-Chvi, H.; Kan, K. Black/Korean conflict in LA. In *Managing Social Conflicts*; Dunn, S., Ed.; Sage: Newbury Park, CA, 1994; 72–87.
11. Buss, T.F. *Capital, Emerging High Growth Firms and Public Policy*; Praeger: New York, 2001.
12. Tseng, Y.-F. Beyond "Little Taipei." *Int. Migr. Rev.* **1995**, 29, 33–58.
13. Franco, A. *The Impact of Miami Cuban Exiles*; Florida International University: Miami, 2001. Dissertation.
14. Saxenian, A.L. *Local and Global Networks of Immigrant Professionals in Silicon Valley*; Public Policy Institute of California: San Francisco, 2002.
15. Scott, R.E. *NAFTA at Seven*; Economic Policy Institute: Washington, DC, 2001; briefing paper.
16. Portes, A. Global villages. *Am. Prospect* **1996**, 74–77, March–April.
17. Fairley, R.W.; Meyer, B.D. *Does Immigration Hurt African American Self Employment?* Northwestern University, Institute for Policy Research: Chicago, 2001.
18. Zsembick, B.A. *The Cuban Ethnic Economy*; Division of Immigration Policy and Research, US Department of Labor: Washington, DC, 1996. Working Paper # 27.
19. Butcher, K.F.; Piehl, A.M. Cross-city evidence on the relationship between immigration and crime. *J. Policy Anal. Manage.* **1998**, 17, 457–494.
20. Davis, D.; Weinstein, D. *Technological Superiority and the Losses from Migration*; Department of Economics, Columbia University: New York, 2002.
21. A price worth paying. *Economist* **2002**, 364, 72, June 1.

BIBLIOGRAPHY

1. Fairley, R.W.; Meyer, B.D. *The Effect of Immigration on Native Self-Employment*; National bureau of Economic Research: Cambridge, MA, 2000. Working Paper 7561.

Economic Development, Infrastructure and

Robert Krol

Department of Economics, California State University, Northridge, California, U.S.A.

INTRODUCTION

An economy's infrastructure provides valuable services to private firms and individuals. Infrastructure investment can influence the growth, level, and spatial distribution of economic activity. The bulk of the research since the 1990s has focused on the impact of public capital investment on productivity and overall economic development.

Initial studies suggested that public capital had a large impact on U.S. economic activity. More recent studies found only a modest overall impact on the economy. The low net return to public capital investment reflects diminishing marginal returns to infrastructure construction, the spatial reallocation of economic activity, higher tax rates needed to finance projects, and reductions in agglomeration economies.

This article is organized in the following manner. First, U.S. infrastructure trends are reviewed. What follows is a discussion of how infrastructure influences the economy. Key empirical studies that examine the impact of public capital on the economy are reviewed. Special attention is given to reasons why the estimated net impact in recent studies is so small. Finally, policies to improve the efficiency of infrastructure provision and its use are discussed.

PUBLIC INFRASTRUCTURE TRENDS

This section discusses the composition and trends in the public infrastructure stock of the United States. To get a sense of public infrastructure trends, [Fig. 1](#) provides plots of a quantity index for federal-state & local fixed assets (the public capital stock) from 1960 to 1999. The non-military public capital stock data used in this section includes equipment and software, structures, highways and streets, conservation and development, sewer systems, water supply, and other structures. The data comes from the U.S. Department of Commerce *Fixed Assets and Consumer Durable Goods in the United States*.

Both indices trend upward over this period. The slowdown in federal infrastructure investment occurred in the late 1960s. This coincided with the completion of the interstate highway system. State and local infrastructure accumulation has not slowed significantly over this period.

[Table 1](#) provides information on the composition of non-military government fixed assets for selected years between 1960 and 1999. The shares do not change radically over the period. For the federal government, the exception was equipment and software, which more than doubled over the period. This is no surprise, as the government significantly increased its use of computers, beginning in the late 1980s and continuing into the 1990s. There is a noticeable decline in the federal conservation component over the period. Structures and highway's shares remained fairly stable during this period. The composition of state and local fixed assets appears more stable than for the federal government. Once again, the clear exception is the equipment and software component of state and local infrastructure.

State and local governments own the vast majority of the public infrastructure stock. As shown in [Table 1](#), the federal government owned approximately 18% of total non-military government fixed assets in 1960. By 1999, the federal government's share had declined to roughly 13%.

INFRASTRUCTURE'S IMPACT ON ECONOMIC ACTIVITY

Public infrastructure is made up of public goods typically subject to economies of scale, such as roads, sewage treatment facilities, and water supply systems. Increases in infrastructure can have a significant impact on economic activity in various ways. The effects on output can be direct and indirect.^[1] In the case of a direct effect, the infrastructure provides intermediate services to the private sector. For example, a new or wider road can reduce congestion, reducing travel time. Local businesses can now make more deliveries with the same number of trucks and drivers, resulting in higher output. Investments by firms in additional trucks would have produced the same result—an increase in the number of deliveries the firm is capable of making.

The indirect output effect works through the impact of greater infrastructure on the marginal product of private inputs hired by firms. If a complementary relationship exists between public capital and private inputs, increases in public capital result in a higher marginal product of private inputs. The marginal

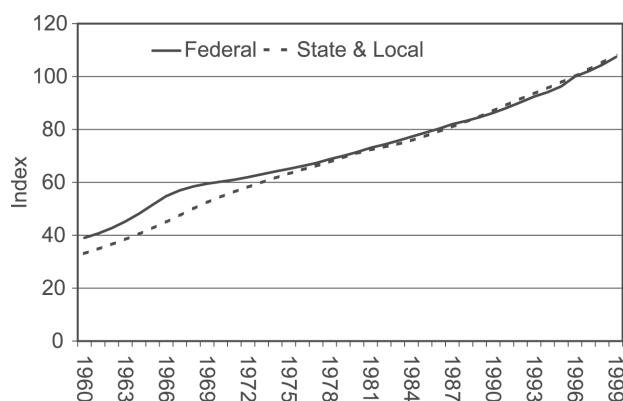


Fig. 1 Public infrastructure stock.

product of the private inputs rises because they have more public capital to work with in the production process. The higher marginal product of private inputs raises the firm's demand for those inputs. Continuing with the road example, the improved road system makes existing trucks and drivers more productive. As a result, the firm buys more trucks and hires additional drivers, causing the firm's output to rise.

The previous example illustrated a complementary relationship between public and private capital. Alternatively, higher public capital can substitute for private capital, reducing private inputs in production. For example, a city can decide to produce electricity itself, rather than have it produced privately. In this case, public capital crowds out private capital. It may even reduce economic activity if government is a less efficient provider.

New infrastructure can influence the spatial distribution of economic activity in a region.^[2] The building

of a new road can cause businesses to relocate to take advantage of lower transportation costs and greater customer flows. The new road raises the marginal product of private inputs near it, creating an incentive for the private inputs to relocate closer to the road. Economic activity increases near the new road, but declines in areas away from the road. In this case, the road provides only localized benefits to firms near the road. As a result of spatial reallocation of economic activity, net output increases little in the region as a whole.

AGGREGATE AND REGIONAL EMPIRICAL STUDIES

This section reviews empirical studies that have played an important role in shaping the literature on infrastructure investment (see Ref.^[3] for more details). Much of the research in this area was stimulated by a controversial 1989 study by David Aschauer.^[4] Aschauer estimated an aggregate production function for the United States using annual data for the period from 1949 to 1985. He found public capital to have a large positive impact on total factor productivity and the productivity of private capital. In addition, he showed that the slowdown in the accumulation of public capital during the 1970s contributed significantly to the overall decline in U.S. productivity growth during the 1970s and 1980s.

John Tatom^[5] questioned Aschauer's results on three counts. First, Aschauer ignored significant time trends in the data; this could lead to a spurious finding of a relationship between public capital and

Table 1 Composition of infrastructure stock

	1960	1970	1980	1990	1999
Federal					
Equipment and software	0.057	0.094	0.075	0.108	0.131
Structures	0.513	0.505	0.52	0.516	0.513
Highways and streets	0.027	0.034	0.037	0.034	0.028
Conservation and development	0.397	0.354	0.352	0.321	0.29
Other structures	0.001	0.012	0.015	0.021	0.038
State and local					
Equipment and software	0.016	0.017	0.019	0.029	0.032
Structures	0.604	0.601	0.604	0.605	0.612
Highways and streets	0.249	0.258	0.229	0.214	0.21
Conservation and development	0.001	0.009	0.011	0.012	0.011
Sewer systems	0.047	0.042	0.055	0.054	0.052
Water supply	0.043	0.034	0.036	0.035	0.034
Other structures	0.034	0.038	0.047	0.052	0.05
Federal/total	0.183	0.161	0.146	0.138	0.13

productivity when none is present. Second, Aschauer ignored significant energy price changes that might have affected productivity during the sample period. Tatom found that including energy prices in the model and adding time trends reduced the measured impact of public capital on productivity by one-half. Tatom also provided evidence that some of the variables in the model are not stationary. Because of this, the model should be estimated in first differences. When Tatom estimated the model in first differences, public capital is no longer significant.

John Fernald^[6] reexamined the link between public capital and productivity. Fernald used a growth accounting approach to assess the impact of the growth in highway construction on U.S. industry-level productivity growth for the period from 1953 to 1989. He tested whether the increase in the highway capital stock had a disproportional impact on productivity growth in industries that are vehicle intensive. He found this to be true.

His results explained the productivity slowdown in the United States in the 1970s. Road construction was unusually productive before 1973, but not since that time; road construction today provides below normal returns. In other words, building the interstate highway system was exceptionally productive, but a second one would not provide comparable returns.

Rather than focusing on aggregate measures, some researchers began examining the impact of public capital on the economy at the state and local levels. Analysis at the state level increases sample size, improving statistical inference. It also makes sense because decisions and ownership of the public capital stock are primarily at the state and local levels.

Alicia Munnell^[7] constructed estimates of state-level public capital to examine the relationship between infrastructure and economic development. Munnell estimated production functions using state level data for the period 1970 to 1986. She found that a 1% increase in a state's public capital stock raises per-capita real gross state product by 0.15 of a percent point. This estimate is smaller than some aggregate results, but is statistically significant.

Munnell's results were challenged by Holtz-Eakin^[8] and Krol.^[9] Both researchers pointed to the fact that Munnell failed to control for unobserved differences in state production functions. Ignoring these differences can significantly bias estimates of public capital productivity. Both Holtz-Eakin and Krol found that once unobserved production function differences are controlled for, public capital is generally no longer a significant determinant of productivity. These results question the widespread view that large increases in state infrastructure investment are an important source of state economic development.

Morrison and Schwartz^[10] focused on the manufacturing sector, examining the impact of additional infrastructure investment on regional manufacturing costs. Estimating cost functions for regions of the United States for the period from 1970 to 1987, they found that additional public capital significantly lowers manufacturing costs in all regions. An additional \$1 million worth of infrastructure lowered manufacturing costs by about \$160,000 to \$180,000 per year. But a positive marginal product is not enough in and of itself to justify additional investment; there must be a positive net return after the costs of financing the investment are considered.

Tax and Spatial Issues

The studies discussed in the previous section indicate that additional infrastructure investment will lead to only modest increases in productivity and output. Three factors contribute to muting the net affect of public capital on productivity and output. First, financing additional infrastructure can result in higher taxes and interest rates. Second, new infrastructure causes economic activity to be reallocated, resulting in small net regional increases in output. Third, because new highways lower transportation costs, spatial density falls, reducing agglomeration economies.

New infrastructure projects must be financed through taxation or borrowing. Taxes levied on private economic agents distort decisions, resulting in a less efficient allocation of resources. In addition, the higher taxes raise the cost to businesses in the economy. The higher costs due to greater taxation can offset the lower production costs associated with the benefits of additional infrastructure.

Borrowing funds in capital markets to finance an infrastructure project can raise interest rates (interest rate effects are likely to be modest in global capital markets). As in the tax example, this raises business costs that can offset the benefits derived from additional infrastructure. In both cases, the net effect on business costs and production may be close to zero.

Morrison and Schwartz^[10] and Seitz^[11] found evidence to support this conclusion. Morrison and Schwartz found the annual benefits to businesses from additional public investment (lower costs) to be less than the annual marginal costs of funds in most years. Seitz found any increase in manufacturing employment resulting from additional infrastructure to be offset by the negative employment effects of the higher taxes.

The spatial reallocation of resources that results from the construction of a new highway was examined Chandra and Thompson.^[12] They investigated this relationship for non-metropolitan counties with new highway constructed during the period between 1969

and 1993. Rather than estimating a production function, they instead estimate a county earnings function, controlling for national, regional, and local factors that influence economic activity.

They find that the construction of a new highway in a county draws economic activity away from adjacent counties. Earnings in counties where the highway was constructed rise relative to counties that do not receive a new highway. Total earnings are 6% to 8% higher in the long run in the counties that get a new highway. Total earnings fall 1% to 3% in adjacent counties. Services and retail rise 5% to 8% in the county with the new highway and fall 8% to 11% in the adjacent county. They conclude the net effect of the new highway is about zero.

Boarnet^[12] confirmed this result, looking at California counties during the period from 1969 to 1988. He found negative spillovers in the case of the construction of new highways. He found the construction of a new highway to be positively related to within-county output, but negatively related to adjacent-county output. These two studies confirmed the idea that highway construction provides mostly localized benefits, with limited aggregate effects.

Furthermore, the construction of a new highway can reduce productivity because it can lower the benefits derived from agglomeration economies. By lowering transportation costs, the incentive for businesses to locate near one another falls. As a result, firms are in less of a position to capture the local geographic externalities that arise with agglomeration economies.

Ciccone and Hall^[13] found that if you double employment density in a state, average labor productivity rises by 6%. This supports the link between density and productivity. Haughwout^[14] found lower benefits to densely populated urban areas from highway construction. This result links highway construction and the associated lower transportation costs to reduced agglomeration economies. It appears that highway construction can reduce agglomeration economies, which lowers productivity and output growth.

Improving the Provision of Infrastructure

This section examines how infrastructure policy reforms can improve the efficiency associated with the construction and use of highways. Highways and roads represent one of the largest civilian public capital goods in the United States. Winston,^[15] and Boarnet and Haughwout,^[16] discussed how changing the way we finance and price our highways and roads can lead to large net welfare gains. These policy reforms would also moderate the demand for additional highway construction.

Highway construction is financed primarily using funds from the Federal Highway Trust Fund. In most cases, 80% of the funds used to build a new highway come from the federal government. The justification for the federal government's major role in financing highways was the presence of significant spillover benefits captured by individuals living outside the area where a road is built. However, the studies discussed in the last section indicated that highway benefits tend to be highly local.^[2,12] As a result, a case can be made for the elimination of federal financing of highway construction (repealing the federal gas tax and refunding the Highway Trust Funds to state and local governments).

Federal financing raises concerns because it introduces biases toward excessive highway construction. Federal funds focus primarily on construction, rather than maintenance. Funding programs ignore the higher return to road maintenance, compared with the low return on road construction. The reason for this bias is believed to be political. Voters reward elected officials for opening a new spur of a highway, but tend to take regular maintenance to an existing road for granted.

In addition, Boarnet and Haughwout^[15] pointed out that financing new highway projects with pooled tax revenues results in non-economic projects being funded. In many cases, the local benefits individuals capture from a new road exceed local costs. However, this does not make the project economic because the local benefits are often less than the total cost of building the road. As a result, total welfare is reduced when the road is built using federal funds. Instead, if local decision makers are required to finance their own roads, only those that can pass a full cost-benefit analysis will be constructed.

Congestion on many urban highways creates political pressure to build more roads. This approach has done little to solve the congestion problem. Congestion indicates the price of using the highway is too low. An alternative to the costly "build your way out of congestion" approach is to move to an efficient pricing system. More efficient use of existing highways would significantly reduce the need for additional highway construction.

Today, in the United States, a flat gas tax is used to finance highways. This tax does not solve peak-load congestion problems, nor does it accurately reflect the pavement wear associated with using roads. Instead, a toll can be levied directly on users of the highway. Efficient tolls are higher during rush hours. This gives drivers an incentive to take into account the longer commuting time associated with increasing highway congestion. Tolls can be administered electronically to minimize the inconvenience of payment for road users. The potential net welfare gain from imposing congestion tolls in the United States was

estimated to be almost \$4 billion per year in 1998 dollars.^[17]

Not only can highway construction be offset by cheaper maintenance spending, but also maintenance expenditures can be lowered by a tax that reduces highway damage. The current gasoline tax worsens highway wear and tear by trucks. Because the tax is on a per-gallon basis, trucks are built using fewer axles to improve mileage. However, highway damage depends on weight per axle, not total vehicle weight. As a result, the current tax system leads to greater highway damage than one based on a toll system that factors in weight per axle.^[15]

Road maintenance can also be forestalled by building thicker roads and highways. Winston^[15] argued that the adoption of efficient pricing and building thicker roads has a benefit-to-cost ratio of 4. This suggests that these reforms are economic and would result in a net improvement in welfare, if implemented.

On a less optimistic note, Winston^[17] argued that entrenched special interest groups, the construction industry, and public employee unions make serious transportation policy reform unlikely. The only way to get the needed reform is to privatize the construction, operation, and maintenance of the transportation system. He argued that market incentives are the only way to correct for government failure, and to improve the efficiency and quality of service.

CONCLUSION

In this article, I have argued that public infrastructure has both direct and indirect effects on the marginal products of private inputs. If the relationship between public and private inputs is complementary, public capital investment can raise productivity, employment, and output. When public and private inputs are substitutes, the impact of additional public capital on output is weaker.

The empirical evidence suggests that additional public infrastructure has only a modest impact on economic activity. The small impact is the result of diminishing marginal returns to public capital, the spatial reallocation of economic activity, higher taxation to finance the project, and reductions in agglomeration economies.

With respect to highways, many observers have pointed to alternatives to costly highway construction. Net welfare can be increased by using congestion pricing, setting user fees that more accurately reflect road

damage, building thicker roads, and by using cost-benefit analysis in project evaluation.

REFERENCES

1. Tatom, J.A. Should government spending on capital goods be raised? Federal Reserve Bank of St. Louis Review **1991**, 73 (2), 3–15.
2. Chandra, A.; Thompson, E. Does public infrastructure affect economic activity? Evidence from the rural interstate highway system. Reg. Sci. Urban Econ. **2000**, 30 (4), 457–490.
3. Krol, R. Public infrastructure and economic development. In *Handbook of Economic Development*; Liou, K.T., Ed.; Marcel Dekker: New York, 1998.
4. Aschauer, D.A. Is public expenditure productive? J. Monet. Econ. **1989**, 23 (2), 177–200.
5. Tatom, J.A. Public capital and private sector performance. Federal Reserve Bank of St. Louis Review **1991**, 73 (3), 3–15.
6. Fernald, J.G. Roads to prosperity? Assessing the link between public capital and productivity. Am. Econ. Rev. **1999**, 89 (3), 619–638.
7. Munnell, A.H. How does public infrastructure affect regional economic performance. In *Is There a Shortfall in Public Capital Investment*; Munnell, A.H., Ed.; Federal Reserve Bank of Boston: Boston, 1990.
8. Holtz-Eakin, D. Public-sector capital and the productivity puzzle. Rev. Econ. Stat. **1994**, 76 (1), 12–21.
9. Krol, R. Public infrastructure and state economic development. Econ. Dev. Q. **1995**, 9 (4), 331–338.
10. Morrison, C.J.; Schwartz, A.E. State infrastructure and productive performance. Am. Econ. Rev. **1996**, 86 (3), 1095–1111.
11. Seitz, H. Infrastructure, industrial development, and employment in cities: Theoretical aspects and empirical evidence. Int. Reg. Sci. Rev. **2000**, 23 (3), 259–280.
12. Boarnet, M.G. Spillovers and the locational effects of public infrastructure. J. Reg. Sci. **1998**, 38 (3), 381–400.
13. Ciccone, A.; Hall, R.E. Productivity and the density of economic activity. Am. Econ. Rev. **1996**, 86 (1), 54–70.
14. Haughwout, A.F. State infrastructure and the geography of employment. Growth Change **1999**, 30, 549–566, (Fall).
15. Winston, C. Efficient transportation infrastructure policy and pricing. J. Econ. Perspect. **1991**, 5 (1), 113–127.
16. Boarnet, M.G.; Haughwout, A.F. Do highways matter? Evidence and policy implications of highways influence on metropolitan development. Discussion Paper, The Brookings Institution Center on Urban and Metropolitan Policy **2000**, 1–30.
17. Winston, C. Government failure in urban transportation. Fisc. Stud. **2000**, 21 (4), 403–425.

Economic Development, State Government Administration of

Keith Boeckelman

Department of Political Science, Western Illinois University, Macomb, Illinois, U.S.A.

INTRODUCTION

Throughout much of U.S. history, economic development has been a critical state function. Activity in this area dates to colonial times, when states first began to offer land grants, loans, and tax exemptions to subsidize industry.^[1] During the nineteenth century, states used massive infrastructure investments, such as the Erie Canal, to foster growth. After a post-Civil War period of federal dominance in the economic realm, development issues reemerged as a concern for state policy makers during the twentieth century. Since the 1970s, economic development issues have been at or near the top of most states' policy agendas. Three approaches to economic development, known as the first, second, and third waves, characterize the contemporary era. Each is summarized and evaluated below. Following this discussion, current state efforts are examined and future trends are speculated.

It is easier to describe the various approaches to economic development than it is to determine exactly what states are doing in this area or whether their efforts have the desired impact. The fact that many programs exist off budget, use intergovernmental grants, or operate through public-private partnerships complicates description. Evaluating the results of economic development programs is also notoriously difficult. States have open economies subject to a large number of national and international variables. Isolating the impact of policies on the economy is therefore a tricky proposition. In addition, factors that supposedly affect economic development, such as labor costs, are in turn affected by levels of economic development, making it difficult to ascertain what is causing what. These complications can frustrate both elected officials and practitioners in their efforts to develop coherent and effective policies.

FIRST-WAVE ECONOMIC DEVELOPMENT POLICIES

First-wave economic development efforts, also referred to as the "supply-side approach" or "locational

policies," adopt a business recruitment model. These policies try to entice manufacturing plants to locate in a state by reducing taxes, wage rates, construction and infrastructure expenditures, or other costs. Tax incentives are the most common and widely publicized first-wave policies. They typically excuse businesses from corporate income or property taxes, or reduce levies on firms that engage in economically beneficial activities, such as buying new equipment or creating jobs in the state.^[2] Other common first-wave programs include industrial development bonds, right-to-work laws, and enterprise zones. State or local governments use the former to finance the construction or purchase of industrial buildings. Right-to-work laws forbid closed-shop provisions thus weakening the collective bargaining power of unions in order to depress wage rates. Finally, enterprise zones target various tax and other incentives to putatively depressed parts of a state.

Southern states pioneered first-wave policies, beginning in the 1920s.^[3] At about the same time, these states began to create economic development agencies to spearhead business recruitment efforts. By the late 1970s, however, economic difficulties in the Northeast and Midwest led to a nationalization of first-wave policies. This trend was perhaps most evident in the rivalry to attract foreign auto plants, such as Volkswagen, Mitsubishi, and Mercedes Benz. The competition for jobs in lean economic times led to a climate of desperation where incentive programs escalated into an "arms race" as states matched and exceeded the deals their neighbors offered.^[4]

The impact of first-wave economic development policies has been a matter of substantial debate, particularly with respect to the question of whether tax policies affect a state's economy. The answer seems to depend, at least in part, on research methodology. Studies based on interviews of corporate decision makers suggest that tax incentives play a negligible role in determining where businesses locate.^[5] At best, taxes may determine which states are eliminated from contention in the early stages of the decision-making process. For the finalists in the location sweepstakes, however, factors such as transportation

and labor costs become more important in determining which state ultimately wins.^[6]

Econometric studies, however, show a more significant role for tax policy. Overall levels appear to be especially important, although targeted incentives matter in some instances, according to this research. Specifically, high tax rates are likely to reduce employment, gross state product, manufacturing investment, and the number of business locations. Tax incentives are most important in affecting where capital-intensive manufacturers move.^[7] Yet, the number of jobs affected by tax policy is usually rather small.

Any analysis of incentive programs must also note that they often distort a state's tax system by treating similar businesses unequally, because some receive breaks while others do not. Furthermore, they are regressive because profitable companies receive the majority of concessions. In addition, widespread use of incentive programs complicates tax administration, as laws are likely to change more frequently. Tax incentives can also have negative consequences that spill over a state's borders and create a negative-sum game for the nation as a whole. In other words, the money that states spend attracting businesses are a waste of national resources because the plants would locate somewhere in the United States anyway. Thus, if all states increase incentives to respond to the competitive "arms race," they not only lose revenue but they also do not create new competitive advantages.

Economic development programs that focus on labor costs, such as right-to-work laws, affect plant location decisions, particularly for those firms that do not need to be near specific markets or suppliers.^[6] Obviously, however, such policies attract relatively low-wage jobs. Existing studies suggest that enterprise zone programs tend to target businesses that would operate in the area anyway, and rarely hire neighborhood residents.^[8] Evaluations of industrial bond programs show that about 60% of the projects they fund would have occurred anyway.^[2]

SECOND-WAVE ECONOMIC DEVELOPMENT POLICIES

Second-wave policies, also sometimes known as "demand-side" or "entrepreneurial" approaches, began to appear in the 1980s. They emerged due to concerns about the cost and effectiveness of tax incentive programs, as well as the perception that the jobs created by first-wave initiatives were of low quality. These programs emphasize nurturing existing local businesses and promoting entrepreneurial activity, rather than trying to attract companies from outside the state. Resembling Japanese-style industrial

policy, at least superficially, the state government's role can include targeting investment to specific regions and industries or fostering public-private partnerships.

Specific examples of second-wave policies include high-technology initiatives, venture capital efforts, and export development programs. High-technology programs have two basic goals. Some focus on promoting hi-tech jobs in general because officials view this sector as having great potential for growth. Other initiatives attempt to preserve the viability of existing industries through technological improvements. Texas' Advanced Research Program and Advanced Technology Program are examples of the first type, while the Ben Franklin Partnership in Pennsylvania typifies the second approach. Many states administer their high-technology programs through public research universities in order to draw on the expertise of science and engineering faculty.

Venture capital programs provide loans to small, entrepreneurial businesses that may not have access to funds through traditional methods. State pension funds often serve as a source of capital for such programs. Export development initiatives assist firms in developing export markets overseas. About 40 states maintain offices abroad to administer these programs. A few states have also tried to bolster economic performance by restructuring labor-management relations or targeting industries outside the high-technology realm, but these programs have been less widely adopted.

Because they are of more recent vintage, there are fewer independent evaluations of second-wave policies than of the first wave. Existing studies suggest that high-technology business incubator programs that provide space and shared services, such as secretarial and accounting assistance, are effective. Export development initiatives appear to have a greater direct economic impact than do venture capital efforts or university-based technology programs.^[9] The public-private partnerships that many states have created to administer second-wave policies raise unique accountability issues. For example, it is not clear whether such entities must obey "freedom of information" or "sunshine" laws that govern more typical state agencies.^[10]

THIRD-WAVE ECONOMIC DEVELOPMENT POLICIES

Doubts about the effectiveness of both first- and second-wave policies led to a new approach beginning in the early 1990s. "Third-wave" initiatives reflect a retreat from direct economic management in favor

of more general “capacity-building” activities. Specifically, states began to offer different programs and to organize them in new ways. Programmatically, the third wave emphasizes fundamentals, such as infrastructure, educational improvements, and job training. Organizationally, this approach favors giving non-profits some responsibility for economic development administration and cutting back on direct state spending. It also emphasizes greater coordination between state and local governments, and among officials in a particular region.^[11] Studies of the impact of third-wave policies suggest that funding highway infrastructure raises state incomes, but not necessarily employment levels, while educational spending has little impact.^[12] The organizational innovations associated with the third wave are difficult to measure and evaluate.^[11]

STATE ECONOMIC DEVELOPMENT POLICY TODAY AND IN THE FUTURE

The appearance of a new wave of economic development policies does not necessarily wash away its predecessors. Thus, most states use program elements of all three approaches discussed in this article. As noted, measuring levels of state commitment to the various waves of economic development is difficult, in part because so much activity occurs off budget. Nevertheless, it appears that states devote the greatest amount of resources to first-wave programs, which continued to grow at least through the 1990s.^[13] This continuing dominance reflects two political advantages that the later approaches to economic development cannot match. First, supporters of first-wave efforts are often powerful business interest groups. By contrast, second-wave policies such as technology initiatives tend to have few organized beneficiaries, while they foster opposition from labor unions who see few program benefits accruing to relatively uneducated production workers. Second, it is easy for the public to draw a causal connection between a tax incentive followed by a plant location, while the benefits of second- or third-wave policies are rarely as visible.

Despite their continued popularity, there have been increasing concerns about the possible misuse of first-wave programs, especially tax incentives. State officials fear that the arms race mentality results in temptations to “give away the store” by offering packages that exceed job benefits. For example, Alabama’s offered about \$300 million to entice Mercedes Benz to come to the state at a cost per job of about \$200,000 for the 1500 positions created. In addition, states have little recourse when companies receiving incentives fail to deliver on the promised job benefits.

States have adopted various policies in response to these concerns. Perhaps the most common are “clawback laws” that require companies to pay back all or part of incentive packages if the promised number of jobs do not materialize. States are often unwilling to enforce such provisions, however, particularly if they run the risk of antagonizing local industries who are in financial trouble. Meanwhile, the general public is largely oblivious to the cost of tax incentives, perhaps because they are not part of the normal budget process. In fact, many states do not know how much they spend on incentives.^[12] Another, albeit less used, alternative to protect the public interest is to require cost-benefit analyses before offering incentives. Finally, some states have adopted “non-aggression pacts,” whereby they collectively agree to limit their use of incentives. These agreements have usually failed, however, with some pacts lasting only a few months. Efforts to get the federal government to intervene to deescalate the incentive arms race have also been unsuccessful.

Even during the robust economic times of the late 1990s, state interest in economic development did not appear to wane. Thus, it is unlikely to do so in the current, more uncertain, climate. It is not clear if states will stick with existing strategies, or opt for new approaches leading to a fourth wave of activity. Some have suggested that the next logical step for state economic development policy is to address economic inequality issues.^[14] One example of this approach would be to integrate technology development programs with efforts to train former welfare recipients for new jobs. A few states already use their economic development agencies to implement the Temporary Assistance for Needy Families program associated with the 1996 welfare reform law. Getting involved in redistributive policy brings political headaches that economic development officials may want to avoid, however, so it is not clear how far this trend will go.

REFERENCES

1. Brace, P. *State Government and Economic Performance*; Johns-Hopkins University Press: Baltimore, 1993.
2. Buss, T.F. The effect of state tax incentives on economic growth and firm location decisions: an overview of the literature. *Econ. Dev. Q.* **2001**, *15* (1), 90–105.
3. Cobb, J. *The Selling of the South*; University of Illinois Press: Urbana, 1993.
4. Wilson, R.H. *States and the Economy*; Praeger: Westport, CT, 1993.
5. Donahue, J.D. *Disunited States*; Basic Books: New York, 1997.

6. Schmenner, R.; Huber, J.C.; Cook, R.L. Geographic differences and the location of new manufacturing facilities. *J. Urban Econ.* **1987**, *21* (1), 83–104.
7. Wasylenko, M. Taxation and economic development: the state of the economic literature. *N. Engl. Econ. Rev.* **1997**, 37–52, March/April.
8. Porter, M. New strategies for inner-city economic development. *Econ. Dev. Q.* **1997**, *11* (1), 11–27.
9. Bingham, R.E.; Bowen, W.M. The performance of state economic development programs. *Policy Stud. J.* **1994**, *22* (3), 501–513.
10. Fulton, W. The business of bringing in business. *Governing* **1996**, 9, 92, October.
11. Bradshaw, T.K.; Blakely, E.J. What are third wave state economic development efforts? From incentives to industrial policy. *Econ. Dev. Q.* **1999**, *13* (3), 229–244.
12. Fisher, R.C. The effects of state and local public services on economic development. *N. Engl. Econ. Rev.* **1997**, 53–65, March/April.
13. Chi, K.S.; Leatherby, D. *State Business Incentives: Trends and Options for the Future*; Council of State Governments: Lexington, 1997.
14. Bozeman, B. Expanding the mission of state economic development. *Issues Sci. Technol.* **2000–2001**, *17*, 33–36, Winter.

Economic Development: Governance Concepts and Issues

Kuotsai Tom Liou

Department of Public Administration, University of Central Florida, Orlando, Florida, U.S.A.

INTRODUCTION

This article examines the importance of governance in the process of economic development. The article first provides a literature review about the role of government in promoting economic development and major concepts of good governance emphasized in recent studies. Then, the focus is on three policy changes and issues related to the topic of governance and economic development: 1) regulatory reform; 2) privatization policy; and 3) decentralization policy. To conclude discussions about the continuing arguments about the role of government and reforms, the increasing importance of local governments, and the contribution of non-governmental or non-profit organizations in economic development are presented.

OVERVIEW

Government has traditionally played an important role in the process of economic development. From the perspective of public policy analysis, the importance of government to economic development is primarily based on three inter-related roles, functions, or activities. First, government plays a pivotal policy role in identifying and formulating important development policies and programs to promote economic growth and social change. Next, government serves a critical managerial role in assuring the successful institutional development and reform to implement many identified development policies and programs. Last, government provides a supporting role in maintaining positive and stable political and social environments to provide the opportunity and time for the success of development policies and programs.

The study of governance is important to economic development researchers and policy makers. For researchers, the challenge is to systematically collect objective data from different countries and regions to test, verify, or create various development concepts, strategies, and theories. For policy makers, the challenge is to consider different environmental factors in their countries and to design an appropriate model of governmental intervention through the analysis of many development theories, strategies, and experiences. The interaction of researchers and policy makers is critical

to the success of economic development because of the significant consequence of negative development results.

FROM GOVERNMENT INTERVENTION TO GOVERNANCE COORDINATION

The important role of government in the modern economic system has been well recognized and accepted in the literature of public finance, development economics, and political economy. For example, with regard to specific government functions and activities in the market economy, Adam Smith emphasized three duties of the government (state): 1) protecting the society from violence and invasion by other societies; 2) establishing an exact administration of justice; and 3) erecting and maintaining certain public works and institutions.^[1] From the public finance perspective, Musgrave and Musgrave promoted three functions of government: 1) the allocation of social and other goods that the market will not of itself provide; 2) the equitable distribution of income and wealth; and 3) the stabilization of unemployment and control of inflation.^[2] Stern summarized the related literature by identifying five different arguments for governmental intervention in the economy. They were as follows: 1) a concern for market failure (arising from sources such as externalities, public goods, and imperfect information); 2) a concern to prevent or reduce poverty and improve income distribution; 3) the assertion of a right to certain facilities or goods (e.g., education, health, and housing); 4) the importance of paternalism (e.g., relating to education, pensions, and drugs); and 5) the rights of future generations (e.g., concerns about the environment).^[3] The importance of government role in economic development can be discussed in the following two areas of government intervention and governance coordination.

Government Intervention

The issue of state or government involvement in economic growth has been further emphasized and debated in the literature related to economic development. First, the neoclassical arguments of development emphasize the role of foreign trade and investment and the importance of a free market in stimulating

competition during the development process.^[4-7] They claim that the problems of LDCs result from extensive government intervention in promoting import-substitution policies that limit the scope of industrialization. They argue that one of the major factors contributing to the success of East Asia's newly industrialized countries is the adoption of export-oriented policies that encourage the process of technological adaptation and entrepreneurial maturation. They recognize the role of state in the process of development but emphasize a passive and limited role of government in such activities as maintaining stability and providing physical infrastructure.

Next, statist arguments of development indicate that the successful experience of newly industrialized countries is related not only to the operation of the free market but also to the active role of government in directing public and private resources to change the structure of their economy.^[8-11] For example, many of the successful newly industrialized countries emphasize a general incentive policy to encourage the accumulation of production factors (tax measures and research and development) and an industrial targeting policy to promote the growth of particular industries (e.g., subsidizing credit or import protection).

Considering both the market and the state arguments, Aoki, Kim, and Fujiwara promoted a market-enhancing view, which emphasized the role of government policy to facilitate or complement private-sector coordination.^[12] They maintained that the previous approaches viewed the market and government as the only alternative and as mutually exclusive substitutes, and that the market-enhancing approach stressed the mechanisms whereby government policy is directed at improving the ability of the private sector to solve coordination problems and overcome other market imperfections. One example of the mechanisms is the important role of government-private-sector intermediaries (e.g., deliberation councils and national wages council) in facilitating information change to avoid possible market coordination failures and in realizing an efficient allocation of credit under the condition of information asymmetry.

Many studies of economic structure adjustments in developing countries and economic reforms or transitions in postsocialist nations also address the issue of governmental involvement in economic development. With regard to structural adjustments, for example, researchers emphasize the importance of the administrative and political functions of the state in the successful implementation of an adjustment program or a reform policy.^[13,14] Similarly, many studies of economic transitions in postsocialist countries emphasize the issue of stabilization in the process of economic transformation and identify the important role and

involvement of government to assure the success of the transition.^[14-17] In summary, Liou identified five major roles of government in the process of China's post-Mao economic reforms, including: 1) government as a promoter of economic growth (i.e., formulating and implementing growth policies); 2) government as a manager of the economy (i.e., managing fiscal and monetary policies); 3) government as a distributor of income (i.e., distributing resources between different regions and among different social groups); 4) government as a regulator of industry (i.e., regulating for social and economic reasons); and 5) government as a protector of citizens and business (i.e., protecting citizens and business from foreign attacks as well as criminal and illegal activities).^[18]

Governance Coordination

Since the 1990s, many international aid organizations have recognized the problems of governance in their efforts to promote economic and social development in many developing countries.^[19] The governance problems come from two sources: the abuse of public funds by local elite groups and the ignorance of local political problems by the international donors. The donor organizations believe that good governance is necessary for economic development in these countries. They promote good governance concept and emphasize the need for governance reform in these countries.

Despite its popularity, the concept of good governance has been related to different definitions and elements. World Bank, for example, indicated that governance refers to the exercise of the power of the state in managing a country's social and economic resources, as well as the related mechanisms for public accountability, rule of law, transparency, and citizen participation. They identify four specific elements of governance: 1) accountability (i.e., officials being answerable for government behavior), 2) participation (i.e., the involvement of citizens in the development process), 3) predictability (i.e., legal environment being conducive to development), and 4) transparency (i.e., the availability of information to the public and the clarification of government rules, regulations, and decisions).^[20,21]

Pierre and Peters explained that governance is about government's changing role in society and its changing capacity to pursue collective interests under severe external and internal constraints. They also identified four elements of governance: 1) the importance of networks (i.e., the use of networks to dominate public policy); 2) from control to influence (i.e., government's influence through a continual process of bargaining and persuasion); 3) blending public and private resources (i.e., the importance of network framework);

and 4) use of multiple instruments (in developing and implementing public policies).^[22,23]

Besides these basic concepts and definitions, Shepherd pointed out the importance of other governance areas to economic and social development, including, for example: public expenditure patterns and revenue growth, local institutional development and rural local development, inclusiveness and access into sector reforms, the elimination of bad government, and the incorporation of civic society and Non-Government organizations.^[24] Similarly, Kliksberg also emphasized other public management capacity to support economic and social development, including: the importance of coordination between economic and social policy, the improvement in intraorganizational coordination, the need for decentralization, the potential of participation, and the renewal of organizational structure.^[25]

In sum, the governance approach to economic development changes the traditional dominant role of government operation and welcomes the participation of private and non-governmental organizations and alternative methods of operation. Network coordination between government agencies and these organizations is one of the major requirements for the success of economic development.

CHANGES IN REGULATION, PRIVATIZATION, AND DECENTRALIZATION POLICIES

To assure the success of economic growth, many researchers of economic development recognized the importance of an efficient and ethical public management system in developing and implementing sound development policies and programs, as well as the need to reform the public sector to overcome many unethical and inefficient bureaucratic dysfunctions and problems.^[26] To achieve these goals, recent reform policies have emphasized not only reforms or changes of the civil service system, public budgetary system, and state-owned enterprises (SOEs), but also restructuring and adjustment policies in market liberalization, privatization, deregulation, and decentralization. The policy measures emphasized in recent government reforms have broadened the traditional concerns of improving management and operational issues in public organizations to a new focus on good governance for successful development. The policy changes of regulatory reform, privatization, and decentralization are furthered explained as follows.

Regulatory Reform

Public policy researchers have noticed that governments often pursue their policy objectives through

regulation policies and use them to replace traditional public finance instruments.^[27] Government regulation refers to any attempt by the government to control the behavior of citizens, corporations, or subgovernments.^[28] Regulation may consist of such categories as economic regulation, social regulation, and administrative regulation.^[29,30] Economic regulation focuses on direct government intervention in corporations and market decisions, such as pricing, competition, and market entry or exit. Social regulation is related to government protection of citizen and social values, such as health, safety, environment, and social cohesion. Administrative regulation has to do with government formalities and paperwork. The instruments used in these regulation policies include, for example, laws, orders, and rules issued by all levels of government.^[30]

The pervasive use of regulation and the growth of regulatory cost have resulted in many suggestions for regulatory reforms. The purpose of regulatory reform is to improve the quality of regulations in terms of enhancing performance, reducing costs, or finding alternative policy tools. The reform activities range from revising a single regulation to scrapping and rebuilding an entire regulatory system (policies, processes, and institutions). Deregulation is part of the overall regulatory reform, which refers to complete or partial elimination of regulation in a sector to improve economic performance.^[30]

With regard to economic development, regulatory reforms are especially useful in promoting economic growth, firm competitiveness, and consumer welfare. Reform activities for industry development focus on the removal of such barriers or burdens as rigidities, disincentives, and market distortions, and the promotion of competition, entrepreneurship, technological innovation, productivity, structural adjustment, and other important market issues. Reforms that emphasize the increase of the transparency and the reduction of red tape and paperwork are also valuable for investors and ordinary citizens because of the time saved on information communication and collection.^[30]

Privatization Policy

The second major policy change in governance is the increase of privatization policies emphasized by policy makers in developed and developing countries. The word “privatization” has been used in many different ways in relation to public-sector reforms. In the broadest sense, it refers to a whole range of reformative actions designed to subject administrative activity to the disciplines of the marketplace; in other cases, it may be specialized as acts of selling government-owned

enterprises, assets, or shareholdings.^[31–33] In the United States, policy makers and public managers emphasize the use of privatization policy to deal with the problem of fiscal retrenchment (due to such changes as tax revolt, “Reaganomics,” etc.) and continued service demand. Privatization policies consist of such options as contracting out (outsourcing), franchising, local (service) shedding, public–private partnerships, grants, subsidies, vouchers, asset divestment, competitive contracting, volunteer services, and total privatization.

The arguments for privation policies are based on some beliefs, such as government should not provide services or products when these services or products are available in the private market; government is better when it is smaller; private firms are more efficient than government agencies; and private managers are better than public administrators. Specifically, privatization policy offers such advantages as improving government operations, meeting short-term project needs, lowering the costs involved in service delivery, adjusting for resource limitations, and improving service quality.^[34] The potential problem associated with privatization policy is the increasing dangers of sector blurring and the twilight zone, where individuals fall into a situation in which they will receive less than optimal treatment as citizens receiving services from government or as customers from private firms.^[35,36] It is believed that substantial benefits and significant values and protections may be lost in the transition from government to the private sector.^[34,36,37]

Finally, privatization policy has also been emphasized in many developing or economic-transforming countries.^[38–41] For these countries, privatization policy is closely related to the policy of reforming SOE. As a part of economic liberalization measures, privatizing SOE is designed to improve the operation of SOE (maximizing benefits, reducing cost, or enhancing effectiveness) and to reduce the role of government in the market economy. Privatization of SOE consists of such measures as contracting out, restraining and redeveloping labor, denationalizing and disinvesting, imposing budgetary constraints, promoting marketization, and introducing competition.

Decentralization Policy

The last major change in governance is the importance of decentralization policies in many developed and developing countries.^[42–44] In the United States, decentralization policies were emphasized as a result of the huge federal budget deficit, the cutback of federal aid, and the Reagan federalism philosophy. In many developing countries, decentralization policies were

adopted because of the emphasis on structural adjustment policies by international assistance organizations (the World Bank and the International Monetary Fund). The structural adjustment policies call for, among other things, decentralization of national government administration and reduction of the central government’s control over or intervention in the economy.^[45,46]

Decentralization has been defined as the transfer of responsibility for planning, management, and the raising and allocation of resources from the central government and its agencies to field units of government agencies; subordinate units or levels of government; semiautonomous public authorities or corporations; area-wide, regional, or functional authorities; or non-governmental private or voluntary organizations.^[47] Researchers have debated the advantages and problems of the decentralization policy.^[42–44] On the one hand, they recognize that advantages of decentralization policy include emphasizing administrative responsiveness, increasing political participation, and promoting democratic principles.^[42–44] On the other hand, they notice that decentralization policy may generate such problems as increasing disparities among regions, jeopardizing economic and social stability, and affecting administrative efficiency.^[48] They maintain that these problems may result from the lack of resources and information at the local level, and the low administrative skills, training, and educations among local public employees.

To implement decentralization policy, it is important for policy makers to consider the decentralization from a comprehensive strategic planning perspective. The strategic planning perspective emphasizes important policy components, such as understanding political, social–economic, and institutional environments of central and local governments; analyzing constraints and opportunities of these governments; considering policy scope and nature; developing an action plan of decentralization; and focusing on capacity building and empowerment. The component of capacity building is especially important, as it includes such areas as institutional, personnel, fiscal, and information capacity building.

CHALLENGES AND IMPLICATIONS

The previous review of governance concepts and policy changes provides the background information for the following discussion of major challenges in the study of governance and economic development. Three issues are especially important here: 1) the need of government and reform; 2) the importance of regional or local government; and 3) the contribution of non-governmental or non-profit organizations.

The Need of Government and Reform

Economic development researchers are continually debating the role of government in the process of economic development. The market approach of development emphasizes the role of international trade and investment and the importance of a free market in stimulating competition during the development process. The statist approach stresses the active role of government in directing public and private resources to change the structure of their economy. A third market-enhancing approach considers advantages of the market and the state approaches and promotes the role of government policy to facilitate or complement private-sector coordination and overcome other market imperfections.

The recent experience of the Asian financial crisis has not resolved the different arguments about the issue of governance. The causes of and cures for the Asian financial crisis are directly or indirectly related to the government policy and managerial activities that are important to economic development.^[49–53] On the one hand, political leaders and governmental officials have been questioned about such problems as excessive governmental intervention, lack of sound regulation, and supervision of financial sectors, as well as unethical and illegal government–business connections. On the other hand, suggestions to resolve financial crisis call for governmental action in strengthening central banks; establishing a prudent banking regulation system; and reforming administrative, political, and legal systems.

The lesson learned from these different arguments is that government continues to play an important role and government reforms are needed to assure the success of economic development. The challenges for policymakers are not only to understand different arguments of governance but also to design their own development model, which considers the balance of public and private sectors and the trade-off between market and government failure. In other words, they have to promote some specific development policies to coordinate resources in public and private sectors in order to deal with market failure problems and to compete in the global economic system. The development model must consider the reform of public management system to overcome many traditional bureaucratic problems and to promote important development values such as efficiency, equity, and accountability.

The Importance of Regional or Local Governments

The impact of decentralization policy and other changes in governance policy has resulted in the

increasing role of regional (or state) and local governments in promoting economic development programs and activities. In the United States, state and local governments have expanded their efforts in economic development policies and programs to deal with fiscal problems and new challenges.^[54–59] On the one hand, local governments have experienced continuing fiscal difficulties and challenges, such as the increasing difficulties in raising adequate revenues (because of the impact of tax revolts) and the increasing demand for public services (because of the cutback of federal social programs). On the other hand, local officials have become actively involved in economic development promotion activities, because they recognized many business development opportunities resulting from changes in the advancement of communication technology and the globalization of the world economy. The formal change refers to the development of telecommunication systems and web-based Internet services, which significantly reduce the barriers of time and distance among business communities. The latter change has to do with the end of the Cold War and the development of many postsocialist countries, which provides additional businesses and markets for local economy growth.

Researchers of regional economic development have supported the involvement of state and local governments in promoting business development. For example, Bartik used the market failure approach to explain the importance of a regional economic development policy.^[60] This approach has strengths, such as allowing a wise use of government resources, because it focuses on what private markets are unable to do, and leading to goals that are measurable in the common currency of dollar benefits. The limitations of the market failure approach consist of the lack of precise information about the magnitude of some non-market benefits, no consideration of distributional effects of regional economic development policies, and overlooking the benefits and costs of one region's policies in relation to other regions. In the absence of federal government involvement, however, Bartik believed that encouraging regional governments to pursue regional efficient policies is likely, on average, to increase the efficiency of the national economy.^[60]

The economic development strategies emphasized by state and local governments vary depending on the environment of their communities and the different goals of their economic development plan. In general, four types of development strategies have been adopted: 1) subsidizing traditional inputs such as capital (e.g., direct loans and loan guarantees, tax-exempt bond financing, and development corporations), land (e.g., land banking and site development provision), and labor (e.g., low cost/mass production, and high quality/lean production); 2) lowering political costs

of doing business, including tax abatements and incentives and limitations on the regulatory environment; 3) promoting entrepreneurial activities of market development (e.g., export promotion, research, and dissemination) and business services (e.g., policy planning, research and development, and support and consortia); and 4) developing attractive social amenities (e.g., arts and environment) and improving distressed areas (e.g., enterprise zones).^[61] All of these strategies are directly or indirectly based on or related to policy or managerial issues of governance.

The Contribution of Non-Governmental or Non-Profit Organizations

Many market-oriented policy changes, such as deregulation and privatization, have resulted in the rise and contribution of non-governmental or non-profit organizations in the delivery of public goods and services.^[62–66] In many developing countries, policy makers emphasize the development of these organizations as service providers because of their proximity to the persons served (establishing a long-term relationship with the communities); their cost effectiveness (using volunteers and relying on donations); and their flexibility, innovativeness, dedication, and responsiveness.^[62–66]

With regard to economic development, researchers have recognized the important contribution of these organizations to economic growth and social development. For example, in the United States, Rubin emphasized the importance of community-based development organizations (CBDOs) to promote economic development.^[67] As neighborhood-based non-profit organizations, CBDOs help many distressed communities to receive financial support, build affordable housing, provide job-related training activities, and create many employment opportunities. Similarly, Bhatt and Tang examined the development of group-based microfinance to promote economic development in many developing countries.^[68] Recognizing many problems associated with developmental finance in developing countries, they pointed out the contributions of group-based lending programs to help the poor generate income and employment opportunities as well as to encourage grassroots participation and empowerment in disadvantaged communities.

Despite their popularity, many non-governmental or non-profit organizations experience problems such as lacking managerial skills, facing uncertain financing and erratic regulation, encountering fragmentation and no coordination, and falling short on standards of transparency and accountability. To overcome these problems, researchers emphasized the importance of forming partnerships between public development agencies, non-governmental organizations, and private

voluntary organizations on the one hand, and investing in capacity and institution building on the other hand to improve the effectiveness of their organization management.^[67]

CONCLUSION

Government continues to plan an important role in the process of economic development. This article has reviewed major literature related to traditional governmental intervention theories and new governance concepts. It has also examined major policy changes in the areas of deregulation, privatization, and decentralization. The scope of recent policy changes emphasizes not only structural or process changes within the governance system but also the interaction between government and market. The market-oriented reform policies have significant impact on future economic development programs. The importance of local governments and the contribution of non-profit or non-governmental organizations to economic development will certainly change the traditional arguments related to the issue of governance and economic development. Future research on new changes of these organizations will contribute to the overall literature of economic development.

REFERENCES

1. Cannan, E., Ed.; *Adam Smith: An Inquiry into the Nature and Causes of the Wealth of the Nations*; University of Chicago Press: Chicago, 1976.
2. Musgrave, R.A.; Musgrave, P.B. *Public Finance in Theory and Practice*; McGraw Hill: New York, 1980.
3. Stern, N. Public policy and the economics of development. *Eur. Econ. Rev.* **1991**, *35* (2), 241–254.
4. Galenson, W., Ed.; *Foreign Trade and Investment: Development in the Newly Industrializing Asian Economies*; University of Wisconsin Press: Madison, WI, 1985.
5. Haggard, S. *Pathways from the Periphery: The Politics of Growth in the Newly Industrializing Countries*; Cornell University Press: Ithaca, NY, 1990.
6. Wade, R. East Asia's economic success: conflicting perspectives, partial insights, shaky evidence. *World Polit.* **1992**, *44*, 270–320.
7. Wade, R. *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*; Princeton University Press: Princeton, NJ, 1990.
8. Johnson, C. *MITI and the Japanese Miracle*; Stanford University Press: Stanford, NJ, 1982.
9. Ho, S. South Korea and Taiwan: development prospects and problems in the 1980s. *Asian Surv.* **1981**, *21*, 1175–1196.

10. Lin, C.Y. *Latin America vs. East Asia: A Comparative Development Perspective*; M.E. Sharpe: New York, 1989.
11. Aoki, M., Kim, H.-K., Okuno-Fujiwara, M., Eds.; *The Role of Government in East Asian Economic Development: Comparative Institutional Analysis*; Clarendon Press: New York; 1997.
12. Haggard, S.; Webb, S.B. What do we know about the political economy of economic reform? *World Bank Res. Obs.* **1993**, *8*, 143–168.
13. Rondinelli, D.A.; Montgomery, J.D. Managing economic reform: an alternative perspective on structural adjustment policies. *Policy Sci.* **1990**, *23*, 73–93.
14. Fisher, S.; Gelb, A. The process of socialist economic transformation. *J. Econ. Perspect.* **1991**, *5*, 91–105.
15. Fisher, S.; Sahay, R.; Vegh, C.A. Stabilization and growth in transition economies: the early experience. *J. Econ. Perspect.* **1996**, *10*, 45–66.
16. Liou, K.T. Strategies and lessons of China's post-Mao economic development. *Policy Stud. Rev.* **1999**, *16*, 183–208(Spring).
17. Liou, K.T. *Managing Economic Reforms in Post-Mao China*; Praeger: Westport, CT, 1998.
18. Liou, K.T. The role of government in economic development: the Chinese experience. *Int. J. Public Admin.* **1998**, *21*, 1257–1983.
19. World Bank. *Sub-Saharan Africa: From Crisis to Sustainable Growth*; World Bank, 1989.
20. World Bank Governance. *The World Bank's Experience*; World Bank, 1994.
21. Asian Development Bank. *Elements of Governance*; Asian Development Bank, (2006) (www.adb.org/Governance/gov_elements.asp) 2006-7-26.
22. Pierre, J.; Peters, B.G. *Governance, Politics, and the State*; St. Martin's Press: NY, 2000.
23. Peters, B.G.; Pierre, J. Governance without government? Rethinking public administration. *J. Public Admin. Res. Theory* **1998**, *8*, 223–243.
24. Shepherd, A. Governance, good government and poverty reduction. *Int. Rev. Admin. Sci.* **2000**, *66*, 269–284.
25. Kliksberg, B. Rebuilding the state for social development: towards smart government. *Int. Rev. Admin. Sci.* **2000**, *66*, 241–257.
26. Summers, L.H.; Thomas, V. Recent lessons of development. *World Bank Res. Obs.* **1993**, *8*, 241–254.
27. Tanzi, V. Fiscal policy for growth and stability in developing countries: Selected issues. In *Government Financial Management: Issues and Country Studies*; Premchand, A., Ed.; International Monetary Fund: Washington, DC, 1990; 15–27.
28. Meier, K.J. *Regulation: Politics, Bureaucracy, and Economics*; St. Martin's Press: New York, 1985; 1 pp.
29. Buchholz, R.A. *Business Environment and Public Policy: Implications for Management and Strategy*; Prentice-Hall: Englewood Cliffs, NJ, 1992.
30. Regulatory Reform: Overview of OECD's Work. Paris (1998).; <http://www.oecd.org/subject/regreform>.
31. De Ru, H.J.; Wettenhall, R. Progress, benefits and costs of privatization: an introduction. *Int. Rev. Admin. Sci.* **1990**, *56*, 7–14.
32. Drucker, P.F. *The Age of Discontinuity*; Harper and Row: New York, 1969.
33. Savas, E.S. *Privatization in the Public Sector*; Chatham House: Chatham, NJ, 1982.
34. Donahue, J.D. *The Privatization Decision: Public Ends, Private Means*; Basic Books: New York, 1989.
35. Bozeman, B. *All Governments are Public: Bridging Public and Private Organizational Theories*; Jossey-Bass: San Francisco, CA, 1987.
36. Moe, R.D. Exploring the limits of privatization. *Public Admin. Rev.* **1987**, *47*, 453–460.
37. Sullivan, H.J. Privatization of public services: a growing threat to constitutional rights. *Public Adm. Rev.* **1987**, *47*, 461–467.
38. Cook, P., Kirpatrick, C., Eds.; *Privatization in Less Developed Countries*; St. Martin's Press: New York, 1988.
39. Cowan, G.L. *Privatization in the Developing World*; Greenwood Press: New York, 1990.
40. Dinavo, J.V. *Privatization in Developing Countries: The Impact on Economic Development and Democracy*; Praeger Press: Westport, CT, 1995.
41. Savas, E.S., Ed.; *Alternatives for Delivering Public Services: Toward Improved Performance*; Westview Press: Boulder, CO, 1977.
42. Chemma, G.S., Rondinelli, D.A., Eds.; *Decentralization and Development: Policy Implementation in Developing Countries*; Sage Publications: Beverly Hills, CA, 1983.
43. Rondinelli, D.A.; McCullough, J.S.; Johnson, R.W. Analyzing decentralization policies in developing countries: a political-economy framework. *Dev. Change* **1989**, *21*, 513–530.
44. Samoff, J. Decentralization: the politics of interventionism. *Dev. Change* **1990**, *21*, 513–530.
45. Please, S. The World Bank: Lending for structural adjustment. In *Adjustment Crisis in the Third World*; Fineberf, R.E., Kallab, V., Eds.; Transaction Books: New Brunswick, NJ, 1984; 83–98.
46. Nellis, J.; Kikerri, S. Public enterprise reform: privatization and the world bank. *World Dev.* **1989**, *17*, 659–672.
47. Rondinelli, D.A.; Nellis, J.R. Assessing decentralization policies in developing countries: a case for cautious optimism. *Dev. Policy Rev.* **1986**, *4*, 3–23.
48. Prudhomme, R. The dangers of decentralization. *World Bank Res. Obs.* **1995**, *10* (2), 201–220.
49. Dolven, B. Hang looser. *Far East. Econ. Rev.* **1998**, *161* (14), 12–14.
50. Granitsas, A. Going for gold. *Far East. Econ. Rev.* **1998**, *161* (14), 10–12.
51. Goldstein, M. *The Asian Financial Crisis: Causes, Cures, and Systemic Implications*; Institute for International Economics: Washington, DC, 1998.
52. International Monetary Fund. The Asian crisis: causes and cures. *Finance Dev.* **1998**, *35* (2), 18–21.
53. Sender, H.; Lumpur, K.; Lee, C. Rotten to the core. *Far East. Econ. Rev.* **1998**, *161* (14), 14–16.
54. Blakely, E.J. *Planning Local Economic Development: Theories and Practice*; Sage Publications: Thousand Oaks, CA, 1994.

55. Eisinger, P.K. *The Rise of the Entrepreneurial State: State and Local Economic Development Policy in the United States*; University of Wisconsin Press: Madison, WI, 1988.
56. Levy, J.M. *Economic Development Programs for Cities, Counties and Towns*; Praeger: New York, 1990.
57. Luke, J.S.; Ventriss, C.; Reed, B.J.; Reed, C.M. *Managing Economic Development: A Guide to State and Local Leadership Strategies*; Jossey-Bass: San Francisco, CA, 1988.
58. McGowan, R.P., Ottensmeyer, E.J., Eds.; *Economic Development Strategies for State and Local Governments*; Nelson-Hall: Chicago, 1993.
59. Reese, L.A. The role of counties in local economic development. *Econ. Dev. Quart.* **1994**, *8*, 28–42.
60. Bartik, T.J. The market failure approach to regional economic development. *Econ. Dev. Quart.* **1990**, *4*, 361–370 (November).
61. Clark, C.; Montjoy, R. Globalization and the revitalization of U.S. Economic Competitiveness: Implications for economic development policy. In *Handbook of Economic Development*; Liou, K.T., Ed.; Marcel Dekker: New York, 1998; 169–170.
62. Carroll, T.F. *NGOs: The Supporting Link in Grassroots Development*; Kumarian Press: West Hartford, CT, 1992.
63. Clark, J. *Democratizing Development: The Role of Voluntary Organizations*; Kumarian Press: West Hartford, CT, 1991.
64. Fisher, J. *Nongovernments, NGOs and the Political Development of the Third World*; Kumarian Press: West Hartford, CT, 1998.
65. Henderson, K., Dwivedi, O.P., Eds.; *Bureaucracy and the Alternatives in World Perspective*; Macmillan: London, 1999.
66. Reilly, C., Ed.; *NGOs, New Paths to Democratic Development in Latin America*; Lynne Rienner: Boulder, CO, 1995.
67. Rubin, H.J. Economic partnering with the poor: why local governments should work with community-based development organizations to promote economic development? *Int. J. Public Admin.* **2000**, *29*, 1679–1709.
68. Bhatt, N.; Tang, S.Y. Group-based micro finance and economic development. In *Handbook of Economic Development*; Liou, K.T., Ed.; Marcel Dekker: NY, 1998; 115–138.

Economic Development: International

Cal Clark

Department of Political Science, Auburn University, Auburn, Alabama, U.S.A.

INTRODUCTION

Issues of international development have given rise to two important sets of policy debates over the postwar era. The first involves how to best promote national development. During the second half of the twentieth century, several distinct stages evolved, beginning with a consensus that was followed by a series of heated debates. Economic development practices in the United States constitute the second major policy domain linked to international development. Until the 1980s, economic developmental strategies within the United States had little concern or connection with international development, but this has changed radically since then, both because globalization has had a major impact on the American economy and because there is growing recognition of parallels in the central issues and dilemmas.

INTERNATIONAL ECONOMIC DEVELOPMENT STRATEGIES

For the first two decades of the postwar era, there was general consensus that economic development in the form of industrialization was highly desirable and that the economic, social, and political dimensions of development were generally reinforcing. This *modernization theory* assumed that the western experience constituted the universally valid path to development. Thus, it advocated the creation of western economic and political institutions (e.g., capitalism and democracy) in the developing world, as well as the replacement of traditional societies and cultures by modern values, which would promote economic and political change. While the modernization approach advocated westernization and capitalism, its developmental economics generally presumed that the state had to play a leading role in the industrialization of “late developers.”^[1]

In terms of economic development itself, growth is relatively limited in agricultural societies, but once a nation achieves the *take off* of industrialization, productivity and consequently GNP growth accelerate, as sketched in Fig. 1.^[2] This figure also shows that the nature of industrialization has changed dramatically during the 19th and 20th centuries in terms of what industry was the most advanced or technological

driver: first textiles, then iron and steel, then automobiles, and most recently high-tech and advanced electronics. Conventionally, it was assumed that the growth of the tertiary or service economy that succeeds industry in this model would result in decelerating productivity gains and economic growth, as indicated by line 1 in Fig. 1. However, the recent surge in the high-tech and information industries has led to the argument that these new technologies have generated another upswing in productivity and growth (denoted by line 2 in Fig. 1) that has been labeled the new economy.^[3] This also spotlights the increasing centrality of what has been termed *the creative class* for sustaining development in postindustrial societies.^[4]

By the late 1960s, the continuing impoverishment of much of the Third World indicated that development was far from as simple and automatic as the modernization school was wont to assume. Consequently, it came under sharp challenge by *dependency theory*, which argued that the capitalist economic relations advocated by the dominant paradigm actually prevented and perverted development in much of the developing world due to the market power exercised by corporations in the industrial core.^[5,6] The strident debate between these two schools of thought waned during the 1980s, though, when economic failure in the Soviet bloc and many statist economies discredited the dependency approach.

The winding down of the debate between the dependency and modernization schools produced two very different theoretical reactions that, in turn, set off a new debate about the political economy of development. On the one hand, laissez-faire economics became something of an orthodoxy in many parts of the development community—among economists, political leaders in the industrialized world, aid donors (e.g., the World Bank and International Monetary Fund), and many reformers in the Third World itself. In contrast, the most dynamic and successful region in the global economy (i.e., East Asia) pointed in exactly the opposite direction because of the strong state leadership there in promoting economic upgrading by aggressive industrial policy. Thus, the advocates of laissez-faire were soon under strong challenge from the proponents of the developmental state model. The cleavage between the laissez-faire and developmental state models, in essence, can be viewed as a

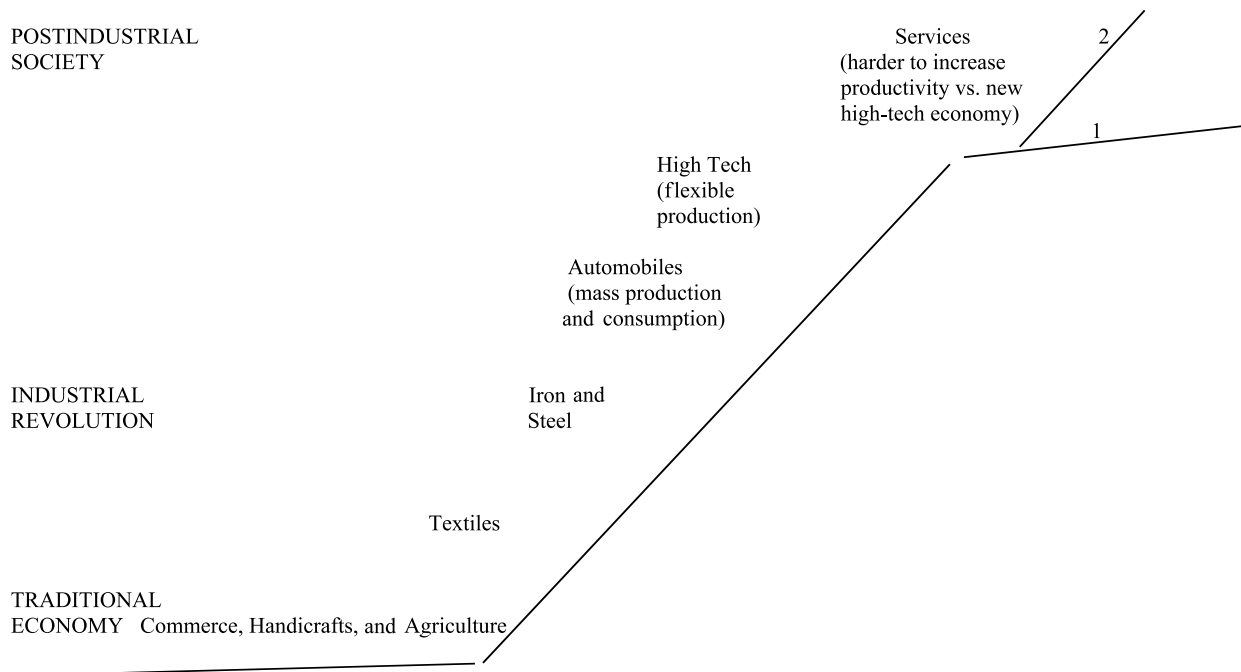


Fig. 1 Changes in leading economic activity and the rate of productivity and GNP growth. Line signifies rate of annual GNP and productivity growth. Line 1 represents conventional theory; line 2 represents the theory of the new economy. (From Ref.^[12], p. 155.)

splintering of the original consensus in modernization theory that capitalism and state economic leadership could be combined productively. Rather, one side in this debate believed that the market should rule, while the other side stressed the importance of a nation's industrial policy.^[7]

In addition, a new sharp controversy erupted in the late 1990s, this time between the advocates and critics of globalization or the rapidly growing economic interdependence (i.e., trade and capital flows, and the activities of multinational corporations) across national borders that has marked the late 20th century. Both the scholarly and the popular discussion of globalization are quite polarized. For some, globalization suggests that the new millennium (the 21st century) will actually live up to its name with greater prosperity in many nations, less hierarchic relations in the business world, the spread of democracy in an historic third wave, and many fewer threats to world peace in a global village. Conversely, others view globalization as the primary driving force behind alarmingly regressive change. In particular, they cite the growing inequality and social turmoil in the developed world, the exploitation of developing nations, and the subversion of indigenous cultures and values by coca-cola capitalism, all stemming from the decreased ability of governments to protect the public interest from corporate policy.^[8-10] Although far from being resolved, the debate over globalization highlights Schumpeter's

key insight that economic development involves creative destruction.^[11] New industries arise but only as old ones are destroyed, creating massive costs of adjustment and social turmoil. Certainly, the faster pace of creative destruction associated with globalization is creating a massive challenge for many governments.

ECONOMIC DEVELOPMENT POLICY IN THE UNITED STATES

Since the 1980s, the growing exposure of the U.S. economy to international competition, popularly termed globalization, has brought momentous changes affecting the American economy, society, and lifestyles. The social and economic changes associated with globalization have certainly been substantial and, in some ways, unsettling. In the 1950s and 1960s, the United States clearly had the most productive and competitive economy in the world based on mass production techniques of standardized goods. During the 1970s and 1980s, in contrast, America's leading economic position came under challenge for two interconnected reasons. First, the transportation and communications revolutions made it easier to locate mass production factories away from the existing industrial centers to take advantage of lower wages. Second, affluence made consumers increasingly eager to buy more specialized and higher-quality goods than the traditional mass production

techniques could fashion. The recession of the early 1990s, in particular, raised fears about America's ability to retain its industrial leadership and prosperous lifestyle. By the mid-1990s, however, many U.S. corporations had responded quite well to this challenge, reorganizing themselves to promote high-quality and high-tech production and moving into advanced and sophisticated service industries, although more recently even some of these sectors have fallen victim to significant outsourcing. Moreover, this transformation from an Industrial Age to an Information Age economy has done little, if anything, to help many of the communities and workers who had depended on the traditional industries for their well being.^[8,9,12–15]

The governmental response to the challenge of globalization differed dramatically between the national and state–local levels. The United States has long had a tradition or public philosophy of limited government, which has resulted in a government that is smaller and weaker than almost any other country in the developed world.^[16] This laissez-faire tradition, therefore, militated against a turn to industrial policy to help reinvigorate the nation's international competitiveness and manage the considerable transformation of the American economy that was under way. In fact, the federal government took quite limited action. In ironic contrast, however, almost all state and many local governments made an explicit commitment to promoting their communities' economic development,

in almost polar opposition to the laissez-faire approach of national policy. This is ironic because the federal government is seen as more liberal than most state and local ones and, thus, would be expected to pursue more interventionist economic policies.^[17]

State and local economic development activities and programs span a wide gamut, although two principal types or strategies can be discerned, as summarized in Table 1. The first represents the traditional strategy of attracting new investment through lowering the cost of doing business in a given locality either by subsidizing firms' inputs (e.g., capital, land, and labor) or by limiting the political costs of doing business (e.g., minimizing taxes and regulation). Because these incentives were generally targeted at large mass production industries, the strategy of using them came to be called (somewhat sarcastically) smoke-stack chasing. The other, newer strategy, which might be termed promoting entrepreneurship, seeks to stimulate business expansion and creation by developing new markets, by providing services that enhance business operations (e.g., human capital development, support for research and development), or by offering a community with attractive amenities.^[12,18–21]

An initial wave of state–local economic activism began in the years before World War II and continued for the first decade or so after the war. The central thrust of this wave was a drive by poorer and less industrial regions, primarily the South, to catch up

Table 1 Changing models of state–local economic development policy

	Smoke-stack chasing	Promoting entrepreneurship
Theory of growth	Growth is promoted by lowering the factor costs of production by government subsidies of capital and land and by low taxes.	Growth is promoted by discovering, expanding, developing, or creating new markets for local goods and services.
Focus of efforts	1. Stimulate relocation of large, established firms. 2. Government supports low-risk enterprises. 3. Any firm is a suitable target for aid.	1. Stimulate new and small business formation and expansion. 2. Government nurtures high-risk enterprises and activities. 3. Use strategic criteria for aid.
Desirable firms	Large existing ones	New start-ups, small firm, high tech, environmentally friendly
Desirable labor force	Low-cost labor	Skilled and flexible
Primary government contribution	Low-cost land and tax subsidies	Access to advanced technology and finance capital
Governments role	Government should follow and support private decisions about where to invest, what businesses will be profitable, and what producers will sell—basically laissez-faire	Government should help to identify investment opportunities for the private sector (i.e., new markets, products, and industries)—movement to create public–private partnerships
Location assets	Comparative advantage based on physical assets	Comparative advantage from social and environmental amenities
Market focus	Local and regional	National and international

Source: From Ref.^[12], p. 173.

with the more affluent parts of the nation. It focused on the conventional strategies of subsidizing business inputs and lowering political costs. The South's ability to lure industry from the Northeast and Midwest proved so successful by the 1950s that the old industrial heartland decided to fight fire with fire by emulating the South's locational incentives. By the early 1960s, the Northeast and Midwest had matched, if not surpassed, the South in incentive policies. A new wave in state and local economic development policy commenced during the 1970s and 1980s in response to the growing economic stress generated by America's shift in the global economy and by the declining amounts of federal aid. Traditional locational incentives assumed a major role in this offensive, but there was a turn to promoting entrepreneurship strategies, in large part because globalization was luring most of the smoke stacks out of the country altogether. This change in development strategy did not turn out to be necessarily permanent, as the recession of the early 1990s brought a renewed emphasis on competitive smoke-stack chasing.^[4,12,19–23]

CONCLUSION

The policy issues outlined in the last two sections remain unsettled as the 21st century opens. Globalization, for example, has certainly been a double-edged sword in regard to its impact on the United States, as the new technologies that are driving it both undercut the nation's position in traditional industries and create the foundation for a new economic surge in the high-tech and advanced services sectors, in line with the image of creative destruction. In terms of policy, the commitment of the United States to a relatively small government and laissez-faire policies has created a paradox for reacting to the challenges of globalization. Business is able to be flexible and innovative, allowing the United States to become the world leader in most of the emerging Information Age industries. However, the country is constrained in its attempts to improve education, build new infrastructure, and retrain workers whose skills have become obsolete, all vital tasks for retraining competitiveness at the top of the international product cycle, leaving the U.S.'s future competitiveness as something of an open-ended question.^[4]

REFERENCES

- Gerschenkron, A. *Economic Backwardness in Historical Perspective: A Book of Essays*; Harvard University Press: Cambridge, 1962.
- Rostow, W.W. *The Stages of Economic Growth: A Non-Communist Manifesto*; Cambridge University Press: Cambridge, 1960.
- Atkinson, R.D.; Court, R.H.; Ward, J.M. *Technology, Innovation, and New Economy Project*; Progressive Policy Institute: Washington, DC, 2000.
- Florida, R. *The Flight of the Creative Class*; Harper Business: New York, 2005.
- Caporaso, J.A.; Levine, D.P. *Theories of Political Economy*; Cambridge University Press: Cambridge, 1992.
- Randall, V.; Theobald, R. *Political Change and Underdevelopment: A Critical Introduction to Third World Politics*; Duke University Press: Durham, NC, 1985.
- Clark, C.; Roy, K.C. *Comparing Development Patterns in Asia*; Lynne Rienner: Boulder, CO, 1997.
- Friedman, T.L. *The Lexus and the Olive Tree*; Farrar, Strauss, Giroux: New York, 1999.
- Gilpin, R. *The Challenge of Global Capitalism: The World Economy in the 21st Century*; Princeton University Press: Princeton, 2000.
- Mittelman, J.H. *The Globalization Syndrome: Transformation and Resistance*; Princeton University Press: Princeton, 2000.
- Schumpeter, J.A. *Capitalism, Socialism, and Democracy*; Harper & Row: New York, 1950.
- Clark, C.; Montjoy, R. Globalization and the revitalization of U.S. economic competitiveness: Implications for economic development policy. In *Handbook of Economic Development*; Liou, K.T., Ed.; Marcel Dekker: New York, 1998; 151–182.
- Piore, M.J.; Sabel, C.F. *The Second Industrial Divide: Possibilities for Prosperity*; Basic Books: New York, 1984.
- Thurow, L.C. *Building Wealth: The New Rules for Individuals, Companies, and Nations in a Knowledge-Based Economy*; HarperCollins: New York, 1999.
- Womack, J.P.; Jones, D.T.; Roos, D. *The Machine that Changed the World: The Story of Lean Production*; MacMillan: New York, 1990.
- Kingdon, J.W. *America the Unusual*; St. Martin's Press: New York, 1999.
- Graham, O.L., Jr. *Losing Time: The Industrial Policy Debate*; Harvard University Press: Cambridge, 1992.
- Bingham, R.D., Mier, R., Eds.; *Theories of Local Economic Development: Perspectives from Across the Disciplines*; Sage: Newbury Park, CA, 1993.
- Brace, P. *State Government and Economic Performance*; Johns Hopkins University Press: Baltimore, 1993.
- Clarke, S.E.; Gaile, G.L. *The Work of Cities*; University of Minnesota Press: Minneapolis, 1998.
- Eisinger, P.K. *The Rise of the Entrepreneurial State: State and Local Economic Development Policy in the United States*; University of Wisconsin Press: Madison, 1998.
- Cobb, J.C. *The Selling of the South*; Louisiana State University Press: Baton Rouge, 1982.
- Kanter, R.M. *World Class: Thriving Locally in the Global Economy*; Simon & Schuster: New York, 1995.

Economic Efficiency

Brian E. Dollery

School of Economics, University of New England, Armidale, New South Wales, Australia

Joe L. Wallis

Department of Economics, University of Otago, Dunedin, New Zealand

INTRODUCTION

Economic efficiency is defined in three main ways in economic discourse. First, technical or productive efficiency refers to the use of resources in the technologically most efficient manner. Obtaining the maximum possible output(s) from a given set of inputs, or technically efficient production, was first defined with precision by Farrell^[1] and approximates what laypeople commonly conceive of as “best practice” in production. A somewhat looser term, covering the social and technological dimensions of productive activity, was subsequently developed by Leibenstein^[2] and took the name “X-inefficiency.” Leibenstein^[2] argued that although X-inefficiency (sometimes also termed organizational slack) derives primarily from a lack of motivation by productive agents, factors such as the incomplete specification of labor contracts, incomplete knowledge of production functions, and the lack of complete markets for some inputs, including market information, can also explain the existence of X-inefficiency. The second measure of economic efficiency, known as allocative efficiency, refers to the efficient distribution of productive resources among alternative uses so as to produce the optimal mix of output. In the jargon of economics, under conditions of “perfect competition,” the optimal output mix arises through consumers responding to prices that reflect the true costs of production, or “marginal costs.” Allocative efficiency thus involves an interaction between the productive capacity and consumption activity of society. Dynamic or intertemporal efficiency represents the third way of defining economic efficiency. The notion of dynamic efficiency can be traced to Joseph Schumpeter’s^[3] emphasis on innovation and his argument that the perfectly competitive conditions of allocative and productive efficiency were not necessarily the most conducive to long-term innovation and economic growth. In contrast to productive efficiency and allocative efficiency, dynamic efficiency is a much less precise concept with no universally agreed upon formal definition. In general terms, dynamic efficiency refers to the economically efficient

usage of scarce resources through time, and thus, it embraces allocative and productive efficiency in an intertemporal dimension.

BACKGROUND INFORMATION

Economics is said to be founded on the acultural and ahistorical principle of relative scarcity. In this sense, relative scarcity characteristically refers to the universal relationship between limited means and unlimited wants representative of the human condition. Available means or resources are thus scarce relative to the various ends to which they can be devoted, which necessarily implies economizing in the allocation and use of scarce resources. If means were unlimited, or human wants limited, then efficiency in the use of resources would not be necessary. But, given the pervasive existence of relative scarcity in the real world, the efficient employment of limited productive means to fulfill unlimited consumption wants becomes a critical issue. It is thus clear that the concept of efficiency plays a central role in economic analysis and economic discourse.

Interaction between “rational maximizing economic agents” occurs within the context of exchange relationships, with the market mechanism acting as the basic coordinating structure. Although economics focuses on voluntary interaction in exchange relationships, it is important to emphasize that other coordinating mechanisms coexist with exchange relationships in human society. Kenneth Boulding,^[4] in his *Economics*, distinguishes between three broad groups of social organizers. First, the “threat system,” based on interaction of the type “you do something I want, or I will do something you do not want,” underlies the existence of organized government, thus providing for the existence and enforcement of property rights and other legal entitlements and obligations. Second, the “integrative system” embraces such things as love and hate, altruism, affection, and so forth, and consequently, generates social structures like the family, community, and nation, which serve to foster and

legitimize the operations of the threat and exchange systems. Finally, and in direct contrast to the threat system, the “exchange system” is based on mutual gain epitomized by interaction of the type “you do something for me, and I will do something for you.” Exchange or market relationships are thus premised on the twin notions of voluntarism and mutual benefit.

Given the historical existence of market interaction involving the voluntary transfer of property rights between buyers and sellers, economic theory must explain why exchange relationships arise and endure. In essence, it is argued that rational maximizing economic agents exchange property rights through the market mechanism, because they benefit from such exchange. Mutual benefits occur due to the gains from trade accruing to all market participants as a consequence of the law of comparative advantage. The law of comparative advantage holds that individuals, firms, and nations can raise their incomes by specializing in that sphere of economic activity where they possess a greater degree of relative (and not absolute) efficiency, and by exchanging property rights for the output of other economic agents specializing in areas where they, in turn, possess a greater degree of relative efficiency.

While microeconomic theory is centrally concerned with the behavioral implications of the universal existence of relative scarcity, microeconomic policy focuses on measures designed to reduce the extent of scarcity and to alter the burden of scarcity among members of society. Put differently, microeconomic policy seeks to diminish the degree of relative scarcity by improving economic efficiency and to redistribute the burden of scarcity by modifying income differences between individuals. For example, the program of privatization in the United Kingdom, public sector reform in New Zealand, and airline deregulation in Australia represent cases of efficiency—enhancing microeconomic policies. By way of contrast, federal government efforts aimed at the equalization of the financial capacities of the various provinces and territories in Canadian fiscal federalism, drought relief programs in South Africa, and unemployment benefit systems in Western Europe, are instances of microeconomic policy interventions directed at redistributing the burden of scarcity among different individuals and groups in society. Thus, whereas microeconomic theory attempts to explain how markets work, microeconomic policy deals with how well markets can work.

Economic efficiency is defined in the following three main ways in economic discourse.

PRODUCTIVE EFFICIENCY

First, technical or productive efficiency refers to the use of resources in the technologically most efficient

manner. Obtaining the maximum possible output(s) from a given set of inputs, or technically efficient production, was first defined with precision by Farrell^[1] and approximates what laypeople commonly conceive of as “best practice” in production. A somewhat looser term, covering the social and technological dimensions of productive activity, was subsequently developed by Leibenstein^[2] and took the name “X-inefficiency.” Leibenstein^[2] argued that although X-inefficiency (sometimes also termed organizational slack) derives primarily from a lack of motivation by productive agents, factors such as the incomplete specification of labor contracts, incomplete knowledge of production functions, and the lack of complete markets for some inputs, including market information, can also explain the existence of X-inefficiency.

Regardless of whether we accept productive efficiency or the broader concept of X-inefficiency, this notion of economic efficiency is insufficient, because the efficient production of goods does not consider the consumption desires of society. It is pointless to produce goods efficiently if people would rather consume some other combination of goods. Accordingly, additional measures of economic efficiency are necessary.

ALLOCATIVE EFFICIENCY

The second measure of economic efficiency, known as allocative efficiency, refers to the efficient distribution of productive resources among alternative uses so as to produce the optimal mix of output. In the jargon of economics, under conditions of “perfect competition,” the optimal output mix arises through consumers responding to prices that reflect the true costs of production, or “marginal costs.” Allocative efficiency thus involves an interaction between the productive capacity and consumption activity of society.

Historically, economic theory has been chiefly concerned with allocative efficiency, and only in recent times, has attention been focused on productive or X-inefficiency. Perhaps the main reason for this is the central position traditionally occupied by the perfectly competitive model in neoclassical theory. Under perfect competition, market forces ensure an absence of X-inefficiency, because, with decreasing returns, firms must produce at the minimum point of long-run average cost. Any producer that exhibits X-inefficiency will thus not pass the “survival of the fittest” test.

A major accomplishment of economic theory resides in the establishment and refinement of the properties of allocative efficiency. Two approaches to the problem of allocative efficiency have been developed.

Pigouvian Approach

First, the partial equilibrium or Marshallian approach represents the primary method of studying particular markets in isolation. This technique examines the equilibrium conditions in a single market on the assumption that the prices of all other commodities and factors of production are given. It is evident that a partial equilibrium approach does not allow for feedback effects between markets. In practice, the partial equilibrium or Marshallian approach is especially suitable for markets with output that is not a significant item in total expenditure or highly substitutable for any other single commodity.

In terms of the Marshallian approach, the conditions for allocative efficiency were first specified by Pigou.^[5] In essence, Pigou argued that allocative efficiency occurred when the benefit to society of consuming some good or service exactly equaled its cost to society. In technical language, allocative efficiency for private goods occurs where marginal social benefit equals marginal social cost; that is, no divergence between private and social benefits and costs exists. In intuitive terms, this means that the resultant price and quantity accurately reflect the degree of relative scarcity for this good or service.

Paretian Approach

The second major approach to the problem of allocative deficiency is the general equilibrium or Walrasian approach, which examines equilibrium conditions in all markets simultaneously. Because all markets are analyzed at the same time, interrelationships between markets and feedback effects occupy a central position in the general equilibrium approach.

Whether a partial or general equilibrium approach should be employed depends largely on the problem at hand. For instance, if the primary concern is a proposed policy change that has a direct impact on many sectors of the economy simultaneously, then a general equilibrium approach is clearly appropriate. An example of this kind of problem would be the introduction of, or modifications to, general sales tax. On the other hand, if the main concern was a proposed policy change that has a direct effect on only one sector, then a partial equilibrium analysis would probably suffice. Excise taxes on items such as cigarettes and liquor constitute examples of this type of problem.

In terms of the general equilibrium or Walrasian approach, the conditions for allocative efficiency are somewhat more complex. In fact, allocative efficiency in the general equilibrium context requires the simultaneous concurrence of three conditions. First, economic efficiency in production must occur such that

no intersectoral reallocation of resources can increase the output of any economic good without decreasing the output of some other economic good. In economic jargon, this means that the marginal rates of technical efficiency of input factors must be equal in all production sectors at current factor market prices. Second, economic efficiency in consumption must occur such that no interpersonal reallocation of commodities can increase the well-being (or utility) of some consumer without decreasing the well-being of some other consumer. In technical terms, the marginal rates of substitution for all goods and services must be equal for all consumers at prevailing market prices. And finally, overall economic efficiency in production and consumption requires an optimal conformity between economic efficiency in production and economic efficiency in consumption, such that a change in the composition of output cannot increase the utility (or satisfaction derived from consumption) of some consumer without decreasing the utility of some other consumer. In technical language, a given set of market-determined prices will equal the marginal rate of transformation in production and the common marginal rates of substitution in consumption. If these three conditions are met, then society is said to have achieved high-level optimality in the sense that allocative efficiency occurs in all spheres of economic activity.

Walrasian general equilibrium is said to be Pareto efficient in the sense that it is impossible to improve anyone's welfare by altering production or consumption without impairing someone else's welfare. The concept of Pareto efficiency derives from the work of Vilfredo Pareto,^[6] in his 1906 *Manual of Political Economy*, and forms the basis of much of contemporary welfare economics. A central proposition of neo-classical welfare economics holds that if all prices are market determined, then a Pareto-efficient general equilibrium will ensue. In other words, the operation of competitive markets in a capitalist economy automatically ensures that allocative efficiency will occur in all markets. No corrective intervention by governments can improve upon this outcome, because it is already Pareto efficient.

DYNAMIC OR INTERTEMPORAL EFFICIENCY

Dynamic or intertemporal efficiency represents the third way of defining economic efficiency. The notion of dynamic efficiency can be traced to Joseph Schumpeter's^[3] emphasis on innovation and his argument that the perfectly competitive conditions of allocative and productive efficiency were not necessarily the most conducive to long-term innovation and economic growth. In contrast to productive efficiency and allocative efficiency, dynamic efficiency is a much less

precise concept with no universally agreed upon formal definition. In general terms, dynamic efficiency refers to the economically efficient usage of scarce resources through time, and thus, it embraces allocative and productive efficiency in an intertemporal dimension. Sometimes, the concept of dynamic efficiency is given more specific meaning in the literature. For instance, the macroeconomic debate on the optimal rate of saving, or the decision to postpone some part of current consumption to a future date, has been referred to as intertemporal efficiency. Similarly, the comparative institutions approach, often associated with New Institutional Economics, focuses on the efficiency of alternative institutional arrangements at exhausting the welfare gains attendant upon exchange relationships.^[7]

REFERENCES

1. Farrell, M.J. The measurement of productive efficiency. *J.R. Stat. Soc.* **1957**, *120* (Part III), 253–282.
2. Leibenstein, H. Allocative v. X-efficiency. *Am. Econ. Rev.* **1966**, *56* (2), 394–407.
3. Schumpeter, J.A. *Capitalism, Socialism and Democracy*; Allen and Unwin: London, 1943.
4. Boulding, K.E. *Ecodynamics*; Sage: New York, 1978.
5. Pigou, A.C. *The Economics of Welfare*; Macmillan: London, 1920.
6. Pareto, V. *Manual of Political Economy*; Girard and Briere: Paris, 1906.
7. Vira, B. The political coase theorem: identifying differences between neoclassical and critical institutions. *J. Econ. Issues* **1977**, *31* (3), 761–789.

Economic Theories of Public Leadership

Brian E. Dollery

*School of Economics, University of New England, Armidale,
New South Wales, Australia*

Joe L. Wallis

Department of Economics, University of Otago, Dunedin, New Zealand

INTRODUCTION

The concentration of the operational authority over public organizations in the hands of chief executives (CEs) is a striking feature of the new public management (NPM) that has emerged from about two decades of public sector reform in English-speaking nations. This article is concerned with the extent to which the new-styled CEs should be expected to exercise leadership and management.

NEW PUBLIC MANAGEMENT AND ORGANIZATIONAL LEADERSHIP

Although there are variations in the way NPM has been applied in different countries, it generally moves away from a traditional, highly prescriptive, administrative framework by emphasizing the importance of “hands-on” professional management and the “freedom to manage,” the introduction of performance appraisal with explicit performance standards, and a greater use of output controls with their stress on results rather than procedures.

The restructuring required to implement these doctrines has often been radical as large bureaucratic structures have been broken up into “single-objective, trackable and manageable units”^[1] within which it is easier to match resources to defined tasks and to shift from controlling input to monitoring output. This reflects a fundamental realignment of the core administrative values that are embodied in organizational design. According to Hood,^[1] the quest to implement the doctrines of NPM may advance what he calls the “sigma-type” cluster of administrative values that highlight “economy” and “parsimony,” even where this makes it more difficult to realize “theta-type” values of “honesty” and “fairness,” and “lambda-type” values of “security” and “resilience.”

Numerous researchers have pointed out that NPM has been influenced as much by economic theories as by management theories. For example, Hood^[1] argued

that the core doctrines of NPM are derived from a “marriage of two different streams of ideas”—economics, especially new institutional economics and public choice theory, with their emphasis on contractual solutions to agency problems in the public sector, and “managerialism,” with its emphasis on a “professional management” that is “portable,” “paramount,” “discretionary,” “central,” and “indispensable.”

The compatibility of these two streams stems from the way economic contractualism seeks to depoliticize the CE’s role so that they can be “freed to manage” through the application to public organizations of “best practices” developed in the private sector. If economic theory sees the public sector as a vertical chain of principal–agent relationships, then the goal of NPM-style reforms is to delineate accountability for “outcomes,” “outputs,” and “inputs.” This process has been carried furthest in New Zealand where the State Sector Act of 1988 made Cabinet ministers accountable for “outcomes” so that they would have to contract with specific agents, including the CEs of government departments, to supply the “outputs” they deemed necessary to achieve these “outcomes.” The CEs could then be made accountable for the delivery of these outputs, but given the discretion to manage their inputs and delivery mechanisms in the manner they deemed to be most efficient. Through successive funding cuts it was hoped that these CEs would be placed under relentless pressure to save costs by restructuring their organizations.

The staff redundancies in the organizations being restructured have often been substantial. However, the costs incurred by redundant staff, in retraining for and seeking new jobs, and by the state, in providing them with income support while they are unemployed, are typically not internalized by the CEs of the restructured organizations. These officials are more likely to take into account the transaction costs involved in establishing the contractual and accounting systems designed to make them accountable for the efficiency with which they deliver specified “outputs.” To rationalize the restructuring required to implement NPM, they can draw on both economic theories and

managerialism to argue that a shift to NPM enables their organizations to potentially realize “productivity gains” with a net present value that exceeds these transactions costs.

Managerialism does, however, differ from economic theories in terms of the way it relates these productivity gains to the quality of leadership exercised by the CEs responsible for transforming public organizations according to NPM principles. Although economists have traditionally neglected the study of leadership, managerialists have been able to draw from extensive studies in management and organizational behavior that have identified leadership as having a significant impact on organizational performance, particularly during periods of discontinuous organizational change.^[2]

In surveying this literature, Bryman^[3] pointed out that there are as many definitions as there are theories of organizational leadership. This is because leadership is a multifaceted phenomenon so that the definition can often do no more than highlight the aspects of the phenomenon that the subsequent theorizing sets out to elaborate. We recognize this by advancing a definition of leadership that brings together the political and social aspects that are the subject of a range of other leadership theories. We then show how economic theory can be modified to explain the social influence aspect. We then argue, by way of conclusion, that there is a need to look beyond economics to explain the irreducibly political dimension of leadership in order to assess the contractualist view that the role of the CE should be depoliticized.

THEORIZING FROM DEFINITIONS OF LEADERSHIP

Many writers on organizational leadership seek to define this phenomenon in a way that distinguishes it from “management.” Two aspects of leadership are often highlighted. The first is reflected in the oft-quoted slogan that “management is about doing things right while leadership is about doing the right thing.” This focuses on the judgment-making aspect of leadership. To exercise leadership in these terms, a CE must make judgments that affect the direction of an organization’s development. Interpretive studies of leadership^[4] stress the political dimension of this judgment-making process. They depict leaders as making these judgments through their involvement in the multidimensional processes of social interaction through which acceptable meanings of issues, events, and actions are created and sustained.

Second, many writers conceive leadership as a distinctive type of social influence relationship. To lead

is to influence, guide, engage a following, and build their commitment to realize a particular vision. According to Tichy and Devanna,^[5] leadership involves pulling an organization into the future “by creating a positive view of what the organization can become and simultaneously providing emotional support for individuals during the transition process.”

For CEs to be said to be exercising leadership in this way, they must be striving to influence the intrinsic and not just the extrinsic motivation of actual and potential followers. As CEs, they will be in a position to exercise “reward power” by using extrinsic rewards to induce subordinates to perform the tasks they set them, or “coercive power” by administering a set of extrinsic penalties for non-compliance with their directions. They cannot be said to be exercising leadership *per se* if they choose to exercise only these forms of power. Leadership more essentially involves influencing the intrinsic motivation of followers, through processes of “internalization” when they amplify values and beliefs that are shared by both leaders and followers, and “identification” when they engage in behaviors that reinforce and strengthen the sense of personal identification and loyalty that followers place in them.

Comparative studies have often found these processes to be more advanced in private than in traditional public organizations. For example, a Canadian study by Zussman and Jabes^[6] found that the percentage of middle managers who reported that their chief executive officers provided leadership in these ways was significantly lower in public organizations than private organizations, and that the level of public service leadership falls off much more rapidly as one goes down the organization than is the case in the private sector. The implication of these studies is that the implementation of NPM may give the CEs of public organizations the opportunity to close this “leadership gap” by adopting private sector best practices.

In this article, we attempt to elaborate a definition of leadership that brings both its political and social influence aspects together. Accordingly, we propose that a CE can be said to be exercising leadership when this official makes judgments about the direction in which an organization should be moving in a way that engages the hopes of its members to a degree that induces them to strive together to move the organization in this direction. We also propose that economic theory can only make a contribution to understanding this concept of leadership if it is initially assumed that the judgment about organizational direction has already been made. This assumption can later be relaxed when we consider the implications the judgment-making process has for the politics-administration dichotomy.

THE ENACTMENT PROCESS

Once leaders have made a judgment about the direction in which they want to steer the organization, they will need to engage its members by calling them together, at all levels of the organization, to consider how they are to enact their leaders' judgment and advance their quest. This enactment process can therefore set in motion interactions that can take the form of "expression games." According to Goffman,^[7] these involve "senders" who express themselves in particular ways and "receivers" who take in and react to such expressions, forming an impression of the "senders." The expression games generated by the enactment process can give participants the opportunity to express the hope they have placed in the leader's quest.

The CEs of large public organizations will typically only be able to engage directly in such expression games in interactions that bring together actors at higher levels of the organization. These interactions can, however, serve two functions. First, they can give the CEs the opportunity to model the behavior they expect of "followers." Second, they can give other participants the opportunity to identify themselves as such followers. A "follower" in this context is someone whose hopes are aligned with those of the leader. To a varying degree, such followers can look to the leader to strengthen their hopes. The leaders can also look to those followers who leave an impression of their trustworthiness to reproduce these expression games in interactions at lower levels of the organization.

Through the reproduction of these expression games, the enactment process may be accompanied by a diffusion of leadership and a development of a following at all these levels. To understand why this may have a significant impact on the organization's culture and performance, we need to examine the motivational effect of the hopes that are evoked and influenced through this process.

AN ECONOMIC THEORY OF HOPE AND LEADERSHIP

The traditional reluctance of economists to study the social influence aspect of leadership may have been based on the perception that, in seeking to influence followers, leaders are trying to change their preferences. The study of leadership would therefore seem to be out of bounds to the majority of economists who subscribe to the convention that economic analysis should either 1) take the preferences of individuals as given and not look inside the "black box" within which they are formed and transformed; or 2) assume that they are stable and explain apparent preference

change in terms of adjustments in the shadow prices of inputs in household production functions.^[8]

However, attempts have been made by Hermalin^[9] and Casson^[10] to analyze how leaders can influence followers by inducing them to internalize the externalities of their discretionary effort decisions. Hermalin^[9] argued that situations of asymmetric information may ironically give leaders the opportunity to alleviate the weak incentives followers have to supply effort to a team when they cannot be compensated according to their individual contributions. In these situations, leadership by example may be interpreted as a signal that leaders have better information about the value of effort devoted to their common activity so that "the harder the leader works, the harder the followers work."

This model is, to some extent, complemented by an earlier one^[10] in which the effect on individual utility functions of the emotions of shame and guilt followers experience when they fail to internalize effort externalities, is susceptible to "moral manipulation" through rhetoric the leader addresses to the group as a whole. Casson went on to show how leaders can compare the net benefits of raising the intensity of moral manipulation with those associated with more intense monitoring in a group in which individuals have varying "moral sensitivities."

In our view, neither model lays an adequate foundation to explain how CEs can influence the hopes that followers place in their quest to move the organization in a particular direction. The general approach Elster^[11] recommended in his article, "The Emotions and Economic Theory," may be more helpful in this regard. Elster basically argued that economists should reject this simple cost-benefit approach to modeling the effect of the emotions on behavior in favor of one, derived from cognitive dissonance theory, that views an emotion as a particular type of action tendency engendered by antecedent beliefs and the investment of emotional energy.

This perspective may explain why the members of an organization who have focused their hopes on advancing a quest to realize a particular vision may have an action tendency to interact in a way that reinforces their beliefs and enhances their emotional energy. These hopes will be based on beliefs that the advancement of the quest is "neither impossible nor inevitable" and that it is "worthwhile" or "important" in the sense that it is worthy of pursuit in a special way incommensurable with other goals.

If one cannot hope without holding a belief in the worth and possibility of advancing the quest that is the object of hope, it would seem that to hope is to locate oneself in a "space of questions" about both these beliefs and the authenticity of one's claims to hold them. These questions may give rise to a

dissonance that can only be alleviated by engaging in interactions with actors who can both supply clearer, more articulate reasons for holding the beliefs that trigger hope and opportunities to authenticate these beliefs by striving to advance the quest in which this hope is placed. Leader–follower relationships will thus emerge in groups where one person, the leader, plays a focal role in supplying such “reasons” and “opportunities.”

The extent to which the process of enacting the judgments of a CE engages the hopes of an organization’s members may thus be gauged in terms of the way it draws leader–follower groups into a network that penetrates every level of the organization. To understand how the expression games pursued in such groups can enhance the capacity of their members to supply effort toward advancing the CE’s quest, it may be helpful to refer to Collins’^[12] theory of “Emotional Energy as the Common Denominator in Rational Social Action.”

Collins focuses on the reinforcement and strengthening of emotional energy that occurs when “interaction rituals” (IRs) pass thresholds of “boundedness” and “density.” Leader–follower expression games may pass the threshold of boundedness when they give participants the opportunity to express an observable “passion” to advance shared hopes. Where such interactions also pass the “threshold of density” such that the participants are close enough for a sufficient period of time to ensure that they can be moved by one another’s passion, the participating group’s focus of attention and common emotional mood is likely to go through a short-term cycle of increase and mutual stimulation until a point of emotional satiation is reached. The interaction will leave each participant with an “energetic afterglow” that “gradually decreases over time” so that individuals have an incentive to reinvest their emotional energy in subsequent interactions. It may therefore accumulate across IRs so that “an individual may build up a long-term fund of confidence and enthusiasm by repeated participation in successful IRs.”^[12]

The leader–follower expression GAMES that are set in motion during an enactment process may thus significantly boost the emotional energy reserves of followers throughout the organization. This concept of leadership differs from that advanced by Hermalin and Casson in that it emphasizes the influence of leadership on the capacity rather than the willingness of an organization’s members to supply effort in excess of that for which they can be monitored and compensated. As economic theories, all three concepts tend to ignore the complexities involved in the irreducibly political process of judgment making. The implications of this neglect in assessing the leadership role of CEs must be considered by way of conclusion to this article.

CONCLUSION: THE POLITICAL DIMENSION OF PUBLIC SECTOR LEADERSHIP

The process that is resolved when a CE makes a judgment about organizational development would seem to involve the type of “autonomous politics” that, according to Barber,^[13] characterizes occasions when “some *action of public* consequence becomes *necessary* and when men must thus make a *public choice* that is *reasonable* in the face of *conflict* despite the *absence of an independent ground*” (original emphasis). Moreover, to the extent that this judgment sets in motion an enactment process than transforms the culture of an organization by focusing the hopes of its members on advancing a particular quest, it would seem that there is an irreducibly political dimension to the leadership that steers this process.

This is because there are likely to be competing perspectives on the appropriate culture for a public organization and no independent grounds for judging between them. It would therefore not be reasonable for a CE to make a judgment in this regard without striving to appreciate the divergent perspectives different stakeholders bring to this issue and giving them an adequate opportunity to make case for the values, interests, and perspectives they are seeking to advance.

Significantly, the tenure of a CE may become unsustainable, particularly where they are not entitled to a renewal of fixed-term contracts, when these judgments erode their relations of trust with portfolio ministers. These relationships are likely to be based on mutual impressions of reasonableness. They are clearly more personal and less “arms length” than typical contractual relationships. The view that contractualist reforms of NPM can revive the traditional politic–administration dichotomy therefore seems to break down in the analysis of the leadership role of the CEs of public organizations. This role cannot be exercised outside a political relationship of mutual trust between these CEs and their principals. The political dimension of organizational leadership in the public sector must thus ensure it has to be collectively supplied.

REFERENCES

1. Hood, C.A. Public administration for all seasons? *Public Adm.* **1991**, 69 (1), 3–19.
2. Bass, B. *Bass and Stogdill’s Handbook of Leadership*; Free Press: New York, 1990.
3. Bryman, A. *Leadership and Organisation*; Routledge and Kegan Paul: London, 1986.
4. Kay, R. Leadership and voluntary organizations. In *Voluntary Agencies: Challenges of Organization and Management*; Billis, D., Harris, M., Eds.; Macmillan: London, 1996.
5. Tichy, N.; Devanna, M. *The Transformational Leader*; John Wiley and Sons: London, 1986.

6. Zussman, D.; Jabes, J. *Perceptions of Leadership: A Comparison Between Government and Business Organization*, Paper Presented at RIPA Conference on Government and the Private Sector, Canterbury, 1988.
7. Goffman, E. *The Presentation of Self in Everyday Life*; Anchor Books: New York, 1959.
8. Stigler, G.; Becker, G. De gustibus non est disputandum. *Am. Econ. Rev.* **1977**, *67* (2), 76–90.
9. Hermalin, B. An economic theory of leadership. *Am. Econ. Rev.* **1998**, *88* (5), 1188–1206.
10. Casson, M. *Economics of Business Culture: Game Theory, Transactions Costs and Economic Performance*; Clarendon Press: London, 1991.
11. Elster, J. Emotions and economic theory. *J. Econ. Lit.* **1998**, *36* (1), 47–74.
12. Collins, R. Emotional energy as the common denominator of rational social action. *Ration. Soc.* **1999**, *5* (2), 203–220.
13. Barber, B. *Strong Democracy: Participatory Politics for a New Age*; University of California Press: Los Angeles, 1984.

Economics, Transaction Cost

Glenn A. Daley

RAND Graduate School, Santa Monica, California, U.S.A.

INTRODUCTION

Transaction cost economics (TCE) is the study of how markets and organizations are affected by the costs of creating and maintaining economic relationships, apart from the production costs of goods and services. The initial concept of TCE was formulated by Ronald H. Coase in 1937,^[1] but only developed into a distinct field of study since the early 1970s, in the works of Oliver E. Williamson^[2] and others. It has become a cornerstone of the new institutional economics. Starting with the study of transaction costs as economic phenomena, TCE has developed into a broad discipline for the study of governance in organizational and social terms, not just traditionally economic terms.^[2]

THEORY OF TCE

It is obvious that transaction costs affect economic choices. However, there is disagreement among economists about the generalizations of TCE regarding organizational and institutional effects. There are also difficulties in defining transaction costs precisely and measuring their aggregate level in a firm or in the national economy. Thus, TCE is a developing field of study; its proponents continue working to refine the theory, test it empirically, and apply it to a widening variety of economic and organizational phenomena.

In a nutshell, the theory of TCE is that firms and other organizations, as well as social institutions and modes of governance, represent various adaptations to the existence of transaction costs. Further, TCE claims that the characteristics of transactions deeply affect the types of governance used in organizations and in society. This provides insight into economic behavior that is not easily explained within the models of classical economics, showing that seemingly inefficient practices might actually perform efficiency-promoting functions in a transaction cost framework. This approach has been applied to phenomena as diverse as contracting for uncut diamonds and theater bookings,^[3] the behavior of illicit drug dealers,^[4] and even vote trading in the U.S. Congress.^[5]

In his essay, Coase asked why firms exist at all, given the classical concept that economic efficiency is best achieved through the market pricing mechanism.^[1] A firm is characterized by the fact that it suspends the market pricing mechanism for its own internal relationships, while depending on the market in its external relationships. But what advantage is gained by suspending the market pricing mechanism internally? Why are these particular relationships internalized, whereas others are kept external? The classical notion of economies of scale does not adequately answer these questions. It is conceivable that even a large-scale production technology could be adopted by a “virtual” firm consisting solely of independent contractors, using rented capital equipment, and engaging in competitive market transactions with each other on a daily basis. We almost never see this happen, though, and the question is why not?

Coase’s answer was that there are costs to using the market pricing mechanism. A firm can reduce these costs through internalization and integration, producing a certain good or service in-house rather than buying it in the market. This raises another question: If internalization reduces costs, why isn’t everything produced by one giant firm? The answer is that integration brings its own penalties, such as increased costs of coordination and supervision, various other kinds of overhead, and the loss of flexibility due to maintaining a permanent workforce instead of hiring contractors by the day. Thus, what matters is finding the level of internalization that minimizes the net of transaction, integration, and production costs. As in much of economics, the emphasis is on comparing alternative choices at the margin, the “differential” transaction costs of a particular organizational structure or mode of governance, as compared with feasible alternatives.

The field of TCE currently faces a broad agenda of empirical research, both to measure different types of transaction costs and to test specific predictions of the theory. Because it is impossible to observe the costs of an alternative not chosen, it has been difficult to quantify differential transaction costs; much of the published literature consists of case studies rather than quantitative studies. But there is an increasing body of quantitative studies with fruitful results.

One such study by Masten, Meehan, and Snyder^[6] showed that organization costs represented about 14% of total costs for the acquisition of components by a naval shipbuilder, that internalization of the work currently performed by subcontractors would increase internal organization costs by about 70%, and subcontracting work currently performed internally would result in market organization costs almost triple the current costs. This suggests that the firm has found an optimal level of internalization. This study also confirmed that variations in the cost and governance characteristics of transactions were associated with the level of internalization of those transactions. The researchers found that integration decisions were motivated more by their effects on internal organization costs than by their effects on market mechanism costs. This empirical result supports a broader view of TCE as including not only direct market costs, but also all costs associated with alternative modes of organization.

WHAT ARE TRANSACTION COSTS, AND WHY DO THEY OCCUR?

The classical microeconomic model focuses on production costs—the capital, labor, and raw materials that go into creating goods and providing services. In contrast, transaction costs include the costs of locating and evaluating goods, buyers, and sellers; the costs of negotiating and making contracts; and the costs of monitoring and enforcing the terms of contracts over time. The idea of a transaction cost is easily seen in examples such as brokerage and banking fees, sales commissions, and the expense of hiring attorneys to review contracts. However, the concept also includes costs incurred in the process of creating and maintaining alternative arrangements for the governance of transactions.

Whereas classical economics treats management costs as a high-level labor input to production, TCE treats management and other integration costs as the costs of internal transactions. Integration costs (i.e., internal transaction costs) are distinct from the more obvious category of market-related transaction costs, and attempts have been made to measure them independently. But both categories must be considered. Internal transaction costs, including management, are substitutes for market costs; in other words, they are the costs of choosing integration rather than the market pricing mechanism as a mode for organizing economic activities. Thus, transaction costs are any costs that arise from the need to create and govern economic relationships between parties, whether they result directly from transactions or indirectly from the type of structure adopted to govern transactions.

Bounded Rationality and Opportunism

In the classical model, a transaction is a simple event. Both buyer and seller know everything they need to know. The choice of trading partners is based on price alone, but because competition forces the price to be the same for everyone, all sellers are interchangeable, and all buyers are interchangeable. Every good is purchased in separate transactions, with no bundling or tying together of goods. There are no linkages between transactions across time; tomorrow's purchase depends only on tomorrow's price and not on the terms of today's transaction. Both parties can specify exactly what they want from the exchange, and observe instantly whether it is satisfied. The transaction is concluded as money and goods change hands, so there is no need for enforcement after the fact. Due to all these features, the transaction is frictionless and costless.

These assumptions are clearly unrealistic for numerous reasons, but transaction cost economists emphasize two key problems. First is the fact that people are limited in their ability to know and control every detail necessary to achieve their intended outcomes. Where the classical model assumes human rationality, the new approach assumes only *bounded rationality*, a term introduced by Herbert Simon^[7] to indicate human fallibility in the context of rational intentions. Many transactions result in unsatisfactory outcomes because the parties cannot predict the future, know all the relevant facts, and control every detail; knowing this, buyers expect vendors to provide customer service departments, warranties, and return policies.

The second problem is that people tend to take advantage of situations for their own benefit, especially any ambiguity in the terms of an agreement or any lack of observation by the other parties to an agreement. This tendency is called *opportunism*. It includes fraud, deception, and misconduct, but it also includes the many subtle ways that people find to manipulate transactions for their own self-interest. An example is an employment relationship where employees are tempted to shirk responsibilities whenever a supervisor is not looking and an employer is tempted to pressure employees to work additional hours without claiming overtime.

These problems have been addressed to some extent by accommodating new assumptions into the classical model, leading to what is called the neoclassical model. For instance, instead of assuming that the buyer knows what is being purchased and receives it without risk, we say that the buyer has expectations about the utility to be received and adjusts demand in light of the risk. The new model also accommodates asymmetric information, where one of the parties cannot observe all relevant information about the good.

Examples include quality in the case of a complex good such as a used car, performance in the case of a service such as auto repair, and responsible behavior in the case of insured risks such as driving safety. Problems such as these are labeled *adverse selection*—the effect of unobserved information on market choices prior to the transaction—or *moral hazard*—the effect of unobservability on behavior after the transaction.

The neoclassical approach emphasizes structuring the terms of a transaction in advance so as to obtain the best expected return considering these risks. In employment and other performance contracts, the approach of structuring incentives to motivate the optimal response from other parties is called the *principal-agent* model. For example, an employment contract may include both a fixed wage and a variable bonus contingent upon the output of a work team; the bonus places part of the risk of non-performance on the employee, and thus creates an incentive not to shirk.

According to TCE, these elements of the neoclassical approach still do not solve many aspects of the problems of bounded rationality and opportunism, and inadequately describe the ways in which people adapt to these problems in the real world. Asymmetric information models address some but not all the difficulties of *observing* the performance of a contract. However, they do not address the difficulty of *specifying* the performance of a contract. Many economic relationships, such as construction contracts, depend on highly complex and uncertain plans for the future. In other cases, such as personal services, the expected outcome may be a highly subjective experience. In such cases, it may not be possible—or affordable—for a buyer to specify in perfect detail what would constitute the expected outcome of the transaction.

Incomplete Contracts

A contract that specifies every possible detail of its fulfillment and how that fulfillment will be recognized, so that there is no room for opportunism or unanticipated error, is called a *complete* contract. All other contracts are *incomplete*. It is obvious that we seldom do business using complete contracts. The purchase of a simple object that can be inspected at the time of the sale, perhaps an apple or a greeting card, is based on an unwritten complete contract. But where there are hidden qualities, even in something as simple as a bottle of soda or an article of clothing, we rely not on the completeness of contracts but on other aspects of the relationship. We consider such factors as our familiarity with the vendor, our recognition of a commercial brand, and our trust in the truth of labels. Even a lengthy contract for a major industrial project cannot

specify every eventuality; rather, it specifies a *process* for resolving unspecified issues during the life of the contract. In TCE, such dispute resolution processes, truth-in-labeling regulations, commercial brands, and ongoing vendor relationships are recognized as substitutes for complete contracts.

Contracts become even more complex and costly when they are connected to other transactions. This is clear when different goods must interact, such as computer hardware and software; critical elements of one contract may depend on the fulfillment of other contracts with different parties, multiplying the consequences of bounded rationality and opportunism. There is also interconnectedness in multiple transactions with the same partner. When a transaction is repeated many times, uncertainty is reduced and the contract becomes a matter of mutual understanding, making it costly to switch vendors even to obtain a lower price. The terms of contracts may be modified on the basis of expectations about later transactions, while the terms of subsequent contracts may be altered due to previous experiences, so that a transaction becomes part of a pattern of favors granted or withheld over time. Thus, the transacted price of a good or service can deviate considerably from the market price.

Asset Specificity

Another characteristic of transactions, closely related to the above characteristics of complexity and repetition, is called *asset specificity*. It describes the degree to which a good, service, vendor, or employee cannot be interchanged with other goods, services, vendors, or employees. At one level of specificity, semiconductor fabricating equipment cannot be interchanged with metalworking machine tools; at an even higher level, the dies used in manufacturing a firm's proprietary product cannot be exchanged with dies from other firms. With regard to employment contracts, an accountant cannot be interchanged with a mechanic. An accountant with many years of experience in a firm has knowledge that cannot be replaced even by hiring another accountant from outside. Furthermore, the accountant with such inside knowledge is more valuable to this particular employer than he or she would be to a different employer.

Clearly, the higher the degree of asset specificity, the greater the mutual dependence between the parties to a transaction. In such a dependent relationship, one or both of the parties would lose value by switching partners, and thus has an incentive to maintain the relationship. Because people are forward looking and rational, but aware of bounded rationality and opportunism, they will form contracts to balance the potential

cost of losing such relationships with the ongoing costs of maintaining the relationships. Both of these costs are high where asset specificity is high; both are low where asset specificity is low. Thus, the employment contract for a high-level systems engineer with knowledge of proprietary technologies involves higher transaction costs, and more complex organizational arrangements, than the contract for a janitor.

GOVERNANCE AND ORGANIZATIONS

As a result of bounded rationality and opportunism, we generally recognize that we cannot depend on commitments others make or expect others to depend on commitments we make. There are social settings where a person's word is taken as a dependable bond, but closer study reveals that members of such communities depend on cultural norms or institutions and their ongoing relationships with each other to guarantee these commitments. People use diverse and creative methods to establish the credibility of commitments in advance, perhaps by making visible, irreversible investments or by setting up structural impediments to opportunistic mind changing. These methods often involve behavior that does not fit well into a competitive market model. Identifying such non-market guarantees of commitments in our society is one of the major contributions of TCE and new institutional economics.

The neoclassical approach focuses on the incentives and uncertainties in the contract itself, and emphasizes the design of the contract prior to its adoption. It assumes that people attempt to solve the problems of uncertainty and motivation in the terms of the contract. However, even under carefully structured incentives, there will be incomplete contracts and windows for opportunism. The neoclassical model assumes that contracts are well specified and can be enforced in courts of law. Where even this remedy breaks down and the parties cannot find a way to make satisfactory enforceable contracts, there must be some kind of market failure. There will be no agreement and thus no market; or there will be an agreement with externalities in which some part of the burden of non-compliance falls on parties without their consent, and thus not captured within the agreed-upon price.

However, the TCE model acknowledges the incompleteness of contracts, placing them in the context of ongoing economic relationships. Thus, in addition to the incentive structure of the contract, it examines the nature of the ongoing relationship and describes how the contract is governed over time. Clearly, resolving conflicts through the courts is expensive, and most of the time we use less costly arrangements or institutions, such as personal negotiation, mediation, and

arbitration. Within a community of individuals or interacting organizations, much governance can be carried out through social influence, reputation effects, and informal rewards and punishments. Within a firm, governance of internal transactions can be carried out through direct supervision with explicit power to reward and punish.

In fact, one of the defining characteristics of a firm is the degree of control it has over its assets and its employees.^[8] This control is largely free from oversight by the legal system, much more so than in the case of external market transactions. This brings us back to Coase's initial question about the existence of firms. Now we can say that firms provide a structure for governance that is an alternative to the market system and the legal system. Internalization occurs when the total of transaction costs, including integration costs, governance costs, and the potential costs of unfulfilled contracts, can be reduced by bringing transactions inside the firm's control.

Although it focuses on the transaction, TCE provides a powerful tool for the analysis of organizations. A firm is not just a contract, it is a long-term organizational structure for making and governing contracts, one of the many types of arrangement available. In contrast to the principal-agent approach that focuses on designing a contract to optimize its expected outcomes, the TCE approach shows that the very choice of the organizational structure in which contracts are embedded is an adaptation to the incompleteness of contracts. This adaptation can be understood as optimizing the expected net of transaction costs, integration costs, governance costs, and the costs of unfulfilled contracts over a long-term sequence of many transactions.

Such adaptations are not limited to the polar extremes of hierarchical firm versus competitive market. Many large firms are not strictly hierarchical in practice; the study of management control systems provides insights into the practical task of governing transactions within firms to obtain the motivational advantages of market competition and incentives without losing the cooperative and transaction cost advantages of hierarchical organization.^[9] There is a broad range of hybrid types of governance combining various elements of hierarchical integration and competitive market pricing. Claude Menard^[10] outlined how labor arrangements in organizations may vary between quasimarkets and direct hierarchies, with other forms in between, depending on the nature of the transactions, especially their uncertainty and asset specificity.

In addition, different types of adaptation occur at different stages in the same contract. Before a long-term contract is awarded, the process of obtaining competitive bids is essentially a market process; at that stage, the bidders are more or less interchangeable, and

asset specificity has not yet become a concern. Where there are enough bidders and enough opportunities for them to compete, the price for the contract as a package will be determined by a market pricing mechanism. In what is called the *fundamental transformation*, the situation changes after the contract is awarded, and the winning bidder establishes an asset-specific long-term relationship requiring non-market forms of governance. This transformation clearly happens with large industrial contracts, and to a lesser extent when employees are hired. But transactions such as commodity purchases that lack the characteristics of uncertainty, complexity, and asset specificity do not go through the transformation; they remain governed by the market pricing mechanism and conform more closely to the standard economic models.

PUBLIC ADMINISTRATION AND PUBLIC POLICY IMPLICATIONS

Efficiency in Government

Transaction cost economics suggests that government agencies, as well as firms, are significantly molded by transaction cost effects. Government agencies can be viewed as governance arrangements for public goods transactions, an especially necessary type of governance because of the high incentives for opportunism, called “free riding” in this context. Public sector governance is also affected by asset specificity; many of the human skills and physical assets employed by government could not be interchanged with private sector skills and assets without some loss of value. In public administration, the creative use of marketlike incentives where appropriate can help to make agencies more efficient, but these must be tempered by awareness of the TCE reasons that markets do not work for all types of transactions.

In an era of government reform, devolution, and downsizing, it pays to study how outsourcing and governance decisions are made in the private sector. Major changes in the structure and practices of government organizations inevitably bring about changes in transaction costs. Identifying these costs, and comparing them between existing arrangements and alternative proposals, can make an important contribution to the evaluation of reorganization plans.

Efficiency in Markets

In classical and neoclassical welfare economics, a competitive market equilibrium is regarded as the gold standard of economic efficiency. In this view, any organizational arrangements, social institutions, or

economic practices that interfere with or fall short of such an outcome must to that extent be inefficient. For example, the traditional view of antitrust policy has been that unfamiliar or non-standard business practices must be anticompetitive in intent or effect, and thus inefficient. However, TCE has shown in theory and in several empirical studies of utility markets^[10] that surprising and unusual forms of business relationships, hard to account for with standard economic models, can often be explained and justified as efficiency-promoting adaptations to transaction costs and the need for governance of complex long-term relationships. Where this is true, prohibiting such practices through antitrust regulation might increase costs to the consumer, not the effect intended by the regulators.

The theoretical and empirical task of applying TCE to various contexts is unfinished. But the antitrust example indicates that at least some apparent market failures are actually adaptations to transaction costs. This suggests that policy makers should consider the economic functions served by existing institutions and relationships, which can then be taken into account in proposing or evaluating government interventions that might alter transaction cost structures.

However, government investments in public goods and externalities that reduce everyone’s transaction costs may contribute more to the general welfare than interventions in such apparent market failures. Transactions take place in the context of a society’s legal system, its information infrastructure, and its regulatory and tax burdens on business firms. Thus, policies that might increase transaction efficiency across the board include dispute resolution alternatives in the legal system, improved communication and access to information through technology such as the Internet, simplifications of reporting requirements for businesses, programs to retrain workers displaced by economic and technological change that makes their specific skills unneeded. In general, this approach suggests that consumer and business policies should—wherever possible—reduce rather than increase the uncertainty of transactions, and increase rather than reduce the flexibility of parties to govern their own contracts.

REFERENCES

1. Coase, R.H. The nature of the firm. *Economica*. November 1937, IV, 386–405.
2. Williamson, O.E. *The Mechanisms of Governance*; Oxford University Press: New York, 1996.
3. Kenney, R.; Klein, B. The economics of block booking. *J. Law Econ.* October 1983, 26, 497–540.
4. Turvani, M. Illegal markets and new institutional economics. In *Transaction Cost Economics: Recent*

- Developments*; Menard, C., Ed.; Edward Elgar: Cheltenham, UK, 1997; 127–148.
5. Weingast, B.R.; Marshall, W.J. The industrial organization of congress; or, why legislatures, like firms, are not organized as markets. *J. Polit. Econ.* **1988**, *96* (1), 132–163.
 6. Masten, S.E.; Meehan, J.W.; Snyder, E.L. The costs of organization. *J. Law Econ. Organ.* **Spring 1991**, *7* (1), 1–25.
 7. Simon, H. Theories of bounded rationality. In *Decision and Organization*; McGuire, C.B., Radner, R., Eds.; American Elsevier: New York, 1972; 161–176.
 8. Williamson, O.E. Hierarchies, markets and power in the economy: an economic perspective. In *Transaction Cost Economics: Recent Developments*; Menard, C., Ed.; Edward Elgar: Cheltenham, UK, 1997; 1–29.
 9. Anthony, R.N.; Govindarajan, V. *Management Control Systems*, 9th Ed.; Irwin McGraw-Hill: New York, 1998.
 10. Menard, C. Internal characteristics of formal organizations. In *Transaction Cost Economics: Recent Developments*; Menard, C., Ed.; Edward Elgar: Cheltenham, UK, 1997; 30–58.

Education Policy

Frederick M. Hess

*Virginia Center for Educational Policy Studies, University of Virginia,
Charlottesville, Virginia, U.S.A.*

Joleen R. Okun

University of Virginia, Charlottesville, Virginia, U.S.A.

INTRODUCTION

Education policy includes the world of elementary and secondary, postsecondary, and early childhood education. It involves measures ranging from organization of traditional schooling to the setting of college tuition to the provision of programs intended to help nurture preschoolers. In this sense, education policy is the public policy area in which American government is most heavily invested. In 2001, all levels of government—federal, state, and local—spent well in excess of \$647 billion on educational provision. Approximately 60% of this amount is devoted to providing K–12 schooling, which serves 33.5 million children, while the remaining 40% of these funds are used to support college and university programs.

During the late 1990s and through 2002, the American public consistently listed education as one of its top two or three national concerns. This marked a sharp break with the nation's past, as schooling (and, by extension, education policy) had traditionally lagged near the bottom of the public's policy concerns. Education began to gain increased attention in the mid-1980s, and it began to move up the list of prominent public concerns during the late 1980s and early 1990s.

EDUCATION POLICY

The substance and shape of U.S. education policy has shifted radically over the course of the nation's history. Government involvement in educational provision was haphazard until the late nineteenth century, and systematic public provision of both K–12 schooling and higher education only took on their contemporary form during the first half of the twentieth century. Before that time, schooling remained primarily a locally funded and locally governed service that rarely attracted significant policy attention. Schools were only standardized at the state level, as sustained federal involvement in higher education did not begin until after World War II and did not begin in K–12

or preschool education until the launch of the Great Society programs in the mid-1960s. In fact, the U.S. Department of Education was not created until 1979. Given the scope and largely decentralized nature of public schooling, affecting change is often difficult. This is especially evident in light of the fact that the United States currently has over 15,000 local school districts and 50 state boards of education that coordinate their efforts with the Department of Education.

Brief History of Education Policy

From the nation's founding, education occupied a central place in the imagination of America's leaders. Founders like Benjamin Rush and Thomas Jefferson championed the importance of an educated citizenry, calling for a system of public schools and the creation of state-supported higher education. Throughout the colonial period and the early nineteenth century, however, education policy consisted of little more than the state supplying land for school houses and a small amount of money to support the operation of grammar schools.

The first important move toward what can more conventionally be understood as “education policy” was Horace Mann's crusade as President of the Massachusetts Board of Education (1837–1849) to reform the state's schools, enhance the caliber of teachers, and increase the level of instruction. Much of the support for Mann's effort, and for similar efforts elsewhere, was inspired by a desire to “Americanize” the new waves of immigrants populating the United States at that time. In fits and starts, such efforts proceeded across the nation in the latter half of the nineteenth century, as state after state adopted compulsory education programs that required children to attend at least some grammar school. Between 1890 and 1940, these measures were developed into an ongoing commitment to nearly universal education. The population of students attending school and graduating from the K–12 system grew at a dramatic rate, partially propelled by state provisions for compulsory attendance. Although the first such law was

passed in 1853, it was not until the turn of the century that an overwhelming majority of states had enacted such legislation.

During this time period, the federal government slowly began to increase its involvement in education policy. This began in 1862, when Congress passed the Morrill Act, a provision that earmarked land to support the construction of public universities across the nation. The act marked the first time the federal government sought to involve itself directly in schooling or education. At the K–12 level, the first significant federal efforts would not occur until 1917, when Congress provided fiscal incentives for vocational education in the form of the Smith–Hughes Act.

Different Forms of Education Policy

Education policy serves a broad range of student needs, encompassing much of a broader role outside traditional K–12 schooling and higher education. In the twentieth century, the federal government began to further address the needs of discrete populations by implementing new programs. Those detailed have been among the most popular and widely cited examples of such interventionism.

Early childhood

The federal government made its first foray into preschool education by establishing the Head Start program for disadvantaged preschool children under the Economic Opportunity Act of 1964. Aimed initially only at poor children, its purpose was to organize programs that would prepare preschool children for elementary school. Money was appropriated through the Office of Economic Opportunity, which made individual grants to cities and other localities to set up Head Start centers. In 1969, the program was transferred to the Department of Health and Human Services. It was later extended to children above the poverty level, whose parents, however, had to pay according to their income.

Special education

It was not until the late 1960s that education policy addressed the equalization of student opportunity throughout the United States. A new concern for persons with disabilities was addressed by the Johnson administration. The watershed in the development of these programs was the federal government's landmark legislation in 1975, entitled the Education for All Handicapped Children Act (EAHCA), which was incorporated into a new law, the Individuals with Disabilities Education Act (IDEA) of 1990. In 1997,

Congress reauthorized and amended IDEA, expanding the purpose of the law to emphasize outcome measures for preparing disabled children for employment and independent living rather than that of merely requiring access to public education.

Changing Governmental Role in Education Policy

The federal government initially served a distributory, rather than a regulatory, role in the American educational system. From the early nineteenth century, Congress issued state land grants to provide for common schools. In conjunction with the federal requirement that all states draft a constitution, this was the only role the federal government played in the establishment and administration of state-run school systems.

Education occupied a prominent place in state constitutions, yet schools were largely locally run. Many voters were distrustful of strong, centralized authority during the nineteenth century and resisted moves toward further state control. Throughout this era, educators sought to replace haphazard standards of teacher training and certification, supervision, and curriculum by consolidating school districts and extending state authority over public education. These efforts led primarily to symbolic efforts that were not standardized throughout the state educational system.

However, with the advent of the Progressive era in the early 1900s, greater standardization was achieved. School leaders used state laws to bureaucratize and centralize public schools. This push for change relied heavily on professional expertise rather than lay participation in decision making. Educators and their allies contended that a modern system of education demanded increased consolidation, standardization, and centralization. In light of this, state legislators enacted such changes in spite of public preference for a more locally run educational system. At this time, schools became more heavily bureaucratized, with state-mandated provisions for curriculum, teacher status, administrative organization, and consolidation.

Shift to centrally dictated outcomes, but increased operational freedom

One of the central reforms in education policy throughout the 1990s was a standards movement, which was implemented as states and school districts sought to enhance student achievement. Although earlier reform movements were characterized by either a “top-down” (i.e., mandates) or “bottom-up” (i.e., local control) approach, standards-based reform combines both approaches to enable states and districts to

define the focus and expectations for educational outcomes and to hold educators accountable for meeting these aims. This paradigm involves both a set of clearly defined student performance goals for schools to strive toward (usually content standards), and an accountability system that provides a set of incentives for schools or districts to achieve the standards and accompanying penalties for failing to move toward them (rewards and sanctions). Yet, it also provides districts and schools with greater autonomy to make decisions that they believe will improve instruction and student outcomes. Policy makers recognize that instructional improvement needs to be motivated and developed at the school and classroom level, and they are developing ways to give schools and teachers increased autonomy to make decisions that affect student learning.

New Trends in Education Policy

Accountability

Since the 1990s, a growing number of states have created comprehensive accountability systems designed to increase student learning in public schools. These accountability systems are based on “high-stakes” standardized testing of a state curriculum. Local educators face consequences based on how well their students do on these exams. They receive rewards for good student performance and are subject to interventions to rescue children from low-performing schools. This shift toward comprehensive accountability systems includes set goals for student achievement, specific outcome measures for these goals, incentives for educators who yield good student outcomes, and interventions for students who may be underperforming in so-called failing schools.

State policy makers have been reluctant to extend a great deal of flexibility to districts, and are increasing using high-stakes testing to measure student outcome and guide curricular decisions. States for many years have been accumulating testing requirements that their legislatures, state officials, or local administrators have chosen. As these outcome measures continue to be widely used in a growing number of states, assessment has been inextricably aligned with the centralized testing of student achievement with standardized instruments. Student performance goals were made more explicit so that testing could be more precisely focused, and efforts were made to align curricula with the testing.

School choice

As accountability measures are increasingly being used to determine how well public schools are serving

students, the issue of providing parents with options to ensure enhanced achievement has also taken flight. During the 1990s, school choice has become a pervasive topic in education policy, as over two-thirds of state legislatures passed provisions allowing the establishment of charter schools.

In 1994, federal charter school legislation was passed that exempts these institutions from state and local regulations that inhibit flexible management, yet provides that these schools be guided by public control. A charter school possesses much greater autonomy than a traditional public school. Charter schools are free from many district, state, and union regulations or requirements, including those governing curriculum, teaching methods, contracting for services and facilities, and the hiring of personnel. In exchange, charter schools are held accountable for student performance. Yet, although charter schools have moved to a national forum as a form of school reform, this movement has developed in different ways and at varying rates at the state and local level. The federal government has had only a limited role in the development of charter schools, leaving state charter legislation broad and varied. Such measures may constrain who is eligible to obtain a charter, how many such schools may operate, and to what measures of accountability these schools will be responsible.

A less widely implemented and more ardently contested method of school choice has been the proposed use of vouchers to allow parents to more actively guide their children's education and provide more options to those who could otherwise not afford private schooling. This means of school choice involves encouraging open competition within the school system so that schools are held directly accountable for meeting students' needs. Proponents argue that giving parents a tax credit and allowing them to use it to dictate their child's school of choice allows both schools the opportunity to change to attract students, and parents and students the opportunity to decide what educational opportunity best suits their needs. Voucher plans range from public plans, which would allow students to attend schools according to either inter- or intradistrict regulations to private plans, which would allow students to attend private or parochial schools.

Home schooling

A trend antithetical to the historical growth of compulsory schooling for all children has grown increasingly prominent. In the latter part of the 1990s, state legislatures increasingly amended state compulsory attendance laws to permit home schooling. Regulations on attendance, curriculum, and other requirements vary from state to state, as home schooling became

increasingly popular in the first years of the twenty-first century. The Department of Education estimated in 2001 that over 500,000 students were currently home schooled in the United States.

Tuition support/HOPE scholarship

The 1997 Taxpayer Relief Act provided two major provisions, which make pursuing higher education less financially burdensome, the Hope Scholarship Credit and the Lifetime Learning Credit. The Hope Scholarship Credit allows taxpayers to claim up to \$1500 per student for each of the first 2 years of study at an approved institution of higher learning. The Lifetime Learning Credit applies to those who are beyond the first 2 years of higher education or pursuing part-time enrichment to improve or upgrade their job skills. This provision provides families with a 20% tax credit for the first \$5000 of tuition and fees through 2002, and for the first \$10,000 thereafter.

CONCLUSION

The field of education policy has become a much more varied, contested process. New initiatives prove to play out differently among various states, as the government attempts to better serve the educational needs of

children in the United States. The current direction appears to be toward enhanced accountability for certain outcomes, while retaining a greater degree of choices for parents and students. The system of public education has grown from a provincial, non-mandated institution to a wide-ranging system that nearly every American experiences on a multitude of levels throughout their lives.

BIBLIOGRAPHY

1. Alexander, K.; Alexander, D.M. *American Public School Law*, 5th Ed.; Wadsworth Group: Belmont, CA, 2001.
2. Linn, R.L.; Herman, J.L. *Standards-Led Assessment: Technical and Policy Issues in Measuring School and Student Progress*; CSE Technical Report 426, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation, Graduate School of Education and Information Studies, University of California: Los Angeles, 1997.
3. National Center for Education Statistics. *NAEP 1996 Trends in Academic Progress*; Campbell, J.R., Voelkl, K.E., Donahue, P.L., Eds.; U.S. Department of Education: Washington, DC, 1997. NCES 97-985.
4. Popho, C. Standards, assessment, accountability: the tangled triumvirate. *Phi Delta Kappan* **1997**, 78 (9), 673–674.

Education, State Administration of Primary and Secondary

Thomas C. Sutton

Department of Political Science, Baldwin-Wallace College, Berea, Ohio, U.S.A.

INTRODUCTION

The administration of primary and secondary public education in the United States is conducted primarily by state governments in conjunction with local school districts. All states with the exception of Hawaii employ the structure of local school districts as the direct administrators of public primary and secondary education. Local administration of public schools dates back to the seventeenth-century New England colonies, where communities would collect a property tax for the purpose of hiring a “tutor” to teach the children of the village and surrounding farms. The importance of public education has been emphasized by states since the Declaration of Independence, evidenced by the inclusion in all 50 state constitutions of a clause stipulating that the state government is responsible for ensuring the provision of a system of public education. The constitutional provision for each state established the legal responsibility of state governments for ensuring the effective administration of public education, either directly or through local school districts.^[1]

State involvement in public education has grown significantly during the twentieth century, driven in part by public calls for improvements to public education. With education being viewed as a key element of both civic engagement in a democratic society and as the basis for skill training necessary for gainful employment in a rapidly changing economy, American society has made many demands on school districts. New policy developments and mandates such as integration, the development of special education services for students with physical and learning disabilities, vocational education, improving curriculum and facilities, and increasing teacher preparedness have been developed usually as initiatives of state legislatures and administered through state departments and boards of education.

STATE AND LOCAL ADMINISTRATIVE STRUCTURES

Although local school districts usually predate the formation of states and state constitutions, they are considered to be legal extensions of state authority.

State administration of primary and secondary public education (herein referred to as public education) begins with the granting of charters allowing the formation of geographic units as school districts. School district boundaries conform with either county or municipal boundaries. School district residents elected representatives to a board of education, which is responsible for oversight of all administrative concerns of the district, including financial management, personnel, facilities, curriculum, and student conduct. School boards range in number from 3 to as many as 15 members who serve on a voluntary basis (some large municipal districts provide a small stipend) and are usually elected at large. Boards of education hire a superintendent who acts as the primary administrator of the district, and who in turn is responsible for hiring building principals; teachers; service staff, such as custodians, bus drivers, and secretaries; and supporting administrative staff to handle district finances, curriculum, and personnel issues.^[2]

Paralleling the local district model, states created boards of education responsible for the oversight of public education. State boards range from elected members representing state districts and at large seats to entities appointed by governors and legislators, to mixed boards of elected and appointed members. State boards oversee a state superintendent of public instruction who may be appointed by the board and/or the governor, or may be elected. The state superintendent directs a state department of education that is responsible for administering all state public education mandates, policies, and finances. Occasionally, such departments will have authority for a specific program removed due to concerns of accountability by legislators. An example of this is Ohio, which began a significant program of state funding for local school facility construction and repair in 1997, with the creation of a separate School Facilities Commission that reported directly to the governor and the state legislature.^[3]

Despite the somewhat unique structure within state government, it is often the governor rather than the board of education and/or superintendent who is the most prominent advocate for public education, especially in periods of reform. Governors of southern states took the lead in the 1980s, seeking to improve teacher compensation and training and school curriculum in the wake of the 1983 “Nation at Risk” report

issued by the National Commission on Excellence in Education. A second effort at reform was led by the National Governors Association in 1991, with its issuance of "Goals 2000," a set of benchmarks for student achievement and preparedness endorsed by President Bush as the blueprint for national policy priorities in public education.^[4]

At times, other statewide offices are involved in public education administration. State auditors may become involved through periodic examination of school district finances and intervention in districts with systems that are weak or in crisis. State attorneys general have often served as the lead defense counsel for states that have been sued for issues ranging from integration to provision of special education programs to school finance inequities. Secretaries of State oversee implementation of local elections and levy votes through county election boards, while lieutenant governors are often given responsibility for oversight of special education initiatives related to safety, such as reducing school violence and services for juvenile delinquents.

State legislatures play an important role in public education administration through their growing involvement in policy development and evaluation. State legislatures have passed statutes stipulating exact credit and course distribution requirements for high school diplomas; required degrees, coursework, internships, and testing for teachers and administrators; and mandatory proficiency testing as a graduation requirement, along with the formulas for state aid to local school districts. The activist role of governors and state legislatures is an important reason why state school boards and departments of education have been relegated to largely passive roles as implementers of policy.

SCHOOL FUNDING

School districts are funded through two primary sources: property taxes and state aid. Historically, property taxes served as the primary source of income for districts. Currently, the percentage of funding from property taxes ranges from 40% in Michigan and Kentucky to 70% in New Hampshire, Wyoming, and Vermont.^[5] Property taxes are levied as millage, which is a unit equaling one-tenth of one cent on each dollar of assessed property value. States allow a minimum of between 10 and 20 mills to be levied by districts without voter approval. Additional millage requires voter approval and is usually for a specific period of time, requiring periodic renewal votes. The rapid rise of property values in the 1970s led to a property tax revolt by voters exemplified by California's Proposition 13, which froze property taxes and required voters to approve all new increases, including changes to adjust for inflation. As a result, districts in states such as

California, New York, and Ohio frequently have new tax levies on the ballot that are necessary to pay for cost increases due to inflation rather than for improvements to the district's educational capacity.^[6]

States use various funding formulas to supplement local property taxes. The most common is the foundation aid formula, which attempts to guarantee a minimum per-pupil spending level for every district, adjusted for variations in local cost-of-doing-business factors. A few states use a foundation grant program, which provides flat grant amounts of varying levels determined by the size and property wealth of a district, as opposed to attempting to establish a minimum spending floor for every student.^[7]

The basis of funding in property taxation inevitably results in fiscal disparities between districts with low and high tax capacity and property wealth. These disparities have persisted despite growing levels of state aid, with districts in some states having as wide as a \$10,000 difference in spending per pupil between the poorest and richest public school districts. As a result, school districts began suing states in state courts, alleging that such disparities violated state constitutional guarantees of accessible public education. In 1973, the *San Antonio v. Rodriguez* case was appealed to the U.S. Supreme Court, seeking a ruling on whether funding disparities violated the due process and equal protection clauses of the Fourteenth Amendment. The court ruled that public education is not a guaranteed right under the U.S. Constitution, sending such cases back to state courts. Since the San Antonio decision, 38 states have experienced school finance litigation, with 23 state supreme courts ruling systems of school funding to be in violation of state constitutional provisions for accessible public education.^[8] Cases in some states have dragged on for years through multiple appeals and rehearings.

Urban school districts face special funding challenges. Higher poverty rates produce low residential property values as well as a higher percentage of students with special social and educational needs, incurring higher costs than suburban districts. Added to these factors are the higher costs of urban areas, the presence of strong teacher and service worker unions that drive up wage levels, all compounded by the endemic inefficiency of large administrative bureaucracies that are part and parcel of large urban districts.^[2] A few of the school funding litigation cases have specifically focused on the needs of urban districts, in particular, the New Jersey case of *Abbott v. Burke* (1985).^[9]

CURRICULUM AND PERSONNEL

States are responsible for setting standards for school district curriculum, teacher, and administrator

certification, and are often also involved in oversight of facilities. Curriculum is determined by state standards that mandate a set number of courses in each discipline at the secondary school level necessary to obtain a high school diploma. Primary school curriculum is usually measured by student achievement benchmarks for each grade, such as achieving literacy by fourth grade and arithmetic proficiency by fifth grade. Secondary level (grades 7–12) course credit is measured in Carnegie Units, which are determined by the number of minutes a course meets each week during the four quarters of a school term. States also set the number of days schools must meet each academic year, usually ranging from 175 to 200 days. Districts in many states determine start and end dates, as well as holiday schedules. States will allow a set number of cancelled days, beyond which days a district is closed must be made up by extending the school year or shortening vacation periods (a particularly nagging problem for northern state districts with frequent snowstorms).

States also administer certification for teachers and school administrators. Teachers are required to have a minimum of a Bachelor's degree in education, combined with field-specific coursework in the disciplines the teacher will be using in their career. Most states have shifted from permanent certification to certificates for periods of between 5 and 10 years, requiring teachers to achieve additional professional development to obtain continued certification. Professional development usually takes the form of graduate-level coursework. When combined with district pay schedules that pay incentives to earn graduate degrees, the result is that a majority of career teachers now hold Master's degrees in education. Administrators are usually drawn from teacher ranks. They are also required to earn a Master's degree or graduate certificate in education administration before becoming administrators. Education subfield degree specializations include special education, reading instruction, gifted and talented education, vocational education, counseling, curriculum development, and management.^[2]

Personnel issues include an impending shortage of teachers and administrators. High job stress, lower compensation than in other fields with equivalent educational requirements, and negative public perceptions of public education have contributed to the shortage, and are issues requiring response in order to effectively recruit and retain future teachers and administrators.

STUDENT ACCESSIBILITY AND SERVICES

Public schools are much more accessible than ever before. In 1954, the *Brown v. Board of Education* decision ushered in 30 years of desegregation orders applied to local school districts throughout the United

States. Although desegregation successfully tore down the system of separate and unequal schools within school districts, the success of tactics such as busing, mandatory student assignments to buildings, and magnet schools is mixed. Demographics of urban districts changed rapidly, leading to the situation in most urban areas of poorer urban districts with a majority of minority students surrounded by predominantly white, middle-class, suburban districts. As a result, providing strong public education to students who need it most continues to be a major challenge, compounded by the reflected demographic changes in state legislatures, which are less interested in spending more tax dollars on failing schools with fewer constituents.

Federal funding tied to mandates for provision of special education combined with the Americans with Disabilities Act of 1991 constitutes a mandate to states and local districts to provide special education to all eligible students. Such services are expensive, and are often not adequately funded by state and federal dollars. States usually operate a formula aid program that sends funding to local districts based on how many students require special education services. School district resources have been stretched in recent years by implementation of the philosophy of “mainstreaming,” which requires that special needs students receive services while attending regular classes with non-special needs students, rather than being taught in separate, specialized classrooms.

ISSUES OF ACCOUNTABILITY

Since the early 1990s, schools and teachers have been under increasing scrutiny over issues of accountability. State legislatures have responded by requiring proficiency testing at various levels of primary and secondary education. Many states require passing a high school proficiency test, usually taken in the ninth or tenth grade, to earn a high school diploma. A major concern has arisen over the social promotion of students who fail to achieve grade-level skill development, but are advanced with their age peers. States are experimenting with proficiency tests in the primary grades to measure literacy and math skill attainment. Students who fail these tests receive extra instruction or summer school, or may be held back until passing the tests.

Proficiency testing has been controversial. Proponents see testing as a way of uniformly measuring and guaranteeing minimum standards for student achievement, thus avoiding the stories of illiterate high school graduates. Opponents argue that proficiency testing limits curriculum, with teachers focusing on “teaching to the test.” Others are concerned that proficiency tests disproportionately affect minority students in

urban districts. President George W. Bush included proficiency testing at every grade level through tenth grade as a major component of his education reform package. The Bush plan is to provide incentives and support to help schools with low test passage rates to improve, but to remove the funding if a school district fails to show improvement after 3 years.

Critics of failing urban school districts advocate the use of vouchers and charter schools as ways to introduce competition into public education. Proponents believe that public schools will improve if faced with losing students and funding to private schools. Vouchers are public dollars given to poor families to use to pay for private school tuition. Voucher programs have been in place since the late 1990s in Milwaukee, Wisconsin; Cleveland, Ohio; and Florida.^[10] The use of public dollars to fund education at religious schools prompted a legal challenge under the First Amendment. The Cleveland voucher program will be heard by the U.S. Supreme Court during the 2001–2002 term. Charter schools are private schools chartered by the state for specialized education that are often exempted from regulations governing school districts, to encourage innovative curricula and teaching. Rather than paying for individual tuition at private schools, states subsidize the expenses of the entire charter school. Although appealing as a form of school choice, charter school evaluations in Ohio and other states have shown mixed results in terms of improving student achievement.

REFERENCES

1. Kirp, D.L.; Driver, C.E. The aspirations of system reform meet the realities of localism. *Educ. Adm. Q.* **1995**, *31*, 589–612.
2. Chance, W. *The Best of Educations: Reforming America's Schools in the 1980's*; The MacArthur Foundation: Chicago, 1986.
3. Browning, R.G. *School Funding Policy Recommendations. Unpublished Report to the Ohio School Funding Task Force*; State of Ohio: Columbus, OH, 1997.
4. Education Commission of the States. *The Progress of Education Reform*; Education Commission of the States: Denver, CO, 1996.
5. DeBartolome, C.A.M. What determines state aid to school districts? A positive model of foundation aid as redistribution. *J. Policy Anal. Manage.* **1997**, *16*, 32–47.
6. Gionocchio, J.S.; Hamilton, H.D. Fiscal policymaking by plebiscite: local tax and bond referenda in Ohio. In *Political Behavior and Public Issues in Ohio*; Gargan, J.J., Coke, J.G., Eds.; Kent State University Press: Kent, OH, 1972.
7. Enrich, P. Leaving equality behind: new directions in school finance reform. *Vanderbilt Law Rev.* **1995**, *48*, 101–194.
8. Evans, W.N.; Murray, S.E.; Schwab, R.M. Schoolhouses, courthouses, and statehouses after *Serrano*. *J. Policy Anal. Manage.* **1997**, *16*, 10–31.
9. Odden, A. School finance reform in Kentucky, New Jersey, and Texas. *J. Educ. Financ.* **1993**, *18*, 293–317.
10. Chubb, J.E.; Moe, T.M. *Politics, Markets and America's Schools*; The Brookings Institution: Washington, DC, 1990.

Efficiency

Hwang-Sun Kang

Department of Urban Management, Seoul Development Institute,
Seoul, Republic of Korea

INTRODUCTION

This article introduces the concept of efficiency and discusses the validity of the concept being used as the primary criterion of performance measurement in the public sector. From Wilson, the father of public administration in 1887 to Osborne and Gaebler, advocates of reinventing government in 1992, efficiency has been said to be the first ideology of public administration. Wilson pinpointed the role of government: government should be able to implement delegated tasks from the legislature. Efficiency is defined as the ratio of outputs to inputs. It describes the cost of resources used for an activity to achieve given outputs. This entry shows how the formula applies to prevalent management techniques such as cost–benefit analysis and data envelopment analysis. This entry introduces an argument related with efficiency. When a public policy is analyzed based on the ideology of efficiency, it is questionable if government performance could be measured by the definition of efficiency. Although the definition emphasizes final outputs of government activity, many scholars hold that it is extremely difficult to measure the final facet of government activities. It is neither desirable nor reasonable to measure government performance based on economic (monetary) efficiency. This article holds that there are two kinds of efficiency: technical efficiency and social efficiency. When government performance is evaluated based on the rationale of efficiency, those two concepts should be considered.

DEFINITION OF EFFICIENCY

From Wilson, the father of public administration in 1887, to Osborne and Gaebler, advocates of reinventing government in 1992, efficiency has been said to be the first ideology of public administration. Wilson pinpoints the role of government: government should be able to implement delegated tasks from the legislature. A motto of reinventing government: “Cost Less and Work Better,” also represents the underlying idea of the definition of efficiency.

Efficiency is defined as the ratio of outputs to input. It describes the cost of resources used for an activity to

achieve given outputs. Here, it should be noted that the numerator of the formula of efficiency should be the final outputs of government activity. For instance, “caseload per worker” is not an efficiency measure, because caseloads are not outputs.^[1]

Fig. 1 describes the definition of efficiency and its relationship with effectiveness. Ammons^[2] illustrated some examples of efficiency measures, indicating the difference from other related definitions (Table 1).

MANAGEMENT TOOLS BASED ON EFFICIENCY

Although the basic logic of efficiency is simple, it is the ultimate idea of many management techniques used in the public sector. Cost–benefit analysis that is one of the most traditional techniques of public-policy analysis is based on the simple formula of efficiency. Activity-based management skill imported from the private management arena is also using efficiency as its basic logic.

Another powerful skill newly emerging in the public sector, data envelopment analysis (DEA), adopts efficiency as its basic logic. Interestingly, DEA introduces a new term related to efficiency: relative efficiency. The DEA determines a non-parametric best practice frontier and then measures efficiency relative to that frontier. Thus, DEA is also called “frontier analysis.” DEA method does not use any statistical inferences, rather, it is a mathematical programming technique developed by the operational research arena focusing on business firms and not-for-profit organizations to find a best practice unit among multiple units (decision-making units). DEA has many advantages:^[3]

- DEA can handle multiple input and multiple output models.

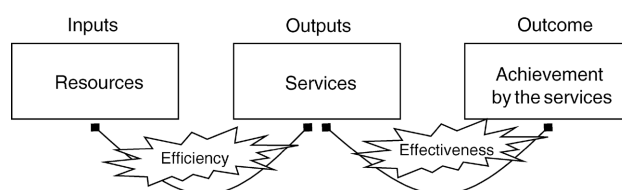


Fig. 1 Relationship between efficiency and effectiveness.

Table 1 Examples of four representative measures of government performance

Municipal function	Workload measure	Efficiency measures	Effectiveness measures	Productivity measures
City clerk	Number of sets of city council meeting minutes prepared	Employee hours per set of city council minutes prepared	Percentage of city council minutes approved without amendment	Percentage of city council minutes prepared within 7 days of the meeting and approved without amendment
Library	Total circulation	Circulation per library employee	Circulation per capita	Circulation per 1000
Meter repair	Number of meters repaired	Cost per meter repair	Percentage of repaired meters still functioning properly 6 months later	Cost per properly repaired meter (i.e., total cost of all meter repairs within 6 months)
Personnel	Job application received	Cost per job application processed; cost per vacancy filled	Percentage of new hires and promotions successfully completing probation and performing satisfactorily 6 months later	Cost per vacancy filled successfully (i.e., employee performing satisfactorily 6 months later)

Source: From Ref.^[2].

- It does not require an assumption of a functional form relating inputs to outputs.
- Decision-making units are directly compared against a peer or combination of peers.
- Inputs and outputs can have different units. For example, X1 could be in units of lives saved and X2 could be in units of dollars without requiring an a priori trade-off between the two.

ARGUMENT RELATED TO EFFICIENCY

There is an argument related with efficiency. When a public policy is analyzed based on the ideology of efficiency, it is questionable if government performance could be measured by the definition of efficiency. Although the definition emphasizes final outputs of government activity, many scholars hold that it is extremely difficult to measure the final facet of government activities. Especially, it is neither desirable nor reasonable to measure government performance based on economic (monetary) efficiency.

Considering this argument, it is necessary to expand the definition of efficiency. There are two kinds of efficiency in the social science arena: technical efficiency and social efficiency.

Technical efficiency refers to the general meaning of efficiency that indicates the rates of inputs to outputs. In contrast, social efficiency discusses not only economic (monetary) efficiency but also social relevance of a government activity. Even if a government activity is implemented efficiently economically, the government activity should be interpreted on whether it is compatible with social needs (social ideology or social relevance).

REFERENCES

1. Berman, E. *Productivity in Public and Nonprofit Organizations*; Sage Publication Inc.: Thousand Oaks, CA, 1998.
2. Ammons, D. *Municipal Benchmarks*, 2nd Ed.; Sage Publication Inc.: Thousand Oaks, CA, 2001.
3. <http://www.emp.pdx.edu/dea/wvedea.html>.

E-Government in East Asia

Hee Joon Song

*Department of Public Administration, Ewha Womans University,
Seoul, Korea*

INTRODUCTION

The East Asian countries, including China, Japan, and Korea, have experienced different modernization processes, which evolved into the diverse political and economic systems today. There exists not only the economic gap between the developed and least-developed countries, but also the political gap between the liberal democratic countries and the socialist dictatorship in the East Asia. Such disparities in political and economic systems cause the differences in objectives, expected outcomes, project contents, and development strategies of the e-government among the East Asian countries.

CHINA

The national informatization (xinxihua) has been regarded by political leaders, such as Deng Xiaoping and Jiang Zemin, as the key tool to facilitate the economic development since 1980s. In 1993, the State Council launched 17 Golden Projects, aiming to develop the IT industries and improve the efficiency of ministries. The projects included 1) the Golden Customs, an integrated data system to link trading companies, banks, and customs; 2) the Golden Bridge, the national network for economic information; and 3) the Golden Card, the unified payment clearance systems.^[1] In January 1999, the State Economic and Trade Commission with the China Telecom initiated the Government Online Project (GOP), in order to interconnect the 40 central ministries, all provinces, autonomous regions, and municipalities.^[2] The GOP aimed to make information on the government policies available on-line for the public to access and to promote the government's image. It also targeted to gradually digitalize internal processes and services of ministries and to build their websites. The Enterprise Online Project (www.sinoeol.com) started in 2000, in order to facilitate B2B and B2C e-transactions. Meanwhile, the Households Online (www.sinohome.com), which initiated soon afterward, focused on establishing household networks. Implementing the 10th and the 11th Five Year Plan from 2001 to 2010, China seeks to significantly reduce the digital gap between China and other developed countries and to rebuild the country

with strong IT power, which would enable "the whole Shaokang Society," where everybody enjoys a wealthy life.

One of the recent outcomes of the above projects has been the launching of an official website of the Central People's Government of the People's Republic of China (www.gov.cn), which opened in January 2006. This website, available both in English and in Chinese, consists of four sections—China Fact File, Services, Business, and China Today—that provide information on the structure, function and key persons of the central government, services for citizens, businesses and foreigners, a variety of search engines, and the communication channel between the government and people. Many local governments, including big cities such as Shanghai, Shenzhen, and Quindao, also have established e-government infrastructures and networks. An example that reflects the successfully implemented e-government projects is the Shanghai city website (www.shanghai.gov.cn), which allows citizens to access 703 government-run businesses and to download 2400 forms of public services. Moreover, the Zhongguancun Business E-Park (www.zhongguancun.com.cn) pursues interactivity, efficiency, and transparency of public service.^[3]

Although the e-government in China has been carried out as a part of the Five Year Plans for economic development, with enthusiastic supports of the political leaders, the front office services has not reached the stage of consolidating the cross-agency business processes yet. Furthermore, it is important to note that Chinese perspective toward the participatory, open e-government, especially that affects the policy-making process, is rather dissenting. While the Chinese government utilizes the Internet as a driving engine to facilitate economic development, the country contradictorily recognizes the technology as a factor that should be monitored and tampered, in order to prevent the democratic participation of citizens that might challenge the regime.^[4]

JAPAN

In 1993, Japan's Cabinet proposed a plan to promote the cross-ministerial informatization, the use of administrative information, the improvement of

public services to citizens, and the innovation of administrative processes. In 1994, the government also developed the New Social Capital Plan to enhance the quality of the high-speed network, to create the early market demand on the high-speed network as a leading user, and to give financial supports to the private sector to reduce its investment burden. The Basic Plan for the Informatization of the Administration (1995–1999) prioritizes in establishing the cross-ministerial information sharing by implementing the network, the e-mail and e-document exchange system, the system to support the national council, the DB system of the white papers and research reports, and the system to search for the location of administrative information. The Basic Plan was extended in 1998 to “the Electronic Government Towards the 21st Century (1998–2000),” which included the on-line publishing of information on the location of administrative information, the provision of on-line public services such as e-application and e-filing, the digitalization of internal processes and documents, and the information and database sharing on-line. While pursuing the informatization of the administration, the ministries developed the systems to manage their internal processes and specific tasks and provided the integrated service connecting the central and local government.

E-Japan Strategy, developed to build the global IT leading country in 2000, included the plans for establishing the network that would enable people to access the Internet any time at a low price, thus improving the e-commerce system and market regulation. This, in turn, would realize the e-government and enhance the human resource development at the same time. In particular, the Japanese e-government replaced the hard-copied document with electronic information, making computerized public services like e-application and e-filing available. The paperless administration also allowed information sharing through the network. E-Japan Strategy II was developed in 2003, aiming to build the convenient society, where people enjoy sound and secure lives and share new satisfaction. It planned 1) to apply IT to seven areas—medical service, food, living, finance for small-and medium-sized enterprises, learning, working, and public service—in order to improve their efficiency and 2) to create the ubiquitous network environment required for the utilization of the next-generation IT. Both strategies contributed to Japan’s success in developing and utilizing the global top-level high-speed Internet infrastructures, lowering the telecommunication prices, distributing multifunctional mobile phones, and improving e-commerce environments.

Finally, in January 2006, Japan set up an agenda, “New IT Reform Strategy: Realizing Ubiquitous and Universal Network Society Where Everyone Can

Enjoy the Benefits of IT” (www.kantei.go.jp) and has pursued the e-government action plan (2006–2010). The plan includes more than 50% increase in the use rate of on-line services, the promotion of e-government services, the reduction of expenditure and service-processing time, and the optimization of cross-ministerial processes (www.e-gov.go.jp/doc/scheme.html).

Overall, Japan did not lead the world in e-government in 1990s, in spite of its high economic and technological capability. While the country successfully digitized the internal process of the central and local governments and promoted on-line publishing and utilization of administrative information, it failed to realize the integration of cross-ministerial work and e-participation and interaction between the government and the citizens. As Japan has steadily expanded the range of on-line services available to its citizens and businesses in 2000s, it is expected to develop the top-level e-government within a short period, having outstanding IT capacity.^[5]

KOREA

The e-government has been the core of various master plans on national informatization since 1980s. The development of Korean e-government can be largely classified into three steps according to its target, content, and characteristics changes that reflect different political leaderships, information technologies, work types to be automated, linkage with governmental innovation, and legal and institutional frameworks.^[6]

Step I (1987–1995) focused on improving the delivery and production of public services, increasing the convenience of people and strengthening the national competitiveness, as it targeted to realize a small but strong government through national informatization. To achieve these goals, the Korean government carried out the first and the second National Basic Computer Network Project (NBCN). The first NBCN (1987–1991) focused on building five basic national databases on administration, finance, education and research, national defense, and public security. During the second NBCN (1992–1995), eight databases on prioritized tasks—national welfare, post offices, procurement, freightage, fishing boats, land registration maps, intellectual property rights, and weather information—were developed.

During step II (1996–2000), the government targeted to realize a small and efficient government, while innovating public service to citizens and promoting the sharing and disclosure of government information within the framework of the first National Informatization Promotion Master Plan. The Korea Information Infrastructure (KII), which started in 1995 and completed in 2005, made it possible to provide

high-speed Internet service to all schools across the country from 1997. The KII also made the front-office services such as procurement, patent administration, customs, national taxation, and railroad services available to individuals and companies on-line. During this period, social actors such as universities and mass media trumpeted the social campaign "Let's lead the world in information society, although we have been late in industrial revolution," contributing to improve national awareness on informatization.

The vision of the Korean e-government in step III (2001–2007) is "to leap to global leading country in 21st century," by improving public services to citizens and businesses, enhancing transparency and productivity of the administration, and realizing people's participation. In addition, the Presidential E-Government Special Committee was established to derive ministerial coordination and cooperation for the cross-ministerial e-government projects.

Step III covers the last 2 yr of the Kim Dae-jung administration (2001–2002) and the whole period of the Roh Moo-hyun administration (2003–2007). Eleven e-government projects were implemented during the Kim administration, enabling cross-ministerial connection. During the Roh administration, 31 e-government projects were implemented and would be completed by the end of 2007, with the goals of innovating business practices, improving customer services and information resource management, and reforming laws and regulations. In short, the e-government of step III can be characterized with 1) the development of government-wide architecture to support cross-ministerial consolidation and linkage of information systems; 2) the establishment of participatory e-governance allowing people's participation in policy-making process; and 3) the complementation and improvement of the existing projects.^[7]

The evolution of Korean e-government reflects the typical development stages of e-government. The internal administration became more efficient and productive at the first stage. The second stage makes the public service user friendly. Finally, the third stage realizes the democratic governance that ensures interactive participation.

CONCLUSION

As information technologies advance and government processes change, e-government continuously evolves. The model of the Organization for Economic Cooperation and Development explains this evolution with three development levels: one-way information provision; two-way consultation; and active participation in policy making. United Nation's web presence

measurement model shows five development levels of e-government: from emerging to enhanced; to interactive; to transactional; and to networked presence. Both models measure the maturity level of e-government with technology advancement and the extension of interactive participation of people.^[8]

According to UN model, East Asian countries show outstanding achievement. Korea went up from 15th to fourth among 191 UN member countries, Japan from 27th to 14th, and China from 93rd to 57th between 2002 and 2005.^[9] Korea is already considered as the global top in e-government readiness, and Japan is almost catching up. Although currently positioned among middle-ranking countries, e-government in China is also rapidly developing.

East Asian countries share several characteristics as well. Firstly, the strong political will and leadership on e-government are critical to success. Both Korean and Chinese leaders pursued e-government as one of presidential agenda, considering it as a lever for economic development and administrative innovation. Secondly, e-government, accompanied by the government-wide business process reengineering beyond the simple automation of the existing agency-centric process, improves the transparency and efficiency of the administration. Chinese President Jiang Zemin stressed in 2002 the administrative reform through e-government, and Korea established the e-Government Special Committee under the Presidential Committee on Government Innovation to tightly link e-government with government innovation. Thirdly, e-government is expected to lead the development of participatory e-governance, as its objective evolves from the delivery of public service to the active participation of people in policy-making process. However, regarding the use of the Internet, Korea, Japan, and China hold different positions. Japan and Korea consider the Internet as a useful means to facilitate participatory governance as well as administrative transparency and accountability. On the other hand, China regards it negative, and even as a challenge to the regime.

Lastly, e-government agenda of these countries include the issue of bridging digital divide in society. Without overcoming discrepancy among social classes in the access to the Internet, e-government services cannot be delivered to the citizens.

REFERENCES

1. Ure, J.; Lovelock, P. E-government in China. In *China's Digital Dream: The Impact of Internet on Chinese Society*; Zhang, J., Woessler, M., Eds.; The University Press Bochum: Bochum, 2002; electronic form; 5–7.
2. Zhang, J. A critical review of the development of Chinese e-government. *Perspectives* **2002**, 3 (7); electronic form; 7.

3. UNDESA. *Compendium of Innovative E-government Practices*, II Ed.; UN: New York, USA, 2006; 81 pp.
4. Ure, J.; Lovelock, P. E-government in China. In *China's Digital Dream: The Impact of Internet on Chinese Society*; Zhang, J., Woesler, M., Eds.; The University Press Bochum: Bochum, 2002; electronic form; 1 pp.
5. Miyazaki, Y. *Japan's e-Government*, ICA Country Report; International Council for Information Technology in Government Administration, 40th Conference, Guadalajara, September, 2006, 1 pp.
6. Song, H.; Shin, S. Historical approach to e-government of Korea: lessons learned and challenges ahead. In *Government Reform: Past, Present, and Future*; Proceedings of the International Conference for the 50th Anniversary of KAPA, Seoul, Oct 16–17, 2006; The Korea Association for Public Administration: Seoul, 1996; 281 pp.
7. Korea, Presidential Committee on Government Innovation. *E-Government of 'the Participatory Government'*; Seoul, Korea, 2005; 50–53.
8. OECD. *Citizens as Partners: Information, Consultation and Public Participation in Policy-Making*; OECD: Paris, France, 2001; 23 pp.
9. UNDESA. *Global E-Government Readiness Report 2005*. UN: New York, USA, 2005; 42 pp.

BIBLIOGRAPHY

1. Japan, Ministry of Internal Affairs and Communications. *Integrated Window of E-Government*, [//www.e-gov.go.jp/doc/scheme.html](http://www.e-gov.go.jp/doc/scheme.html) (accessed January 2006).
2. Japan, IT Strategic Headquarters. *New IT Reform Strategy*, [//www.kantei.go.jp/foreign/policy/it/ITstrategy2006.pdf](http://www.kantei.go.jp/foreign/policy/it/ITstrategy2006.pdf) (accessed January 2006).
3. OECD. *The e-Government imperative: main findings. Policy Brief*; 2003.

Electric Industry Restructuring

John D. Wong

Stacey M. Brewer

*Hugo Wall School of Urban and Public Affairs, Wichita State University,
Wichita, Kansas, U.S.A.*

INTRODUCTION

Prior to the 20th century, electricity was primarily an unregulated industry in the U.S.A., largely owned by private entities. These privately owned utilities provided for the generation, transmission, and distribution of electricity.^[1] The early structure of the electric utility industry was predicated on the concept that a natural monopoly was an efficient, low-cost means of generating, transmitting, and distributing electricity. Because monopolies in the U.S.A. were outlawed, regulation of the utilities became the rule. Federal and State government regulations were developed to control operating procedures, prices, and entry into the industry to protect consumers from potential monopolistic abuses.^[2]

However, the electricity industry has undergone considerable changes in the last two decades. Major industry restructuring has separated the once vertically integrated electric utilities into distinct entities. To facilitate competition at the wholesale level, the Federal Energy Regulatory Commission (FERC) has required transmission-owning utilities to “unbundle” their transmission and power marketing functions and provide non-discriminatory, open access to their transmission systems by other utilities and independent power producers. At the retail level, some states have required utilities to divest their generation assets as part of restructuring. In addition, power marketers—who often do not own generation, transmission, or distribution facilities—buy and sell power on wholesale markets and market electricity directly to customers.^[3]

Several factors have caused this structure to shift to a more competitive marketplace. First, technological advances have altered the economics of power production. Second, rate increases caused primarily by increases in utility construction and fuel costs prompted many to question the efficiency of the existing regulatory environment. Third, the emergence of nonutility power producers demonstrated that traditional vertically integrated electric utilities were not the only source of reliable power.^[2]

THE FEDERAL POWER ACT OF 1935

Much of the present structure of federal electric utility regulation dates back to the Federal Power Act of

1935. The act granted the Federal Power Commission (FPC) authority to regulate wholesale power sales that were made by investor-owned utilities (IOUs). Before this legislation, interstate transactions could not be regulated. The FPC was given a broad authority in matters in which state regulation was found ineffective or unconstitutional. The FPC was authorized to promote interconnections between utilities under its jurisdiction or between a jurisdictional utility and other entities in wholesale electricity transactions.

When the Federal Power Act was established, the issue of mandatory or unrestricted wheeling was raised. At that time, the Congress chose to narrowly define the act. The act allowed the Federal Power Commission to order interconnections or energy sales only on the application of a state commission or utility and only where such action would not jeopardize service to existing customers or force the expansion of a utility's generation or transmission capacity. Mandatory wheeling, or the use of the transmission facilities of one party to transmit the power of another, per se, was prohibited, except during emergency situations, when the FPC could order temporary interconnections, sales, or transmission of energy.

After the establishment of the Public Utility Act of 1935, the electric utility industry stabilized. Electricity prices declined, demand expanded, capacity and reserves were adequate while the industry received a stable level of profits. Because of this stability, federal regulation was limited.

In 1962, the FPC reorganized the Bureau of Power to strengthen its regulation of wholesale electricity markets and to improve the efficiency of the power industry. The FPC's jurisdiction covered the wholesaling of electricity by any utility that is interconnected to utilities in other states. As a result, practically every wholesale electricity transaction came under the jurisdiction of the FPC and its successor, the Federal Energy Regulatory Commission, created in 1977.

THE PUBLIC UTILITY REGULATORY POLICIES ACT OF 1978

A turning point in the evolution of federal regulation of electric power was the Public Utility Regulatory

Policies Act of 1978 (PURPA). Prior to PURPA, the new wholesale power producer could not sell directly to retail customers because it did not have a state retail franchise, nor wholesale power could be delivered to wholesale power purchasers because local retail utility companies owned and thus controlled local connecting power transmission lines. After it came into force, PURPA amended the FPA and gave the FERC power to order transmission service when certain specific conditions are met. Under PURPA, mandatory wheeling may be ordered by FERC after: 1) a complaint has been filed by another utility, after; 2) notice and an opportunity for a hearing, and upon; 3) the determination of a number of factors including:

- That the request has been made of the utility that sells power for resale and it is either unwilling or unable to provide the transmission service.
- That the requested order conserves energy, promotes the efficient use of resources, or improves reliability.
- That the power to be transmitted will not replace power presently provided to the recipient by the transmitting utility.
- That the proposed transaction will reasonably preserve existing competitive relationships among the utilities involved.
- That the order will not result in any “reasonably ascertainable uncompensated economic loss” by any of the involved utilities.

Although PURPA provided explicit “wheeling authority,” severe limitations were placed on its use. It was incomplete because transmission was limited to local distribution and that distant transmission of wholesale power was open only through voluntary agreement. Federal Energy Regulatory Commission did not have wheeling authority to force a transmission access to wholesale distributors. The restrictions placed on such sales limited the full capabilities of the wholesale power generation market.

THE ENERGY POLICY ACT OF 1992

The rationale underlying wheeling is that by opening up transmission access to wholesale power distribution, power companies will be able to operate more efficiently. By purchasing wholesale power, the retail electric companies will be able to limit capital investments necessitated by having to build additional generating capacity to meet peak load demands. According to this argument, electricity prices will decrease because of increased efficiency, and stranded investment costs will be reduced. As such, Congress enacted the Energy Policy Act of 1992 (EPAct), which expanded

the Federal Energy Regulatory Commission’s authority to order vertically integrated utilities to allow non-utility power producers access to the transmission grid to sell power in an open market. Although the law specifically prohibits FERC from ordering retail wheeling and the transmission of power to a final customer, the law amended the wholesale transmission provisions of the Federal Power Act to allow a nationwide open-access electric power transmission grid for wholesale transactions. Independent power producers, publicly owned utilities, rural cooperatives, and industrial producers gained the ability to seek from FERC orders that require transmission-owning utilities to provide transmission service at FERC-defined “just and reasonable” rates. The language of the law concerning pricing directs FERC, when it issues a transmission order, to approve rates which permit the utility to recover “all legitimate, verifiable economic costs incurred in connection with the transmission services.” Such costs include “an appropriate share, if any, [of] necessary associated services, including, but not limited to, an appropriate share of any enlargement of transmission facilities.” The language also says that FERC “shall ensure, to the extent practicable,” that costs incurred by the wheeling utility are recovered from the transmission customer rather than “from a transmitting utility’s existing wholesale, retail, and transmission customers.”^[2]

However, FERC’s authority to order access was implemented on a case-by-case basis and proved to be slow and cumbersome. To remedy that, in 1996, FERC issued Order 888 requiring all vertically integrated investor-owned utilities to file an open-access transmission tariff that would provide universal access to the transmission grid to all qualified users.^[2] The order requires all public utilities that own, control, or operate facilities used for transmitting electric energy in interstate commerce to:

- File open-access nondiscriminatory transmission tariffs containing minimum terms and conditions.
- Take transmission service (including ancillary services) for their own new wholesale sales and purchases of electricity under open-access tariffs.
- Develop and maintain a same-time information system that will give existing and potential users the same access to transmission information that the public utility enjoys.
- Separate the transmission from generating and marketing functions and communications.^[4]

Order 888 was an important stimulus in the development and strengthening of competitive wholesale power markets, but discriminatory practices regarding access to the transmission grid still remained, and a more effective effort was needed. In December 1999,

FERC issued Order 2000 calling for the creation of regional transmission organizations (RTOs), independent entities that will control and operate the transmission grid free of any discriminatory practices. Electric utilities are required to submit proposals to form RTOs from October 2000 to January 2001.^[2]

Another equally important component of Order 888 was the requirement for transmission owners to functionally unbundle their activities. Functional unbundling required the transmission owner to take transmission service under the same tariff as other transmission users, to separate rates for wholesale generation, transmission, and ancillary services, and to rely on the same electronic information network that its transmission customers rely on to obtain information about prices and available capacity of the transmission system.^[2]

The second goal of Order 888 was to ensure that electric utilities are able to recover their sunk costs in a competitive industry. These sunk costs are called stranded costs, or transition costs, and they represent a utility's capital investments that are unrecoverable because of the transition to competition. The rationale for allowing stranded cost recovery is that utilities have invested billions of dollars in facilities under a regulatory regime that allowed cost recovery of all prudent investments. To gain support and cooperation for a successful transition to a competitive industry, and to be consistent with the past decisions, FERC believed

it was critical that utilities recover these costs. At the same time, FERC recognized that recovery of stranded costs may delay some of the benefits of competitive power markets.^[2]

Order 888 of FERC spelled out under what general conditions a utility is entitled to recover its stranded costs and from whom. Order 888 specified that cost recovery at the wholesale level is limited to situations where there is a link between the use of FERC's required open-access transmission tariff and the loss of wholesale power customers. Then FERC went further to specify that recovery of wholesale stranded costs should be assigned to the departing customer. At the retail level, FERC determined that states should have primary jurisdiction over cost recovery resulting from retail competition, although it would entertain requests to recover costs resulting from retail competition when a state does not have the authority.^[2]

Whether the EPAct ultimately results in an efficient competitive wholesale electric power market depends in large part on how FERC and state regulators interpret and implement its provisions. Although one of the objectives of the EPAct is to facilitate equal access to transmission services for all wholesale generators at equitable rates, the economic leverage of transmitting utilities remains. The EPAct also amends PURPA to require states to decide whether it is appropriate and consistent with state law to conduct "general evaluations" of major issues before allowing or mandating

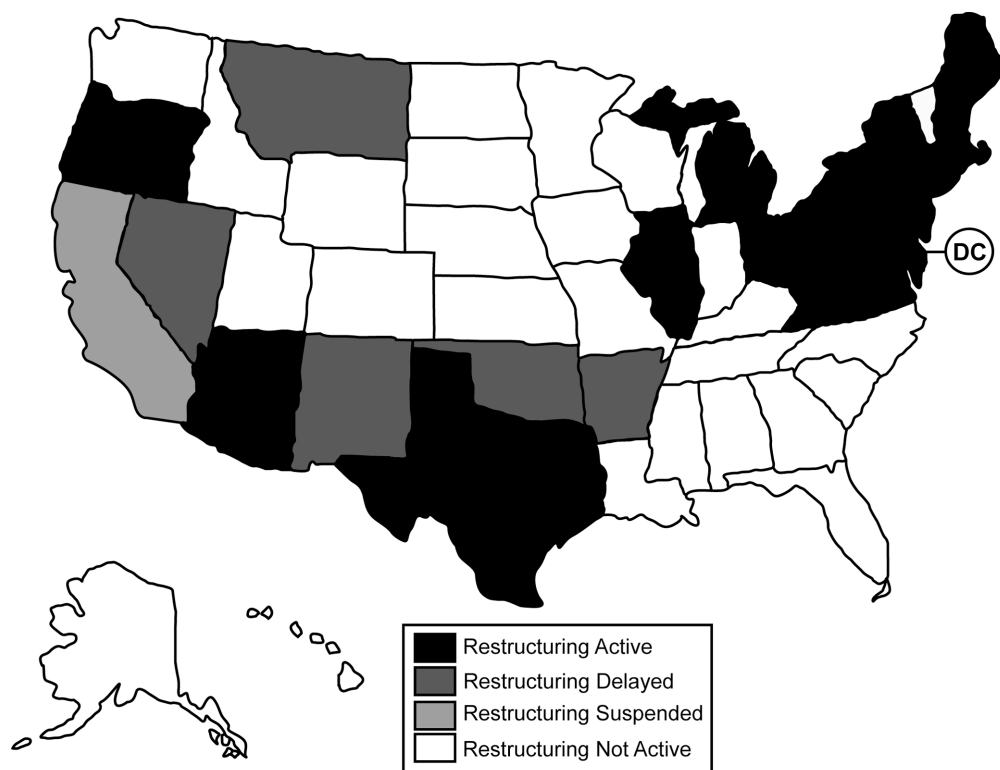


Fig. 1 The status of state electric restructuring as of February 2003.

utilities to enter into long-term wholesale power agreements. Notwithstanding, although the Energy Policy Act does prohibit FERC from ordering retail wheeling, it is under consideration in many states. In Fig. 1 the status of state electric restructuring as of February 2003 is shown.

ELECTRIC UTILITY OPERATION IN RETAIL WHEELING ENVIRONMENT

In a retail wheeling environment, electric utilities can serve customers at any location and utility customers are free to purchase their electric energy from any electric energy supplier, which may not even be an electric utility company in the traditional sense. There will be interchanges of electric energy through the grid just as before; however, now instead of the interchanges between different electric utilities there will be interchanges between different electric energy suppliers and their customers, thus producing the term retail wheeling. This makes the operation of the grid much more difficult.

The wheeling transaction is a matter of precise coordination among the seller, the wheeler, and the purchaser. The seller overgenerates electricity in the amount the purchasing utility has agreed to buy. The purchasing utility then undergenerates the same amount at the same time as the seller. The wheeler allows its transmission lines to be used. This process allows a specific amount of power to flow from the seller to the purchaser over the lines of the wheeler during a specified period.

The transaction involves only the transmission facilities of the wheeler. The generation facilities of the wheeler are not involved in any way. The main responsibility of the wheeler is to make sure that its lines are able to carry the exchanged amount of power without becoming overloaded or overheated, causing damage to its systems or power outages.

In the traditional operating environment, individual electric utilities had total control of their transmission and distribution network and only had to be concerned about the adequacy and reliability of their grid to serve their own customers. However, in a retail wheeling environment many different electric power suppliers, from both inside and outside the state, will be trying to transmit electric energy through the grid to their specific customers located anywhere in the state. In addition, some electric power suppliers in the state will be sending power through the grid to new customers outside the state and there will probably be some power flowing through the state from power suppliers outside the state to customers outside the state. The grid will change from a private conduit with access controlled and monitored by each utility in their own service territory, to a public thoroughfare (or common carrier) where theoretically, at least, any electric energy supplier

has the right to use the grid and inject electric energy for extraction by the particular customers that have decided to purchase electric energy from this supplier. Thus, in a retail wheeling environment the number of transactions between energy suppliers and energy consumers will increase, because any supplier can sell to any customer and any customer can buy from any supplier. There will also no longer be any geographic restrictions placed on customers and suppliers.

As all electric transmission and distribution lines carry a designed maximum amount of power, the issue of who will operate the grid in a retail wheeling environment is an important technical issue. In a retail wheeling environment, the grid operator will serve as a gatekeeper to ensure that as many transactions between electric energy suppliers and their customers as are desired can take place, without in some way overloading a particular line or portion of the grid and thus endangering the reliability of the electric grid system in general. When overloaded, transmission and distribution lines can fail because of heating. When one line fails, the load is taken up by the other lines in the grid. However, if some remaining lines are already near their maximum capacity, they too can become overloaded and fail if they are forced to carry the additional power from a failed line. This can lead to a cascading failure that, if unchecked, could destroy the entire grid system.

A second key technical issue associated with retail wheeling relates to the first issue and involves the development of some criteria by which individual energy suppliers will be allowed access to the grid in the likely event that portions of the grid are operating at maximum capacity and no additional energy transactions can take place without jeopardizing the operation of the overall grid system.

A third technical issue involves the fact that it will be difficult to plan and design new lines in the future because it will be difficult to predict the number of electric energy suppliers and customers that will want to inject and/or extract electric energy into or from this future line. There is also the related question of who pays for the new lines.

Many traditional electric utilities are not likely to survive in a fully evolved retail wheeling environment. While, traditionally, electric utilities have been involved in all three aspects of the production of electric power—generation, transmission, and distribution—electric utilities in a competitive environment are likely to focus operations on a single aspect of the production process. Utilities that have an abundance of new, efficient generating plants are likely to divest themselves of their transmission and distribution systems, and their traditional geographical customer base, and devote themselves totally to the generation of power (generation companies or “gencos”) for the wholesale market.

Similarly, utilities with efficient transmission systems, particularly those covering a wide geographic area, are likely to focus on the transmission of power (transmission companies or “transcos”) from wholesalers to retail distributors. These entities would operate parts of the transmission system within a discrete region as either a common carrier or an open-access contract carrier. A common carrier provides transmission service on a non-discriminatory basis to all those seeking its services within the limits of its capacity. This carrier would not actually own the electricity flowing through its lines, but would simply provide the transmission service for a fee. An open-access contract carrier would commit its transmission capacity, possibly on a long-term basis, by means of contractual agreements. Alternatively, the contract carrier could buy power at one location and sell power at another, thereby acting as a kind of broker for delivered electricity. The transmission company would serve as a bridge between producers and their customers. These transmission companies would operate the grid and ensure its reliability by limiting access and restricting the amount of energy injected into the grid if it was determined that portions of the grid were operating at maximum or near maximum capacity. These companies would also plan, design, and build new power lines as the grid requirements increased. However, because the transmission company has the potential to exercise substantial monopoly power, regulation of access and rates might be necessary to avoid unnecessary duplication of facilities, and to prevent chaos in interconnections.

Distribution companies or “discos” would result where utility companies have decided to focus their business activity on the sale and distribution of power to retail customers. Under this arrangement, consumers would receive two bills. One of these bills would be from the local distribution company and would be a fixed amount each month for access. This would represent the charge for having the lines and electric meter connected to the customer’s site. The second bill would be from the particular electric energy supplier the customer had selected as the source of his/her electric energy. Presumably, this bill would include not only the cost of the energy but also any charges added by the transmission company.

CONCLUSIONS

In a competitive environment, the traditional regulatory compact of an exclusive service territory in exchange for non-discriminatory service to all customers upon demand must yield to a set of transactions that are subject to change, based on price, quality, service, attractiveness, or any combination of reasons. The telecommunications, transportation, and natural

gas industries have undergone deregulation. Customers now have many choices, as do the suppliers who can tailor their strategies to meet their desired market. Competitive electricity markets are clearly in the early stages of development. While restructuring efforts have broadened electricity markets and increased the number of market participants, the extent to which expected benefits have been achieved is uncertain.^[5] The aftermath of the California electricity crisis in 2000^[1,6–9] and the August 14, 2003, electricity blackout in eight northeastern and midwestern U.S. states^[10,11] underpin ongoing uncertainties concerning the efficacy of the restructuring of the electric power industry. However, it is likely that the economic deregulation or restructuring of the electric industry will continue.

REFERENCES

1. Tomain, J.P. The past and future of electric regulation. *Environ. Law* **2002**, 32 (2), 435–474.
2. *The Changing Structure of the Electric Power Industry 2000: An Update*; U.S. Department of Energy, Energy Information Administration, Office of Coal, Nuclear, Electric and Alternate Fuels: Washington, DC, 2000.
3. America’s energy infrastructure: a comprehensive delivery system. In *Reliable, Affordable, and Environmentally Sound Energy for America’s Future, Report of the National Energy Policy Development Group*; U.S. Department of Energy: Washington, DC, 2001; 7–1–7–18.
4. *The Restructuring of the Electric Power Industry: A Capsule of Issues and Events*; U.S. Department of Energy, Energy Information Administration, National Energy Information Center: Washington, DC, 2000.
5. *Lessons Learned from Electricity Restructuring: Transition to Competitive Markets Underway but Full Benefits Will Take Time and Effort to Achieve*; GAO-03-271; U.S. General Accounting Office: Washington, DC, 2002.
6. Fox-Penner, P.; Basheda, G. A short honeymoon for utility deregulation. *Issues Sci. Technol.* **2001**, 17(3), 51–56.
7. Ferrey, S. The eagles of deregulation: the role of the courts in a restructured environment. *Environ. Law* **2002**, 32 (2), 297–329.
8. *Restructured Electricity Markets: California Market Design Enabled Exercise of Market Power*; GAO-02-828; U.S. General Accounting Office: Washington, DC, 2002.
9. Hirst, E. *The California Electricity Crisis: Lessons Learned for Other States*; Edison Electric Institute: Washington, DC, 2001.
10. Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendation; U.S.-Canada Power System Outage Task Force: Washington, DC, 2004.
11. *Electricity Restructuring: 2003 Blackout Identifies Crisis and Opportunity for the Electricity Sector*; GAO-04-204; U.S. General Accounting Office: Washington, DC, 2003.

Electronic Government

Mary Maureen Brown

*Department of Political Science, University of North Carolina at Charlotte,
Charlotte, North Carolina, U.S.A.*

INTRODUCTION

Electronic government (e-gov) refers to government's use of technology, particularly web-based Internet applications, to enhance the access to, and delivery of, government services to citizens, business partners, employees, and other government entities. Electronic government activities focus on three forms of interactions: Government to Citizen, Government to Business, and Government to Government. Within each of these interaction domains, four types of activities typically take place. The simplest type of activity is to provide information over the Internet. Under this scenario, public agencies employ the Internet to distribute information such as school closing notices, public hearing schedules, issue briefs, or regulatory notifications. The second type of activity allows two-way communications between the agency and the citizen, a business, or another government agency. These web pages allow users to engage in dialog with agencies—users can post queries, comments, or requests to the agency, regardless of time of day or day of week. Gaining in complexity, the third form of activity that takes place over the Internet involves conducting transactions. Some governments have begun to port transaction-based services to the web, allowing users the opportunity to access agency services 24 hours a day, 7 days a week. Some of the more common transaction-based activities include:

Procurement	Interagency payments	Employee leave requests
Joint tax filing	Vehicle registration	Worker's compensation
File business taxes	Apply for licenses	Birth/death certificates
Marriage licenses	Divorce decrees	Criminal background checks
Request license plates	Register to vote	School enrollment
Pet licensing	Change of address forms	Electronic fund transfers
Travel expense reimbursement vouchers		

The final form of activity that takes place over the web centers on governance. Online polling, voting, and campaigning are just some of the ways governments are attempting to increase citizen involvement in the political process.

Despite the noted benefits, the implementation and maintenance of online service delivery efforts can encounter formidable challenges. The deployment of web-based services requires an extraordinarily complex combination of technical, organizational, economic, human, and political factors. Some of the areas that dominate discussions pertaining to digital government include better methods of information technology management, models for electronic public service transactions and delivery systems, and avenues for citizen participation in democratic processes. Other topics such as secure and controlled access to timely, accurate, and authoritative public records are central to any web-based initiative. Data quality, willingness to share data across organization boundaries, and threats to privacy and security all serve to stymie digital government benefits.

TECHNOLOGICAL COMPONENTS

Governments connect to the Internet by establishing services with an Internet service provider (ISP). The citizen then accesses the government web site via their ISP subscription. Citizens gain access to all the online services a government might provide on these servers via a single entry screen known as a "portal." A portal is a "single face" where citizens can conduct "one-stop shopping" for their government needs. Portals allow organizations to aggregate and share content—information, services, and applications with customers, partners, employees, and suppliers. Portals are often organized by topic area rather than bureaucratic function and often incorporate various levels of government. For example, in the "my vehicle" section, citizens may find all information pertaining to vehicles—from driver's licensing, to vehicle registration, to traffic citations. In designing a portal, attention should be given to content, presentation, density, and color schemes. Above all, web sites should be easy to navigate and users

should not have to follow more than three links to find the information for which they are looking.

Software Applications

As mentioned earlier, governments employ online services in four ways: 1) to disseminate information; 2) to promote communication; 3) to conduct transactions; and 4) to facilitate governance. A standard design includes a firewall server, a web server, a transaction server, and a database server. Depending on how they choose to employ online services, the following applications may be needed to establish online government services.

1. Electronic payment systems allow the government agency to collect fees and levy taxes for services such as business licensing, vehicle registration, or refuse collection. Electronic payment systems differ in their approach to how they collect fees. Some software systems allow credit card transactions, some electronic checking, and others provide a digital cash approach. Initiatives are also underway to allow tax payers to pay for services via automated teller machine accounts. When deciding which payment scheme (or schemes as the case might be) to employ, careful consideration must be given to confidentiality, integrity, and authentication issues.
2. Commerce server software provides numerous features pertaining to product information, shopping carts, shopper information, order initialization, shipping charges, taxes, payments, and receipts. Commerce server software allows citizens to easily conduct one or more transactions in one visit. This software tracks the transactions and provides a final summary and checkout process for the citizen.
3. Web server software allows citizens to request information and serves up the pages as requested. The server software controls the storage and retrieval of documents and their transfer; it is also responsible for encryption and authentication. Moreover, the software will have incorporated communication programs that allow it to communicate over the Internet. The software packages range in price according to functionality. The software should be simple to configure, as well as easy to administer and manage on a daily basis. Performance is critically important. The software must be capable of receiving hundreds of thousands of transactions per day.
4. Web browser software allows citizens to connect with servers to access any number of web pages, and to follow links from document to document or page to page. Browser software acts as a link between the user and the Internet or online service. The browser will contact the online service and send a request for information or for a web page. The browser then receives the information, usually in the form of a page, and displays it on the user's computer. Browsers often include plug-ins, which are software programs that are needed to display images, hear sounds, or run animation sequences.
5. Transaction server software is often referred to as online transaction processing (OLTP) software. Its basic duty is to process data according to the rules established by an organization. It passes requests from the web server on to the database server. The transaction server confirms orders, sends a request for payment, and captures payment information. It also takes responsibility for adding, deleting, and modifying orders. The transaction server runs the programs that tell the database server how to handle the data. Because OLTP systems are designed for high volume, the underlying data will change on a minute-by-minute basis. Online transaction processing software captures the day-to-day e-government operations by capturing data about basic business processes and events. Many users concurrently access the database to insert, retrieve, update, and delete individual records. Data are stored in operational databases to meet these transactions. This software is designed for a high volume of small transactions. The focus of these systems is on performing reliably under heavy transaction loads. These systems support day-to-day operations, and are designed to be performance sensitive and to work well under repetitive routine transactions. Because they are not designed for analytical purposes, they do not generate decision-based information well.
6. Firewall server software controls the flow of traffic between two or more networks. The firewall is usually placed between the government network and an external network such as the Internet or a partner's network. Firewalls monitor access to and from specific networks. These systems can protect against viruses, intrusions, and other unauthorized system access.

Hardware Solutions

Internet-based solutions are built on the client/server concept. The client/server concept incorporates a number of computer devices that "serve" or "host"

or “operate” an installed software application. For the most part, each server will “host” only one application. The hosting of only one application per server allows activities on the other servers to continue while one is taken out of service for maintenance or repair. Hence, loss of services is kept to a minimum with the client/server model. In short, depending on the type and number of online services the agency will be providing and the number of citizens that will be accessing the services, a separate computer server will probably be required for each of the fore-mentioned software applications.

For the most part, computer servers only differ in three major ways. They differ in their processing speed, their storage capacities, and the speed of their data input–output transporting mechanisms. In determining the hardware requirements for designing the servers, the following issues should be carefully considered:

- *Performance*: The ability to perform numerous operations or to handle a large workload without requiring extensive modifications.
- *High reliability*: The ability to run without a non-recoverable failure for long periods of time.
- *Service/support*: Maintainability and ease-of-use considerations, such as the ability to remove and replace a component (e.g., fan, disk drive, power supply) without having to shut down the system or cause processing functions to stop.
- *Storage*: The storage capacity, both in terms of raw space (i.e., how many gigabytes) and the number of disk drives a server can support inside the chassis.
- *Redundancy*: Whether a given component (e.g., fan, power supply, network interface card, disk drive) has a backup and a fail-over mechanism; redundancy is a method for improving system reliability and availability.
- *Interoperability*: The ability for the different hardware components to interconnect and communicate seamlessly. Software and hardware incompatibilities continue to create difficulties in deploying technology initiatives.

INTERNET ACCESS AND INTERNET SERVICE PROVIDERS

Much as freeways are networks of roads designed to transport people from one location to the another, networks provide the opportunity for data to travel from one location to the next. The Internet is often defined as a highly connected virtual highway that allows data to pass from person to person, organization to organization, and country to country. Data are able to make these cross-country or international trips because the Internet has standardized on a common protocol of

data definitions. To communicate effectively, everyone that transfers data via the Internet adheres to the TCP/IP (transfer communication protocol/Internet protocol) protocols.

Governments, businesses, and citizens gain access to the Internet by subscribing with a locally affiliated ISP. The ISP charges a flat monthly rate for their connection, which usually includes their fee and the cost of a dedicated line from your location to their nearest facility or point-of-presence for connection to the Internet. The dedicated line is typically provided by the local telephone company, and most ISP's make these arrangements for their customers.

Today's backbone consists of high-speed fiber-optic lines. The number of “hops” the data make as they are routed from the government to the backbone will affect the speed at which data are delivered to the citizen. The more computers in between the ISP and the backbone, the slower the connection. When contemplating establishing a subscription with an ISP, several issues should be considered. The ISP should also incorporate numerous security mechanisms to protect the data from unauthorized access and destruction. The ISP should also provide written guarantees on levels of service and technical support. Also, inquire about financial and technical background, and ask about expansion capabilities.

Reliability and speed are two critical issues to consider when deciding on an ISP service agreement. An ISP should provide performance guarantees regarding their uptime or availability. In essence, the ISP should be able to guarantee that their services will be available 99.9% of the time.

Bandwidth refers to the speed with which content can be delivered to the end user. Internet service providers typically provide a range of choices on how the government might connect to the Internet, and the choices will vary according to speed, user load, and cost. Some users connect to the Internet over ISDN telephone lines with connection speeds of 128 Kbps. Telephone companies are offering digital subscriber line connections, which deliver at speeds approximating 1.5 Mbps. Likewise, some cable companies offer high-speed cable access in the 10 Mbps range. Government agencies often link to ISPs with dedicated telephone connections such as T3 lines (deliver at speeds of 45 Mbps), T1 lines (1.5 Mbps), or fractional T1 services (256 Kbps to 1.5 Mbps). The decision as to how much bandwidth is required is based on the number of users that will access the system and the types of services the government wants to deliver.

It is important to keep in mind that many of the citizens will be connected to their ISP by way of dial-up telephone lines. Most consumers connect to the Internet over telephone modems with speed ranges from 14.4 Kbps to 256 Kbps. Because these connections

are slow, web design efforts should account for users employing slower dial-up connections.

Although much attention has been dedicated to Internet activity, government agencies also deploy services using an intranet or extranet approach. In theory, the only difference between the Internet and an intranet or extranet is how the data are routed. With an Internet, the data travel the “open roads” via the high-speed backbone. With intranets and extranets, data are more confined and do not access the backbone. With intranets, data remain within the organization’s network. With extranets, data leave the internal network and travel to outside organizations via dedicated lines. It would not be uncommon for a single government agency to maintain an intranet, an extranet, and also a public Internet.

In summary, as decisions are made regarding the various technical alternatives that a government agency might employ in developing their online service, overarching concerns should always center around system performance issues pertaining to:

- *Scalability*: The system should be capable of handling increasing numbers of users without any disruption to service.
- *Flexibility*: The system must provide a broadly configurable array of hardware and software devices that do not require major reinstallations as enterprise requirements change.
- *Compatibility*: The system must meet expandable configuration requirements, as well as standard industry specifications, to protect future application investment.
- *Manageability*: The system should not demand excessive management time and effort for maintaining online operations.
- *Availability*: The system must be capable of sustaining tens to hundreds of thousands of processing transactions with minimal wait time or downtime.
- *Security*: The system must incorporate an appropriate level of security to prevent unauthorized access to data.
- *Total cost of ownership*: An amalgam of the fully configured system price, including installation, setup, training, infrastructure costs, downtime costs, and ongoing maintenance.

THE POLICY ARENA

Much attention must be given to developing the underlying policy framework that will support online services. Most policy discussions center on issues pertaining to security, privacy, and ethics. However, attention must also focus on substantive, or field/profession-specific, policy issues. As organizations migrate from the first

three channels to the fourth channel, substantive policies related to the actual operation that is being automated may require modification. Nonetheless, this section focuses on policies related to security, privacy, and ethics.

Security

Public agencies are particularly vulnerable to three types of security threats. The illicit modification, destruction, or exposure of data can create serious consequences—jeopardizing not only privacy rights, but inappropriate data access can threaten the safety of citizens. Overriding concerns regarding the moral, legal, and ethical ramifications of security breaches have forced many public managers to take a more risk-averse route, imposing serious limitations on computer information-sharing capabilities.

Unfortunately, as mentioned, the sharing of departmental data can result in the deletion, modification, circulation, and theft of important organizational data. Stallings defined four broad categories of data attacks. Interruption occurs when assets of the system (i.e., certain data elements) are destroyed or become unavailable for use, and interception occurs when an unauthorized party gains access to an information asset.

Conversely, modification relates to tampering with the data, and fabrication occurs when an unauthorized party inserts counterfeit objects or data into the system. According to Icové, Seger, and VonStorch, the motives for these attacks are based on entertainment value, or financial or personal gain.^[1] The results of these penetrations can be serious, and can result in the corruption or disclosure of information, theft of data, or denial of system services. All these threats expose the organization to severe repercussions. Legal liability, as well as the erosion of public confidence that results from an agency’s inability to ensure data security, has widespread societal implications.

Public key cryptography is touted as a potential solution to the security problem. This involves securing data with two keys—one is used to encrypt (scramble) the message, and the other is used to decrypt (unscramble) the message. A key is used to encrypt the message prior to its being sent, and a corresponding key is used to decrypt the message once it has received. In its simplest form, each person in the transaction has a pair of keys—one private and one public. The public key is made available for public access (e.g., to friends, colleagues, etc., or to anyone who might want to send the individual a confidential message). Employing encryption application software such as “pretty good privacy” software, the sender will encrypt the message with the designated receiver’s public key. The only way the message can be decrypted is via the receiver’s

private key, which the receiver keeps confidential. Only the receiver has the private key that is capable of decrypting the message. Thus, files can be encrypted to allow only certain individuals' access to the message content. Interestingly, it is virtually impossible to calculate one key based on knowledge of the other.

Privacy

Privacy is the right to be left alone and the right to be free of unreasonable personal intrusions. Information privacy is the claim of individuals, groups, or institutions to determine for themselves when, and to what extent, information about them is communicated to others. This issue is becoming the most important issue for consumers. Unfortunately, compliance with the Privacy Act of 1974 and its extensions are not simple because the line between legal definitions in ethics is not always clear. Protecting intellectual property on the web is also quite difficult because it is easy and inexpensive to copy and disseminate digitized information. The Organization for Economic Cooperation and Development^[2] recommended the following guidelines:

- Personal data should be obtained by lawful and fair means, and with the knowledge or consent of the data subject.
- Personal data collection should be relevant to the purposes for which that data are to be used. To the extent necessary for those purposes, such data should be accurate, complete, and current.
- The purpose of the data collection effort should be specified, and the subsequent use of such data should be only for the stated purpose as agreed to by the individual.
- Personal data should not be disclosed or otherwise used for any purposes other than that which was set forth in the purpose specification, except with the consent of the data subject by the authority of law.
- Personal data should be protected with reasonable security measures to prevent unauthorized access, destruction, modification, or disclosure.
- The data controller should make available information concerning the existence, use, and nature of all personal data that are collected.
- Individuals should have the right to ascertain from a data controller whether the data controller has personal data relating to that individual and should have access to correct the data, as required.

Ethics

Because the increasing use of information technology has created new ethical challenges, many of today's organizations are developing their own code of ethics

as a guide for their members. Mason, Mason, and Culnan categorized ethical issues into four domains: privacy, accuracy, property, and accessibility.^[3] Privacy relates to the collection, storage, and dissemination of information about individuals.

Data collection principles should focus on collecting only those data that are required to accomplish a legitimate business objective. Collection efforts should not be excessive and should be targeted only toward what is required. All individuals should be allowed to give their consent before data are gathered. However, such consent may be implied from the individual's actions, for example, applications for credit, insurance, or employment.

Accuracy relates to the information's authenticity as it is collected and processed. Steps should be taken to verify the accuracy of the information as it is acquired, processed, stored, and disseminated. Furthermore, the data should be made available to the individual to ensure that they are correct. If there is disagreement about the accuracy of the data, the individual's version should be noted and included with any disclosure of the file. Furthermore, in disseminating information, agencies should provide some indicator as to the accuracy of the information as misleading information can be worse than no information at all.

Property relates to ownership and the issue of intellectual property rights. Organizations need policies concerning who owns what data and how the data are to be employed. To date, many of the issues associated with ownership of electronic data remain under debate. As courts work through the intricacies of ownership, organizations can limit their liabilities by instituting policies pertaining to data ownership rights.

Accessibility relates to the right to access information. All access rights should be based on a legitimate "need-to-know" basis. Public organizations should maintain adequate security practices and procedures.

MANAGEMENT CHALLENGES

Although most public officials appreciate the benefits that e-government solutions can offer, they question their agencies' ability to overcome the hurdles associated with adoption and implementation. A study of state CIOs conducted by NASIRE identified that while funding is always difficult, cultural, organizational, and leadership issues are some of the most significant barriers to adoption. Specifically, the survey revealed that the greatest barriers center on:

- Providing a culture of teamwork, innovation, and change
- Education of executives and managers

- Collaboration and integration efforts
- Skills and expertise
- Diversity and complexity of existing operational processes

These noted concerns demonstrate an overwhelming need to establish a learning environment for coaching, nurturing, and facilitating both the adoption and maintenance of digitally based government services. Specifically, public managers need to be sure to encourage staff to:

- Develop skill sets to understand how to leverage Internet tools.
- Develop mechanisms to ensure responsibility and accountability for planning initiatives.
- Assess impacts as they relate to government outputs and outcomes.
- Understand how program specifications and requirements drive e-government developments.
- Assess impact on *all* citizens.
- Develop and monitor web-based balanced scorecards.
- Develop policies in areas such as security, privacy, electronic payments, standards, investments, digital archiving, encryption, and authentication.

REFERENCES

1. Icové, D.; Seger, K.; VonStorch, W. *Computer Crime: A Crimefighter's Handbook*; O'Reilly & Associates: Sebastopol, CA, 1995.
2. Organization for Economic Cooperation and Development. *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*; <http://www.oecd.org/dsti/sti/it/secur/prod/PRIV-EN.HTM>.
3. Mason, R.O.; Mason, F.M.; Culnan, M.J. *Ethics of Information Management*; Sage Publications: Thousand Oaks, CA, 1995.

Employee Benefits

Jerrell D. Cogburn

*Department of Public Administration, University of Texas at San Antonio,
San Antonio, Texas, U.S.A.*

Dennis M. Daley

*Political Science and Public Administration, North Carolina State University,
Raleigh, North Carolina, U.S.A.*

INTRODUCTION

Together with pay (or “direct compensation”) and work environment, employee benefits (or “indirect compensation”) represent a major component in public employers’ total compensation systems.^[1] As a component of total compensation, employee benefits contribute to public employers’ attainment of important human resource goals; namely, attracting, retaining, and motivating qualified employees. As this suggests, competitive employee benefits programs are integral to public organizations’ human resources goals and, ultimately, to organizational performance.

Employee benefits come in two forms—mandated and discretionary. Mandatory benefits, which employers are legally required to provide, include Social Security coverage, unemployment compensation, worker’s compensation, and family and medical leave. These mandated benefits along with several common discretionary benefits [health care, life and disability insurance, retirement pensions, and paid time-off (PTO)] represent what employers have offered traditionally as employee benefits. In recent years, however, traditional benefits packages have fallen short of meeting the needs and expectations of an increasingly diverse workforce. In order to meet these evolving employee needs and expectations and to remain competitive in today’s labor market, public employers have begun to offer a wider array of discretionary or optional benefits. Many of these benefits can be subsumed under the “work–life balance,” “family friendly,” or “worker friendly” heading because they seek to restore a sense of balance between employees’ work, family, and personal time.^[2,3]

market. The traditional benefits discussed below are offered by almost all public employers.

Health Care Benefits

Health care benefits are among the most commonly offered—and most expensive—of all employee benefits. Medical costs for hospital care can run into thousands of dollars in a matter of a few days even for minor illnesses. If prolonged treatment and intensive care are required then the term “catastrophic” refers not only to an illness’s life-threatening nature but also to its financial costs. Naturally, the prospects of paying these costs can be anxiety-producing for employees. Health insurance is the means by which these fears can be, to some extent, allayed. In addition to major medical expenses, health insurance can also aid in alleviating other health-related threats to motivation and productivity. For example, health insurance plans may include additional provisions for prescription of drug, mental health, dental, and eye care benefits. What is included and the extent of that coverage varies substantially from plan to plan with many having multiple levels of coverage under discretionary or optional provisions.

Deductible (e.g., \$300 per individual annually with a \$750 family cap) and co-payment (e.g., 20% of charges) cost containment provisions also alter the real value and costs of these benefits.^[4] These plans may also limit an individual’s total financial exposure through a total out-of-pocket provision (e.g., \$2500 per individual with a \$5000 family cap). It is because of this extensive variability among plans that it is extremely difficult to do cost comparisons on health benefits.

The basic health care covered under insurance plans is likely to be separated into segments requiring different levels of co-payments. Preventive care, such as is found in an annual physical examination and periodic eye and dental checkups, is often fully reimbursed and exempt from any deductible provisions. Relatively common, minor medical procedures may be reimbursed at a 90% level. More serious or long-term

TRADITIONAL EMPLOYEE BENEFITS

By offering traditional employee benefits, employers attempt to provide a measure of security and safety for employees and to meet a minimum threshold of benefits expected by potential employees in the labor

(but not catastrophic and life-threatening in nature) illnesses may require a 50/50 match. Catastrophic care (e.g., cancer and heart disease) whether as a part of the general policy or as an additional or optional benefit again provides something in the order of 80–90% reimbursement. Since the cost of catastrophic care quickly escalates into hundreds of thousands of dollars, even at these reimbursement levels the co-payment requirements are substantial. Reimbursement is also based on the customary and usual rates; in some instances employees are responsible for amounts in excess of these.

Employer-provided health plans usually provide options for spouse and family coverage (paid in full or part by the organization or entirely at the employee's own expense). Concerns about the health of family members can adversely affect an employee's productivity. Hence, the extension of health benefits to family members is essential.

Sick Leave and Disability Insurance

While health insurance covers the costs of obtaining medical care, it does not itself address the loss of income that can also occur due to illness. As a mandated benefit, worker's compensation legally covers employees for job-related accidents. But there are also discretionary benefits that provide income protection. Sick leave serves to continue an employee's income for short periods of illness. Disability insurance provides financially for long-term job loss.

Sick leave, which normally can be accumulated at a rate of 4–8 hr/mo, provides an employee with pay during a short-term illness. This encourages employees to take care of themselves when necessary, instead of attempting to “gut it out” only to lapse into a more long-term illness. It also provides protection to other employees by removing potentially infectious individuals from the workplace. In addition, sick leave can be used for medical appointments and caring for ill family members.

While part of the mandated Social Security program provides long-term disability coverage, the amounts paid out may not be enough to fully or adequately replace lost wages and salaries. Disability insurance for replacing the lost income (enabling one to continue paying for the ongoing expenses that level of income was financing) is often provided. Disability insurance can be either short-term or long-term. Short-term disability policies can often provide 100% of pay replacement for up to a month and replace 50% of pay over the subsequent 6 mo. Long-term disability (often integrated with, i.e., reduced by, Social Security) can replace two-thirds of pay until the disabled employee reaches 65 yr. In case of permanent disability, long-

term medical care insurance for home health care and nursing homes may be necessary (albeit this is quite expensive and is seldom provided as an organization-paid benefit).

Paid Time-Off

In addition to sick leave (which is also a form of PTO), most public employers offer employees additional PTO from work. Normally, this comes in the form of paid holidays and annual leave. Holiday leave normally coincides with observance of national holidays and amounts to about 10–12 days/yr. Annual leave or vacation time is taken at the discretion of the employee in consultation with the employer. Amounts of annual leave vary (e.g., 1–4 weeks/yr, depending on seniority). Conceptually, annual leave can be considered another health benefit: It benefits the individual to step back from the day-to-day stresses in an organization. Vacations refresh and renew. Time-off also allows employees to devote more time to friends and family, personal pursuits, recreational and leisure activities, and community service.^[5] As is the case with sick leave, vacation time can often be accumulated and applied toward retirement. Since this can create a substantial liability, the amount of leave that can be accumulated may be capped.

All types of leave—vacation, sick, and paid holidays (secular and religious)—may be lumped together in a leave bank. These comprehensive approaches to leave are often called PTO policies. When an organization has a PTO policy, employees are given a set number of days which they may use for sick leave, vacation, or some other form of leave (e.g., bereavement leave or personal leave). The PTO approach has several advantages. For instance, PTO eliminates concerns about sick leave abuse. It also reduces the organization's leave accounting and monitoring requirements. For example, with PTO managers are not required to determine whether an employee's leave should be coded “sick leave of vacation time”—it all comes out of the same PTO pool. Finally, PTO avoids the political questions that can arise regarding which holidays are observed and which are not.

Retirement Pension Benefits

Retirement pension benefits, like health care benefits, are both common and expensive. While the Social Security program will provide modest income benefits for retirement, employer-sponsored pensions provide public employees the security of knowing that they will have the needed income in their retirement years. For government employers, providing for employees' postemployment financial security is a matter

directly related to employee productivity: “To ensure employees’ current commitment and attention on productivity, future security must be guaranteed.”^[6]

There are two major types of pensions, defined benefit (DB) plans and defined contribution (DC) plans. The DB model has been, and continues to be, the predominate pension model in the public sector, whereas the private sector has increasingly opted for the DC model. Under a DB plan, an employee’s retirement benefit is calculated by a predetermined formula which normally includes employees’ age, salary, and years of service. The formula defines what employees who are vested (vesting typically occurs after 5 yr, but can take as long as 10 yr) in a given DB plan can expect to receive as a retirement benefit. Defined benefit plans are favored by many employees for this income predictability, the convenience of having professional money managers investing their assets, and the inflation countering cost of living adjustments retirees often receive. On the downside, DB plans can create large unfunded liabilities for employers are not easily portable for employees wishing to change employers, and do not allow individual employees to make investments that might produce higher rates of return.

In contrast to DB plans, retirement benefits under a DC plan are variable (i.e., undefined): employers and employees make regular contributions—usually at a fixed percentage of employee salary—to individual employee accounts. These contributions are invested by employees in a variety of employer-provided investment options. The accumulation of employer and employee contributions in individual employee accounts and the returns realized through the investment of those account assets determine employee retirement benefits. Commonly cited advantages of the DC model include the portability of benefits (since these are often minimal, e.g., 1 year, vesting requirements), individual control of investments, a clear bottom-line picture of what an employee has available for retirement, and for employers, the elimination of long-term pension obligations, and uncertain pension funding requirements. Disadvantages include the assumption of investment risk by employees and replacing professional money managers’ investment strategies with individual investment by employees possessing varying levels of financial planning and investment sophistication.

Public employers have also introduced hybrid pension plans that combine elements of both the DB and the DC models. Combination plans, for instance, direct the employer’s contribution to a traditional DB plan while the employee’s contributions are directed to a DC plan. With cash balance plans, employees receive benefits credits (i.e., a percentage of pay) based on years of service or, in some cases, years of service

plus age. Like DC plans, employers set up individual employee retirement accounts but these are really “phantom” accounts since the employer’s actual pension plan contributions are determined actuarially (i.e., sufficient contributions to cover future benefits), pooled into a pension fund, and invested collectively as in a DB plan.^[7] Thus, employee “accounts” track the hypothetical cash value of an employee’s pension benefit. This cash balance is the defined retirement benefit the employee is entitled to.

WORK-LIFE EMPLOYEE BENEFITS

In addition to traditional employee benefits, public employers have increasingly opted to enhance their employee benefits packages with a wide array of discretionary benefits. These work-life benefits go beyond the safety and security needs of traditional benefits to address employees’ diverse needs in a holistic manner that balances work, family, and personal needs.^[5] These benefits policies recognize the stresses that employees face in their lives and the impact that these stresses can have on employee performance.^[8]

Employee Assistance and Wellness Programs

Employees face personal problems and difficulties that can affect their ability to concentrate on work. In response, many employers have initiated employee assistance programs (EAP). Employee assistance programs provide employees with counseling and other forms of assistance for a variety of personal problems. For individual employees the availability of counseling, drug and alcohol treatment, financial advice and planning, legal assistance, and other EAP services can be quite encouraging (for example, see Refs.^[9–11]). Employee assistance programs are usually contracted for from outside groups and organizations. For example, a nonprofit, umbrella organization may serve as an overall EAP contact agency and coordinator. In such cases, the independent EAP can provide excellent services due to its community-wide economy of scale. The independence of outside EAP providers also assures employees confidentially of what are still personal problems.

Wellness programs focus on preventative health care. They undertake to encourage behavior that lead to good health and ease stress. They encourage individuals to exercise, eat healthily, and give up hazardous habits (e.g., smoking). Many of these activities are geared to behaviors that are associated with the risk of cancer and heart disease—two of the most expensive insured illnesses. As a part of such efforts, organizations may actually establish gyms or health spas for

their employees or, alternatively, subsidize memberships (with reimbursement linked to actual spa/club attendance). Many large organizations construct walking trails around and build their parking lots at the edge of their “campuses.” As a social activity, employee sports teams may be encouraged. In all of these instances, the goals are to improve employee health (which should result in greater productivity) and to reduce employer and employee health care costs.

Human Capital

Education, training, and professional conferences are all means of developing human capital and enhancing organizational productivity. The knowledge-based organization must invest in its people if it is to ensure its own existence and prosperity. While clearly designed to make the organization better able to cope with its environment, employee development is also a major individual benefit.

Employee development has the added advantage of not only enhancing technical skills, but also of psychologically motivating the individuals involved. The organizational investment is a recognition of the employee’s worth. The added skills, while paid for by the organization, belong to the individual. For the organization to fully obtain the benefits of its education and training programs, it must retain the individual. This implies a long-term relationship and fosters organizational commitment and loyalty.

Tuition reimbursement and educational leave are the two popular means of encouraging employees to add to their knowledge and skills. Prior approval of course work is usually required in tuition reimbursement programs. These programs also usually stipulate that courses are job-related (thus benefiting the employer and the employee) and that the minimum of letter grade (e.g., a “B” or equivalent) be earned. Educational leave may vary from a flextime arrangement (with work hours made up) to granting PTO for courses. A few public organizations (such as the military) even send employees to school as their duty assignment. In all of these cases, employers help individuals achieve their personal goals and, at the same, benefit by having those individuals who apply their new knowledge, skills, and abilities on the job immediately.

Family (or Worker) Friendly Benefits

Employers have begun to realize the interdependence of work life and family life: Even when an employee is physically present at work, he or she may be pre-occupied or otherwise affected by family or personal concerns.^[12] Many employers have responded by

offering family friendly benefits. Originally, these benefits were in response to demographic and social changes that have made women a permanent part of the modern workforce. The strains of dual career couples, juggling careers, and family obligations only exacerbated things. Since women in the workforce still bore the major brunt of family responsibilities, organizations found it necessary and advantageous to make adjustments to accommodate these requirements. The resulting family friendly benefits come in a variety of forms, including flextime (geared to school hours), flexible schedules and compressed workweeks, flexi-place (work at home or telecommuting), job sharing, leave pooling and sharing, elder care, child care, and educational assistance (e.g., tutoring, scholarships, school matching, etc.) for dependents (for example, see Refs.^[13–15]). Employers also may offer dual career couple (or trailing spouse) programs when recruiting new employees. Such programs range from assisting in job searches to actually creating a job for the spouse.

More recently, observers have suggested replacing the term family friendly with worker friendly or work–life balance.^[2] This broader term recognizes that demographic changes continue to affect the workplace, resulting in more single and childless workers and in more unmarried couples. Since flexible scheduling, leave, and other work–life policies are beneficial to employees both with and without families, the term worker friendly may be more appropriate. Also, in relation to changing demographics, some employers are beginning to offer domestic partner benefits which extend the same benefits married couples receive to unmarried couples (i.e., homosexual and, in some cases, heterosexual couples) living together.

Other Benefits

There are a variety of other employee benefits that are offered, to varying degrees, by employers. Directly and indirectly, employers can subsidize living expenses. Some employers provide housing allowances and underwrite mortgages. Some may actually provide the housing itself (e.g., police departments may offer police officers free housing in high-crime neighborhoods). In recruitment, employers often pay for relocation and temporary housing expenses. In some cases, the organization may even assist in the sale (including buying) of an existing house.

Employers also provide a number of convenience benefits. For example, on-site commissaries and cafeterias can reduce food costs. These facilities may also package “take-out” meals for employees to take home at the end of their workday. Other retail services (e.g., dry cleaning, car washing, and/or oil changes, movie passes, tickets to sporting and entertainment events,

etc.) may also be made available to employees at discounted rates.

ISSUES IN EMPLOYEE BENEFITS

Employee benefits are integral to organizations' human resource and performance goals. To remain competitive in the labor market, employers must offer an array of benefits that meet the diverse needs of an increasingly diverse workforce. In designing and administering employee benefits programs, public organizations face a host of challenges. Some of the more salient ones include ensuring that employee benefits are efficient, effective and competitive, protecting employee privacy, and containing costs.

Efficient, Effective, and Competitive Employee Benefits

With employee benefits, it is a truism that there is no such thing as "one-size-fits-all." Despite organizations' best intentions, one generous benefits package is unlikely to meet the needs of all workers. One way this potential motivational hurdle can be cleared is through cafeteria benefit plans. Cafeteria plans attempt to fine-tune the benefits offered by allowing individual employees to allocate their benefit dollars among those options that they themselves deem most useful. Some benefits are obviously mandatory for all employees, others are a part of the standard fare of traditional benefits expected by all employees. However, there are still other benefits that are merely in the desirable category. While many employees may often desire them, for others they are clearly inappropriate. To provide such benefits to all employees is a waste of resources both in terms of the money spent and in terms of the motivation they fail to elicit.

While cafeteria plans offer means for ensuring a fit between employee needs and the benefits provided, a more fundamental approach is to carefully plan, administer, and evaluate the benefits offered. In designing employee benefits, organizations should conduct a needs assessment to include such things as employee demographic data, employee-identified benefits needs, utilization levels for current benefits, and satisfaction with existing benefits.^[16] Once appropriate benefits options have been identified, organizations must then carefully implement and evaluate their benefits policies. This requires several steps: setting policies and values for the benefits program, identifying options or models for specific benefits to be offered, articulating the goals and objectives of the benefits program, planning for program implementation, specifying

outcomes and benefits, and measuring and evaluating benefits.^[16]

Steps such as these can go a long way toward ensuring that an organization has an attractive array of employee benefits that are effective in meeting employee needs and the organization's human resource and performance goals. Having knowledge about the type of benefits offered by the other competing organizations and about which benefits are important to employees enables an organization to design a plan that will then remain competitive in the labor market, meet employees' needs, and produce the desired productivity gains.^[1]

Privacy Concerns

Health concerns may also lead to the organization becoming more intrusive. Lifestyle behaviors, which increase health risks (and potential costs) may be targeted for modification—"lose the habit or the organization loses you." The exercise, diet, and stress reduction options may become required.

Growing concerns are also evident in regard to the use of medical records and genetic testing. Illness is not neutral; they are viewed as "deserving" and "undeserving." Individuals attach social stigmas to illnesses that are the consequences of lifestyle choices and, therefore perceived as preventable. Whether the dissemination and use of such information is deemed to be relevant and job-related is still a subject of debate. Clearly, information regarding these stigmatized illnesses can exert a negative effect on careers.

Finally, privacy provisions of the Health Insurance Portability and Accountability Act of 1996 (HIPPA) create additional requirements for health care insurers and providers and for employers. Specifically, HIPPA requires employers to keep employees' health-related information protected and secure.

Cost Containment

As benefits packages become more generous and as certain benefits costs (e.g., health care) continue to rise, organizations are looking for ways to contain benefit costs.^[17] In addition to the deductible and co-payment mechanisms discussed above, health care cost containment has led to the creation of various managed care options as alternatives to the traditional fee-for-services physician. The Preferred Provider Organization lists doctors who accept patients under a set-payment schedule (analogous to Medicare provisions). The Health Maintenance Organization (HMO) is a business that provides full medical care for a preset fee. Each of these options, more or less, introduces

financial incentives in balance to the provision of medical services.

Under HMO plans doctors are usually on salary rather than “commission” as is the case under fee-for-service arrangements. In addition, it is usually some form of gainsharing or bonus incentive included as an added inducement against overmedication. It is assumed that the doctor’s intrinsic desire to professionally serve their patients will counterbalance these financial incentives to insure adequate patient health care.

A utilization review is conducted prior to authorizing a procedure. The purpose is to ensure that necessary medical services are provided while “what if” and “wouldn’t it be nice” are discouraged. These utilization review procedures often raise the specter of individual’s being denied needed services for the sake of corporate profits. Without a doubt, some of this occurs especially where reviewers lack the relevant professional expertise of either physicians or registered nurses. Even so, the traditional fee-for-service approach often profited by ordered unneeded tests. Since there is an added danger in overmedication and from unnecessary procedures, utilization review is also likely to protect an individual’s health and well-being. Stringent cost containment efforts can adversely affect health care options for employees if physicians and clinics or HMOs refuse to accept government employees (or new employees) due to the payment schedule. This can create a significant access problem.

Cost containment also is an issue for retirement pension plans. Traditional DB plans can create long-term unfunded liabilities for governments. A variety of political, social, and economic developments have strained public employers’ ability to meet pension commitments in recent years. Politically, elected officials may choose to enhance pension benefits for an active political class—public sector employees and retirees—without always fully considering the long-term financial implications of those enhancements. Short-term budgetary pressures also affect retirement pensions. For example, elected officials sometimes choose to defer employer payments to pension funds—or worse yet, “raid” the pension funds for supplemental cash—in order to meet short-term budget shortfalls. Socially, demographic trends in the United States point to an aging population. The prospect of a large cohort of employees becoming eligible for retirement coupled with longer life expectancies means that pension systems face potential strain. Economically, stock market downturns in the early 2000s created alarming (at least in the short-term) pension fund investment losses in the first few years of the 21st century.

To ameliorate these costs concerns, governments are considering a number of options. As mentioned

previously, some governments are considering switching to DC programs to limit their long-term pension fund liabilities. Other governments have issued pension obligation bonds, which allow them to meet current unfunded pension fund liabilities but spread the costs out over the life of the bonds. Finally, some governments are creating two-tiered pensions plans in which newly hired employees receive less generous pension benefits than existing employees.

CONCLUSIONS

Employee benefits are a crucial component of an employers’ total compensation systems. Designing and administering an attractive yet affordable employee benefits program is vital for public organizations seeking to recruit, retain, and motivate qualified employees. What is more, employee expectations and needs continue to evolve, thus placing greater pressure on employers to seek innovative benefits solutions that meet both organizational and employee needs. Public organizations are responding to these challenges with employee benefits packages that meet employees’ security and safety through traditional benefits and employees’ personal needs through a host of work-life benefits.

REFERENCES

1. Society for Human Resource Management (SHRM). In *The Fundamentals of Human Resource Management*; Society for Human Resource Management: Alexandria, VA, 2002.
2. Hoyman, M.; Duer, H. A typology of workplace policies: worker friendly vs. family friendly? *Rev. Public Personnel Adm.* **2004**, *24* (2), 113–132.
3. Roberts, G.E. Municipal government benefits and personnel outcomes: results from a national survey. *Public Personnel Manag.* **2004**, *33* (1), 1–22.
4. Cayer, N.J. Issues in compensation and benefits. In *Public Personnel Management: Current Concerns, Future Challenges*, 2nd Ed.; Ban, C., Riccucci, N.M., Eds.; Longman: New York, 1997; 221–236.
5. Roberts, G.E.; Gianakis, J.A.; McCue, C.; Wang, X. Traditional and family-friendly benefits in local government: results from a national survey. *Public Personnel Manag.* **2004**, *33* (3), 307–330.
6. Daley, D.M. An overview of benefits for the public sector: not on the fringes anymore. *Rev. Public Personnel Adm.* **1998**, *18* (3), 5–22.
7. Green, L.B. *Questions and Answers on Cash Balance Pension Plans*; Compensation and Working Conditions Online; United States Department of Labor, Bureau of Labor Statistics: Washington, DC, 2003; <http://www.bls.gov/opub/cwc/cm20030917ar01p1.htm> (accessed December 2004).

8. Volk, W.; Cayer, N.J. *Work-Life Balance: Integrating Benefits with Expectations*, IQ Service Report 31 (11); International City/County Management Association: Washington, DC, 1999; 1–16.
9. Johnson, A.T. A comparison of employee assistance programs in corporate and government organizational contexts. *Rev. Public Personnel Adm.* **1986**, *6* (2), 28–42.
10. Johnson, A.T.; O'Neill, N. Employee assistance programs and the troubled employee in the public sector workplace. *Rev. Public Personnel Adm.* **1989**, *9* (3), 66–80.
11. Mani, B.G. Difficulties, assumptions, and choices in evaluating employee assistance programs. *Rev. Public Personnel Adm.* **1991**, *11* (1), 70–80.
12. Champion-Hughes, R. Totally integrated employee benefits. *Public Personnel Manag.* **2001**, *30* (3), 287–302.
13. Roberts, G.E. An examination of employee benefits cost control strategies in New Jersey local governments. *Public Personnel Manag.* **2001**, *30* (3), 303–321.
14. Pynes, J.E. *Human Resource Management for Public and Nonprofit Organizations*, 2nd Ed.; Jossey-Bass Publishers: San Francisco, CA, 2004.
15. Cayer, N.J. Public employee benefits and the changing nature of work. In *Public Personnel Administration: Problems and Prospects*, 4th Ed.; Hays, S.W., Kearney, R.C., Eds.; Prentice Hall: Upper Saddle River, NJ, 2003; 167–179.
16. Berman, E.M.; Bowman, J.S.; West, J.P.; Van Wart, M. *Human Resource Management in Public Service*; Sage Publications: Thousand Oaks, CA, 2001.
17. Roberts, G.E. An examination of employee benefits cost control strategies in New Jersey local governments. *Public Personnel Manag.* **2001**, *30* (3), 303–321.

Enterprise: Public versus Private

Thomas S. Nesselroth

Urban and Regional Studies, Department of Economics, University of Wisconsin-Green Bay,
Green Bay, Wisconsin, U.S.A.

INTRODUCTION

For the major part of the twentieth century, a key topic of political economy was the question of the appropriate scope of government suppliers of goods and services relative to private suppliers. In the socialist-planned economies of the former Soviet Union, Eastern Europe, and China, state enterprises dominated most of the economy. Moreover, a great number of developing countries sought economic growth via state-owned firms in key sectors of the economy where often some 10–12% of the national output was produced. Likewise, in the capitalist welfare states—for example, most northern European countries, the United Kingdom, Australia, and New Zealand—the dominant view was that government enterprises should control the commanding heights of the economy. Typical examples were the railroads, the mining sector, telecommunications, banking and insurance, and key industrial sectors such as chemicals and steel. Of course, the nationalization of education was strongly supported without serious debate. In distinct contrast, in the United States, direct government production was largely limited to major parts of the education sector, the postal system, and local government utilities and services such as police/fire protection and refuse collection.

Since the mid-1970s, however, radical change in thinking concerning the appropriate scope of direct government production has occurred.^[1] The collapse of Communism in Europe has led to massive privatization of formerly state enterprises, and the privatization movement has also taken hold in the capitalist welfare states and spread extensively throughout the developing world. In the United States, serious questions are now being raised with respect to the continued direct government production of education, refuse collection, prisons, housing, mail delivery, and air traffic control, among others.

The widespread enthusiasm for government enterprise, especially during the first 30 years after World War II, seems largely linked to the failures of the market economy during the great depression and the apparent success of the Soviet-planned economy. However, advances in economic theory, emerging empirical research, and changing circumstances have greatly weakened the case for direct government provision.

THE DESIRABLE SCOPE OF PUBLIC PROVISION: A MODERN VIEW

What kinds of goods and services should be produced by public enterprises rather than private suppliers?^[1] Let us begin by examining the case of benevolent government; that is, we assume that the government's goal is to maximize society's welfare. It is instructive to note then the conditions under which it makes no difference whether production is public or private. In brief, if the government knows precisely what kind of good or service it wants to produce, then all the relevant product characteristics can be specified in a contract (or regulation), and production can be contracted out to private suppliers. In such cases, governments may continue to *finance* production (e.g., education), but direct government production is no longer essential. For example, it is often believed that a government postal system is required because the delivery of mail to sparsely populated areas is generally unprofitable. Yet, the government can easily specify in a contract that private firms operating a postal concession must deliver mail to all areas.

Of course, the ability to write perfect contracts and regulations often does not exist. There are many instances where the government cannot precisely and completely specify and enforce all the *quality* dimensions that it deems desirable. In essence, many real-world cases involve what may be labeled as *incomplete contracts*. In many cases, the government may desire the supply of specific product characteristics that are impossible to specify and enforce contractually, so-called non-contractible *quality*. For example, the government may be concerned with how well prisons treat inmates, how well schools inculcate character and social responsibility, and how innovative producers are. Consequently, in this view, an important element in choosing between public or private production is how different forms of ownership affect the incentives to deliver the desired non-contractible quality attributes as well as the cost. In this respect, Shleifer states:^[1]

[C]onsider two types of investment incentives: those to reduce costs and those to improve quality or to innovate. When assets are publicly owned, the public manager has relatively weak incentives to make either of these

investments, because this manager is not the owner and hence gets only a fraction of the return. In contrast, private regulated contractors have much stronger incentives because, as owners, they get more of the returns on the investment. Which ownership structure is more efficient depends on whether having high-powered incentives to invest and innovate is a good idea.

However, the case against government enterprise is not so clear-cut when private suppliers have strong incentives to reduce cost, and increase profits, by decreases in noncontractible quality. Put differently, in some cases, efficient production may require *soft* incentives. For example, private schools, hospitals, prisons, and so on may have strong incentives to hire less well-trained and qualified workers to reduce labor costs.

Two market forces, however, often attenuate inefficient decreases in noncontractible quality. First, if the product market is reasonably competitive, consumers will shun suppliers making inefficient reductions in noncontractible quality. Second, in many markets, a key element of long-run profitability is repeat sales. In such cases, private suppliers are given strong incentive to avoid actions that increase short-run profitability at the expense of long-run profitability. Finally, it should be noted that private, not-for-profit firms often emerge in a market economy in those instances where soft incentives seem particularly important. In the United States, for example, a large sector of private, not-for-profit hospitals and universities has long existed.

Assuming benevolent government, Shleifer summarizes the case against government enterprise as follows.^[1]

These considerations point to a rather narrow set of circumstances in which government ownership is likely to be superior. These are the situations in which: 1) opportunities for cost reductions that lead to noncontractible deterioration of quality are significant; 2) innovation is relatively unimportant; 3) competition is weak and consumer choice is ineffective; and 4) reputational mechanisms are also weak. This list gives, in my opinion, a fair sense of how tenuous, in general, is the normative case for government production.

GOVERNMENT FAILURE AND THE CASE FOR PRIVATIZATION

So far, the discussion has assumed a benevolent government seeking to maximize society's welfare. It has now long been recognized, however, that government policies and programs often fail in this respect. In particular, special interest groups in many cases dominate the political process and the political outcomes that emerge are often a pale reflection of society's general interest.

In brief, highly inefficient policies and programs are adopted that to an important extent serve to transfer wealth to politically powerful interest groups. In this respect, government enterprises are often used by the ruling party as a means to increase their political support via above-market wages, excess employment, and outright transfers of wealth. In developing countries, in particular, government monopolies and projects are created expressly to transfer wealth, often to a handful of relatives and key political supporters of the ruling family or party. Shleifer summarized as follows: "In other words, state firms are inefficient not just because their managers have weak incentives to reduce costs, but because inefficiency is the result of the government's deliberate policy to transfer resources to supporters."^[1]

Moreover, in a highly nuanced critique of both neo-classical economics and socialist economic planning, Stiglitz also recognized both a political and a economic case favoring privatization.^[2] To begin, privatization is seen as a means to weaken the coercive powers of the state and to depoliticize enterprise operations. Moreover, Stiglitz argued that a strong economic case can be made for privatization, but one that goes beyond the standard arguments regarding differences in managerial incentives. Stiglitz stressed three points. First, privatization is seen as enhancing the government's commitment to a competitive economy and to sharply circumscribing enterprise subsidies. Second, incentives are strengthened as privatization increases the likelihood that hard enterprise budget constraints will be substituted for soft budget constraints. Third, a dynamically efficient economy requires hard budget constraints to enhance the economy's "selection mechanism, through which it is decided not only which enterprises survive but also which enterprises garner additional resources to expand. Those who cannot meet the market test are weeded out."^[2]

COMPARATIVE ENTERPRISE EFFICIENCY: THE EVIDENCE

De Alessi undertook the first major survey of comparative enterprise efficiency within the context of a broader analysis of the economics of property rights.^[3] De Alessi concisely examined key empirical studies from 1960s to 1970s that assess the relative efficiency of government vs. private enterprises with respect to water and electric utilities, urban transit, airlines, banks, hospitals, fire prevention services, and refuse collection. De Alessi concluded from his review that strong evidence exists for the superiority of private ownership.

In contrast, three major reviews during the 1980s found little or no difference in efficiency between public

and private enterprises.^[4] In brief, these studies suggest that differences in enterprise efficiency are overwhelmingly related to the degree of competition faced in the market, not ownership per se. Vining and Boardman labeled this the *primacy of competition vs. ownership* argument.^[5] In the extreme, the argument implies that in highly competitive markets, without significant distortions, ownership does not matter for enterprise efficiency. More recent evidence strongly counters this view, however.^[6]

To begin, Mueller presented a table of some 50 studies examining comparative enterprise efficiency and concisely summarized the cost differences found.^[6] In only two studies were government enterprises found to be more efficient. In over 40 studies, government firms were found to be the higher cost producers, generally more than 20% higher, and often 40–60% (or more) higher. The case favoring private ownership is further buttressed by several major empirical studies undertaken since the late 1980s.^[6] Of particular note are the related studies of Boardman and Vining, and Vining and Boardman.

A key argument made by these authors is that the relatively few studies that show public enterprise superiority, or at least no difference in efficiency, cannot convincingly demonstrate their claim. In brief, these studies compared firms in sectors with characteristics of monopoly, duopoly, and/or heavy regulation, and thus are not relevant to assessing comparative efficiency in a competitive market. In this respect, Vining and Boardman stated:^[5]

Many studies cited ... pertain to sectors where "firms" have geographic monopolies (electric utilities, water, refuse, fire services) and, therefore, do not directly compete with each other. While these studies tell us something about the effect of ownership on performance in non-competitive, monopoly environments, by definition they cannot inform us about the effect of ownership in competitive environments.

With respect to the empirical evidence, Boardman and Vining used a statistical analysis capable of measuring the independent effect of ownership on comparative enterprise performance.^[6] They estimated their model using several measures of profitability and efficiency as related to three enterprise types: state-owned enterprises, mixed enterprises (a hybrid ownership category where the government has retained an equity position), and private enterprises. Their study used data on the 500 largest manufacturing and mining firms from outside the United States as of 1983. After controlling for a range of factors (e.g., firm size and degree of product market competition), Boardman and Vining found that ownership has a

substantial independent effect on enterprise profitability and efficiency. For example, on average, state-owned enterprises have a return on equity almost 12% less than comparable private enterprises. Relatedly, two studies using a similar methodology, but focusing on the 500 largest non-financial enterprises in Canada (Vining and Boardman) and the 500 largest non-U.S. corporations (Picot and Kaulman), also find credible evidence that state enterprises are less efficient than private enterprises.^[6] More recently, Boubakri and Cosset examined changes in the economic performance of 79 enterprises from 21 developing countries that experienced full or partial privatization between 1980 and 1992.^[6] They find substantial increases in profitability, operating efficiency, capital investment spending, output, employment level, and dividends following privatization.

CONCLUSION

Since the mid-1970s, a radical change in thinking concerning the appropriate scope of direct government production has occurred. The reorientation has been stimulated in part by advances in economic theory—in particular, the theory of ownership and contracting, a greater appreciation of the link between ownership and innovation for a dynamically efficient economy, and a fuller understanding of the adverse consequences of government ownership. Also playing a role is the increasing competitiveness of market economies and improvements in government contracting and regulation. Finally, much real-world experience and a growing body of high-quality empirical research supports the comparative efficiency of private enterprise.

REFERENCES

1. Shleifer, A. State versus private ownership. *J. Econ. Perspect.* **1998**, 12 (4), 133–150. The following two sections concisely summarize Andrei Shleifer's seminal analysis of the desirable scope of public provision from a modern perspective.
2. Stiglitz, J. *Whither Socialism?* MIT Press: Cambridge, MA, 1994; 181 pp.
3. De Alessi, L. The economics of property rights: a review of the evidence. In *Research in Law and Economics*; Zerbe, R., Ed.; JAI Press: Greenwich, CT, 1980; Vol. 2, 1–47.
4. Borchering, T.E.; Pommerehne, W.W.; Schneider, F. Comparing the efficiency of private and public production: a survey of the evidence from five federal states. *Z. Nationalökonom./J. Econ. Theory* **1982**, 127–156.
- Millward, R. The comparative performance of public and private ownership. In *The Mixed Economy*; Roll, E., Ed.; Macmillan: London, 1982; Boyd, C.C. The comparative efficiency of state-owned enterprises. In *Multinational*

Corporations and State-Owned Enterprises: A New Challenge in International Business; Negaandhi, A.R., Thomas, H., Rao, K.L.K., Eds.; JAI Press: Greenwich, CT, 1986; 99–184.

5. Vining, A.; Boardman, A. Ownership versus competition: efficiency in public enterprise. *Public Choice* **1992**, *73*, 205, 216.
6. Mueller, D.C. *Public Choice II*; Cambridge University Press: Cambridge, U.K. Boardman, A.; Vining, A. Ownership and performance in competitive environments: a comparison of the performance of private, mixed, and state-owned enterprises. *J. Law Econ.* **1989**, *32* (1), 1–33; Vining, A.; Boardman, A. Ownership versus competition: efficiency in public enterprise. *Public Choice* **1992**, *73*, 205–239; Picot, A.; Kaulman, T. Comparative performance of government-owned and privately-owned industrial corporations—Empirical results from six countries. *J. Inst. Theor. Econ.* **1989**, *145* (June), 298–316; World Bank. *Bureaucrats in Business*; Oxford University Press: London, 1995; Boubakri, N.; Cosset, J. The financial and operating performance of newly privatized firms: evidence from developing countries. *J. Finance* **1998**, *53* (3), 1081–1110; Cowie, J. The technical efficiency of public and private ownership in the rail industry: the case of swiss private railways. *J. Transp. Econ. Policy* **1999**, *33* (3), 241–251; Plane, P. Privatization, technical efficiency and welfare consequences: the case of the cote d'Ivoire electricity company (CIE). *World Dev.* **1999**, *27* (2), 343–360; Ros, A. Does ownership or competition matter? The effects of telecommunications reforms on network expansion and efficiency. *J. Regul. Econ.* **1999**, *15* (1), 65–92; Scully, G. Reform and efficiency gains in the New Zealand electrical supply industry. *J. Product. Anal.* **1999**, *11* (2), 133–147.

Entitlements and the Congressional Budget Process

Patrick Fisher

*Department of Political Science/Philosophy, Monmouth University,
West Long Branch, New Jersey, U.S.A.*

INTRODUCTION

Entitlements are government spending programs for which Congress has set eligibility criteria—age, income, location, occupation, etc. If meeting the criteria, a recipient is “entitled” to the money. Due to the rapid growth of entitlement programs over the past four decades, less than half of the federal budget today is considered “controllable” through the budget process. The fact that entitlements are broadly considered to be relatively “uncontrollable” is somewhat misleading. While Congress cannot control levels of eligibility in the population, it can control such levels under the law (cost of living adjustments, for example). It would be accurate to say, however, that politically, entitlements are relatively uncontrollable. Entitlements are available for a vote, but only if Congress decides to arrange one. Congress can change or repeal any law it passes, and lawmakers can revisit entitlements at any time to reduce or eliminate them.

THE GROWTH ENTITLEMENT SPENDING

Federal entitlement programs range from gigantic ones, such as Social Security and Medicare, to extremely small ones, such as an indemnity program for dairy farmers whose milk is contaminated by chemicals. Some entitlement payment levels, such as for unemployment insurance, are difficult to predict, while others, such as Social Security, are fairly predictable. While most entitlements go to people, some also go to other units of government. The Title XX Social Services block grant, for instance, goes to states based on population.

What makes entitlements different from the other major forms of congressional spending is that they are mandatory; money must be provided until the program is changed by Congress. By contrast, discretionary spending is good only for one year, and Congress has to renew it annually through the appropriations process. Less than a third of the budget during the Kennedy administration, entitlement, and mandatory spending is now more than one-half of the budget. Discretionary spending (for domestic, defense, and international purposes), on the other hand, has gone from making up 63% of federal government expenditures

in 1962 down to about a third in 2000 (see [Table 1](#)). The tremendous growth of entitlements may be problematic, because entitlements limit the ability of Congress to maneuver in adopting the budget. Congress, therefore, may not be given the flexibility it needs in the budget process. Because most government expenditures are disbursed in long-term authorizations, there is some question about the democratic accountability of the budget process.^[1]

ENTITLEMENTS AND THE TRANSFORMATION OF THE CONGRESSIONAL BUDGET PROCESS

Congress has deliberately authorized entitlement spending because it believed that the government should protect certain obligations from yearly political controversy. Members of Congress have some control over annual entitlement spending, but the total changes in any single year are a relatively small part of the entire budget. Over the long run, however, there is the potential for adequate flexibility within the budget for Congress to make significant changes in the composition of the federal budget.^[2]

Budgeting is a different process from what it once was. The postwar transformation of federal budgeting from a control process to one oriented toward spending growth has upset the relationship between available resources and demands on the budget. Claims on the budget became stronger when budgeting was an expansionary process, and it has not been easy for Congress to reverse course, as the climate has become more constrained.^[3] Budgeting for growth (1950s–1960s) differs from budgeting for stabilization or cutback (1970s–1990s). The expectation of growth makes it easier to please greater numbers of actors, because government can afford to give people what they want. As critics of congressional budgeting argue, this transformed budgetary environment has become a long-term threat to Congress’s power of the purse, because Congress had become unwilling, for political reasons, to control mandatory spending.^[4]

The budget cannot be completely isolated from the economy—the economy drives the budget. The assumptions on which the budget’s prescriptions are based all lie at the mercy of national economic conditions.

Table 1 The growth of entitlement spending: budget expenditures for fiscal years 1962 and 2000

	2000	1962
Entitlements and mandatory spending	55%	31%
Domestic discretionary spending	16%	12%
Defense spending	16%	46%
Interest on the national debt	12%	6%
International discretionary spending	1%	5%

Source: Congressional Budget Office.

Beginning in the early 1970s, demands on the budget soared more than resources did. Congress did not foresee in the 1960s that the budget situation would become much more dire as the economy stagnated.^[3] Politics and economics combine to shape the varying patterns of activity for the legislators working on the budget.^[5] On the one hand, economic goal-setting has often been a frivolous exercise in symbolic politics in the absence of realistic means to achieve goals. Year after year, the most important factor in determining the current year's budget allocations is last year's budget allocations. On the other hand, the establishment of inflexible methods to achieve a goal may simply make other problems worse and threaten the economy.^[6] As economic conditions change, so do priorities; legislators need to maintain their ability to respond to varying economic conditions.

THE POLITICAL SIGNIFICANCE OF ENTITLEMENT PROGRAMS

There are good policy reasons for entitlement spending. Entitlements can be an efficient way of the government providing services, and it seems logical that the government may want to avoid the political difficulty of not being able to maintain its promises, regardless of the financial condition of the government, which is, after all, highly reliant on the state of the economy. One problem with entitlements is that, to a large degree, they go to fund what has been called "middle-class welfare." That is, entitlement benefits are dispensed largely on the basis of criteria other than income (such as age). Programs that require proof of low income to receive benefits are a small proportion of entitlements; more than three-fourths of all entitlement spending (including Social Security and Medicare) is directed toward non-means tested programs, where recipient eligibility is not based on need (see Table 2).

The huge surge in entitlement spending came between 1966 and 1976. During this period, the creation of programs such as Medicare, Medicaid, and food stamps combined with expanded Social Security

Table 2 The top entitlements

Rank	Program	Expenditures (billion of \$)	% of GDP
1.	Social Security	406.0	4.1
2.	Medicare	216.0	2.2
3.	Medicaid	117.4	1.2
4.	Other retirement/disability	87.8	0.9
5.	Farm price supports	30.5	0.3
6.	Unemployment	20.7	0.2
	Total means tested	235.9	2.4
	Total non-means tested	793.9	8.1

Source: Congressional Budget Office.

benefits to provide an enormous boost to entitlement spending. From 1966–1976, entitlements more than doubled in relation to the size of the economy, growing from 5.4% of the GDP to 11.3%, where it leveled until health care costs began driving entitlement costs up again in the 1990s. All together, three entitlement programs—Social Security, Medicare, and Medicaid—make up more than 70% of all mandatory spending.

The rapid growth of entitlements has given them a bad name in some quarters, because they were held to blame for deficits that plagued the federal government through the 1980s and early 1990s. Furthermore, entitlements came to be seen as a means by which the authorization committees could circumvent the appropriations committees. Dramatically increased entitlements spending, without a corresponding willingness to raise taxes, broke the back of the classical budget process.^[1]

Conservatives fear that the result of the growth in entitlements will be higher taxes and deficits; liberals fear that entitlements will squeeze out other programs and, if they keep growing, use up future revenue increases. Yet both conservatives and liberals in Congress have been unwilling to confront the growth in entitlements. As a result, as Social Security has been placed politically out of reach and other entitlements get harder to contain, discretionary spending has become a larger part of what is up for grabs in the short term. This, in turn, has curtailed the power of the appropriations committees. The piece of the budget they control was nearly three-fourths of the budget during the Kennedy Administration; now it is down to little more than one-third of the budget and is dropping.

Some argue that the difficulty in reducing or eliminating entitlements is exactly what the architects of entitlements had intended, and exactly what contemporary Congresses ought to be doing. Entitlement programs dole out benefits automatically so that basic benefits would not be subject to the year-to-year inconsistencies of the appropriations process. Furthermore,

entitlements are growing as they were supposed to, especially during a bad economy. The so-called automatic stabilizers such as unemployment insurance and food stamps are designed to pick up the slack during bad economic periods.

Entitlement programs, however, do not always have the effect that their creators intended. Times change, but often, established entitlements are unable to change as new circumstances dictate. A problem with entitlements is that it is much easier to start an entitlement for an apparently needy group than it is to terminate an entitlement after it no longer makes sense. Most subsidy programs, for example, began for the same reason as programs to benefit individuals—that is, they were a response to hardship cases. In their early days, aid was targeted on the basis of an immediate need or a significant national purpose. Over the years, many of these programs, especially in agriculture, lost their focus as conditions changed. Instead of being phased out, many were expanded for political purposes.^[7]

Another entitlement that has had unintended consequences is the civil service retirement program, where expenditures are currently skyrocketing. Civil service retirement is subject to its own demographic characteristics: the number of retirees has increased dramatically in recent years because of the large number of civilians on the federal payroll during and following World War II. Furthermore, expenditures will increase more rapidly in the future because of the benefit characteristics of the system. Not only do federal retirees receive relatively generous benefits, but the system encourages early retirement, further adding to the costs for the government and taxpayers.^[8]

Some of the largest and most expensive programs in the federal budget, such as Social Security and Medicare, are entitlements that are perceived to be politically hazardous to alter. Social Security, in fact, is called by politicians “the third rail of American politics”—touch it and die. Given a choice, members of Congress would rather look almost anywhere else for ways to cut spending. Unfortunately, the cost of Social Security and Medicare is rising so fast that they risk crowding out the rest of domestic spending. The political power of Social Security was demonstrated with a House vote in 1994 that called for attacking the deficit by reining in entitlement spending, but with a provision excluding Social Security from any cuts. A move to keep Social Security on the chopping block along with other entitlement programs lost by a resounding vote of 392-37. After this vote, the House then proceeded to vote 424-0 to prohibit any move to increase the Social Security payroll tax to help fill the gap in case entitlement spending ever exceeds preset limits.

The problem remains, however, that it is questionable whether or not the budget can be balanced over the long run without reforming Social Security.

Social Security is the largest single spending program in the federal budget, and at a cost of \$406 billion in 2000, it accounted for nearly a quarter out of every dollar the federal government spends. Defenders insist that it should remain as it is because it is self-financed by the payroll tax, which is currently running a huge surplus. A problem with this argument is that it leaves the mistaken impression that the biggest program in the budget would not affect the rest of the budget. Critics argue that the sheer size of Social Security makes it a must for any sizeable budget cuts. As the size of the House votes in 1994 show, however, any meaningful cuts in Social Security in the near future are extremely unlikely.

Medicare's political power was demonstrated after Congress passed a Medicare Catastrophic Coverage Act of 1988, which was to be paid for in large part by an income surtax on middle- and high-income beneficiaries. After ugly demonstrations led by those who were going to have to pay the surtax, however, Congress backed down and repealed the measure. Despite Congress' experience with the Catastrophic Coverage Act, there still remains the fact that you cannot control entitlement spending without altering Medicare and Medicaid.

It is doubtful, however, that Medicare (the federal health insurance for the elderly and disabled) and Medicaid (the joint federal-state health program for the poor) can be controlled without overhauling the nation's entire health system. Because climbing health care costs are expected to increase the deficit significantly as the baby boom generation nears retirement, early in his administration, President Clinton proposed attempting to channel the pressure for continuing deficit reduction into the health care reform debate. Politically, however, Clinton's health care strategy proved to be unsuccessful, and prospects for significant health reform were shattered with the Republican takeover of Congress in 1994. The fact remains, however, that to reduce Medicare and Medicaid effectively, the whole health care system has to be reformed.

Taken together, Medicare and Medicaid are almost as expensive as Social Security. They are also currently by far the fastest growing entitlements. Medicare and Medicaid have been growing at a tremendous rate because of demographics and technology. Demographically, health costs increase as Americans get older; the elderly consume far more health care services than younger people. Medical technology makes health costs more expensive, because it provides doctors and hospitals with more things they can do for patients. Thus, not only are our health costs skyrocketing, but the rising costs are not helping to cover those who need health care the most. The elderly receive a disproportionate share of the federal health insurance, largely because they have a great deal of political power.

The budgetary process, therefore, is an inherently weak mechanism for controlling mandatory entitlements such as Medicare and Medicaid.

When the budget is squeezed, as it has been for the past three decades, the easiest place to formulate budget cuts is in the “discretionary spending” portion of the budget. Discretionary spending includes the choices Congress has to fund state–local assistance and federal domestic programs which cover the spectrum of services—discretionary spending represents the money Congress appropriates every year for everything from battleships to Head Start. The constitutional authority that allows Congress to make annual appropriations, however, has been weakened considerably.

Due to the growth of entitlement spending, the sector of government that Congress can readily get to from one year to another is no longer expanding. In fact, during the 1990s, it actually shrunk in real terms. In 1990, Congress put strict limits on appropriations, which have since been tightened even further, imposing a “hard freeze,” i.e., not allowing increases to keep up with inflation. That policy kept discretionary spending virtually flat through the rest of the decade. That means that there was not enough money to cover the same value of services the government previously provided, much less expand them. The proliferation of entitlement spending at the expense of discretionary

spending may potentially sacrifice the capacity of Congress to build new programs appropriate to changing national needs.

REFERENCES

1. Wildavsky, A.; Caiden, N. *The New Politics of the Budgetary Process*, 3rd Ed.; Scott, Foresman and Company: Glenview, IL, 1997.
2. Leloup, L. *The Fiscal Congress*; Greenwood Press: Westport, CT, 1980.
3. Schick, A. *The Capacity to Budget*; The Urban Institute: Washington, DC, 1990.
4. Ippolito, D.S. *Hidden Spending*; University of North Carolina Press: Chapel Hill, 1984.
5. Fenno, R. *The Emergence of a Senate Leader: Pete Domenici and the Reagan Budget*; CQ Press: Washington, DC, 1991.
6. Leloup, L. A fiscal policy and congressional politics. In *Congressional Politics*; Deering, C.J., Ed.; Brooks/Cole: Pacific Grove, CA, 1989; 262–283.
7. Pascall, G. *The Trillion Dollar Budget*; University of Washington Press: Seattle, 1985.
8. Anderson, A.G. A civil service pay and retirement. In *Entitlement Issues in the Domestic Budget*; Weicher, J., Ed.; American Enterprise Institute: Washington, DC, 1985; 34–56.

Environmental Ethics

Christine M. Reed

*School of Public Administration, University of Nebraska at Omaha,
Omaha, Nebraska, U.S.A.*

INTRODUCTION

Environmental ethics is a branch of applied philosophy that addresses arguments about the value of nature. The literature in environmental ethics is voluminous, reflecting various philosophical traditions and therefore different ways of approaching such questions as whether the source of value in “nature” is objective, subjective, or socially and culturally constructed; whether value in nature is human centered (anthropocentric) or centered outside the sphere of human welfare (non-anthropocentric); and whether value in nature for itself is based on the interests and/or rights of individual animals, or on the preservation of species and ecosystems as a whole.

THE INTRINSIC VALUE OF NATURE

Intrinsic value is the idea that nature has value in itself and not as a means to further ends. One form of the intrinsic value argument locates the source of value in the objective properties of nature, such as the characteristics of a landscape prior to human cultivation.^[1] A second, subjectivist, form of the intrinsic value argument locates the source of value in the attitudinal frameworks of individuals, in particular approval of nonmoral value-adding properties in virtue of which wild nature has intrinsic value. One example of a value-adding property is the property of being naturally evolved.^[2] These metaethical differences often lead to similar normative claims, such as the duty to preserve and restore wild nature. In *Respect for Nature*, Taylor introduces a related, but different, concept: the inherent worth of wild plants and animals of the Earth’s natural ecosystems. He argues that moral agents will respect nature only if they deem non-human living beings to be possessors of inherent worth, as distinct from their instrumental value.^[3]

ESTHETIC VALUE IN NATURE

Another justification for environmental protection is to prevent the loss of esthetic value. According to Allen Carlson, pristine nature has positive esthetic value,

while replacing it with human alterations degrades nature’s essential beauty. For Carlson the characteristics that humans find esthetically pleasing also make the natural world more understandable: its unity, order, and harmony.^[4] A somewhat different argument is that the natural landscape is more beautiful than an artistic representation of that landscape, because esthetic qualities exist more fundamentally in the natural objects making up the natural landscape than in the painting depicting them.^[5] Esthetic qualities associated with natural beauty are nevertheless similar to those qualities attributed to art objects: environments that are magnificent and rich, that present the viewer with new ways of perceiving, that have cultural significance because of a connection with the past, and that challenge the way people live or think.^[6]

ENVIRONMENTAL VIRTUE ETHICS

This approach to environmental ethics locates intrinsic value in the virtues of the moral agent, specifically in environmental virtues. One example of an environmental virtue is humility, the mean between human arrogance on the one hand which closes off an individual to the idea of self as part of nature, and false modesty on the other which leads the individual to lose all sense of individuality in the presence of nature’s vastness and sublimity. Seeing oneself as part of nature is related to other virtues, such as an esthetic sense and capacity for gratitude.^[7] The capacity for gratitude, like the moral intuition of esthetic value in nature, is anthropocentric to the extent that the locus of intrinsic value is in the person rather than in nature itself; however, environmental virtue ethics reflect concern for nature and consider its wanton destruction a vice. Environmental virtues are often reflected in environmental literature, such as *Walden*, in which Thoreau exemplifies temperance, integrity, sensibility to natural beauty, and simplicity.^[8]

ENVIRONMENTAL PRAGMATISM

Some theories about value in nature are based on pragmatism and reject the idea that the locus of value is

either in human consciousness *or* in an independent nature. “Weak” anthropocentrism distinguishes between individual felt preferences for commodities that are satisfied by a market transaction vs. considered preferences that reflect human value formation arising from the experience of natural objects.^[9] Another approach based on environmental pragmatism views understandings of nature as culture-laden linguistic descriptions or metaphors that reflect the human experience of nature, rather than nature as an external reality.^[10] A third contribution challenges the boundary between mind and body, arguing that human cognitive abilities emerge from sensory–motor structures (schema) developed through a process of evolution similar to that of non-human animals.^[11]

CIVIC ENVIRONMENTALISM

The term “civic environmentalism” means that a community has organized on its own, rather than in response to federal mandates, to protect its environment. It refers to a process of local, communal self-reflection, and recognition that certain environmental hazards, such as “non-point” source pollution of streams and rivers, are created over time by the social practices of an entire community.^[12] Environmental ethics and democratic values come together when citizens are motivated by civic virtue to become engaged in the political and ecological welfare of their communities. A related concept is that of “ecological citizenship” in which public participation in local restoration projects is part of a broader project of active engagement in the public affairs of the community. The “ground” of citizenship in modern times is found at the level of local ecology.^[13]

OBLIGATIONS TO FUTURE GENERATIONS

A second example of an approach to linking environmental ethics to democratic values is concern for intergenerational equity: what are the obligations of public officials and of society as a whole to make policy decisions in the present that protect life in the future? One answer is to extend the thought experiment of John Rawls and to imagine people behind a “veil of ignorance” who are establishing procedures for a just society, and who do not know either their economic and social status in society or their place in history. Under those conditions, people will agree to rules of social justice that consider the welfare of future generations.^[14] The scope of intergenerational ecological problems, such as global warming and loss of biodiversity, involves public administration and policy at the

international level, as well as intervention in the private market, to regulate in favor of future generations.^[15]

ENVIRONMENTAL DISCOURSE ETHICS

A third, more radical, way to balance environmental and political values is to maintain concern for processes of governance that are legitimate because they are based on principles of justice, while extending the moral community to include non-humans. Green reason is an extension of the ideal of communicative rationality to non-humans.^[16] It involves defining non-linguistic beings as capable of communication. It also insists that freedom from domination and distortion in human-to-human discourse be extended to interaction with nature. Just as the policy agenda is often dominated by a few powerful interest groups, so too are human representations of the world privileged representations from other marginal positions, especially animal species.^[17] Extending communicative reason to make it truly egalitarian and free from the distortions of political power entails the inclusion of non-human as well as human beings. Discourse, however, is also based on “affect-mediated interpretations” of the needs of non-human others. The “life world” is shared with nonlinguistic beings.^[18]

POSTMODERN ENVIRONMENTAL ETHICS

According to a postmodern approach to environmental ethics, nature is what humans have constructed through their own social practices. The primary ethical responsibility is to be aware of nature as a social construction and to make human relationships with the natural world more just and accountable.^[19] But the preservation of wild nature is an anachronism because “nature” as a reality independent of human or cultural intervention no longer exists. Lamenting the end of nature just serves as an excuse to avoid our responsibility for what society has constructed in both figurative and literal senses of the term. The blurring of the boundaries between the spheres of nature and culture is manifest in the fact that no wilderness areas are left in the United States, apart from some remote areas of Alaska. We are now custodians of a managed world and responsible for all its life forms.^[20]

CONCLUSION: EXPLORING THE LINKAGES

The vast majority of work in environmental ethics consists of various philosophical arguments for locating intrinsic value in natural objects, in the virtues of the moral agent, or in humans as part of nature. There is

a potential conflict between individual human rights and environmental ethics, particularly when environmental ethics are extended to include animals or even ecosystems. Conflicts between human and non-human rights can pit environmental ethics against democratic values; however, efforts to reconcile the two, such as civic environmentalism and environmental discourse ethics, reframe the meaning of democratic practice in a way that enlarges citizenship to include both engagement with government in the larger context of the ecological community, and engagement with non-human others as part of our social life world.

REFERENCES

1. Attfield, R.; Belsey, A. Rehabilitating nature and making nature habitable. In *Philosophy and the Natural Environment*; Cambridge University Press: Cambridge, U.K., 1994; 45–57.
2. Elliot, R. *Faking Nature: The Ethics of Environmental Restoration*; Routledge: London, U.K., 1997.
3. Taylor, P.W. *Respect for Nature: A Theory of Environmental Ethics*; Princeton University Press: Princeton, NJ, 1986.
4. Hettinger, N. Allen Carlson's environmental aesthetics and the protection of the environment. *Environ. Ethics* **2005**, *27* (1), 57–76.
5. Hargrove, E. *Foundations of Environmental Ethics*; Prentice-Hall: Englewood Cliffs, NJ, 1989.
6. Thompson, J. Aesthetics and the value of nature. *Environ. Ethics* **1995**, *17* (3), 291–305.
7. Frasz, G. Virtue ethics: a new direction for environmental ethics. *Environ. Ethics* **1993**, *15* (3), 259–274.
8. Cafaro, P. Thoreau, Leopold and Carson: toward an environmental virtue ethics. In *Environmental Virtue Ethics*; Sandler, R., Cafaro, P., Eds.; Rowman & Littlefield: Lanham, MD, 2005; 31–44.
9. Norton, B. Environmental ethics and weak anthropocentrism. *Environ. Ethics* **1984**, *6* (2), 131–147.
10. Harlow, E. The human face of nature: environmental values and the limits of nonanthropocentrism. *Environ. Ethics* **1992**, *14* (1), 27–42.
11. Preston, C. Animality and morality: human reason as an animal activity. *Environ. Values* **2002**, *11*, 427–442.
12. John, D. *Civic Environmentalism: Alternatives to Regulation in States and Communities*; Congressional Quarterly Press: Washington, DC, 1994.
13. Light, A. Restoring ecological citizenship. In *Democracy and the Claims of Nature*; Minteer, B., Taylor, B., Eds.; Rowman & Littlefield: Lanham, MD, 2002; 153–172.
14. Wenz, P. *Environmental Ethics Today*; Oxford University Press: New York, NY, 2001.
15. Frederickson, F.G. Can public officials correctly be said to have obligations to future generations? *PAR* **1994**, *54* (5), 457–464.
16. Dryzek, J. Green reason: communicative ethics for the biosphere. In *Postmodern Environmental Ethics*; Dryzek, J., Oelschlaeger, M., Eds.; State University of New York Press: Albany, NY, 1995; 101–118.
17. Hayles, K. Searching for common ground. In *Reinventing Nature? Responses to Postmodern Deconstruction*; Island Press: WA, 47–63.
18. Smith, M. Citizens, denizens and the *res publica*: environmental ethics, structures of feeling and political expression. *Environ. Values* **2005**, *14*, 145–162.
19. Cronon, W. Introduction: in search of nature. In *Uncommon Ground: Rethinking the Human Place in Nature*; W.W. Norton & Company: New York, NY, 1996; 23–56.
20. Borgmann, A. The nature of reality and the reality of nature. In *Reinventing Nature? Responses to Postmodern Deconstruction*; Island Press: WA, 1995; 23–56.

Environmental Policy

Thomas Greitens

*American Farmland Trust, Center for Agriculture in the Environment,
Northern Illinois University, DeKalb, Illinois, U.S.A.*

INTRODUCTION

Federal environmental policies center around three pieces of legislation: the Clean Air Act, the Clean Water Act, and the National Environmental Policy Act (NEPA). These acts enhanced federal involvement in environmental policy and established a system of command and control regulation of industry. However, by the late 1980s this emphasis on command and control regulation diminished as policymakers started to emphasize concepts of decreased federal involvement that could help control environmental pollution. Among these concepts were notions of cooperative frameworks between industry and government as well as the concept of organizations implementing environmental management systems to improve economic efficiency through better environmental performance.

This decreased federal involvement in environmental policy is nothing new. In fact, initial federal statutes regarding clean air and water policies did not rely on direct federal involvement. Rather, these statutes typically allowed states to set environmental standards with no federal guidelines. Beginning in the 1960s and 1970s, this began to change as the public increasingly became aware of environmental problems and the federal government realized states implemented environmental standards inconsistently. By briefly analyzing the historical development of federal environmental policies, this changing pattern of federal involvement in environmental policy becomes clarified.

CLEAN AIR POLICY

Before 1955, air pollution statutes were mainly state or local affairs. This started to change with the 1955 Air Pollution Control Act. In this act, the federal government provided states with funds to conduct air pollution research and train personnel.^[1] By 1963, this act was supplanted with the original Clean Air Act. The original Clean Air Act provided additional federal support for air pollution research, funded the development of state pollution control agencies, and allowed

the federal government to assist states when issues of interstate air pollution occurred.^[1] Although significant, these acts represented indirect federal involvement in air pollution policy.

Direct federal involvement began in 1967 with the passage of the Air Quality Act. This act required states to use federal research on air pollutants to develop air quality standards for areas with major air-pollution concerns.^[1] Unfortunately, many states did not develop air quality standards by the time the act expired in 1970.^[1] This set the stage for the 1970 amendments to the Clean Air Act; what many refer to when speaking of the Clean Air Act. Under these amendments, the federal government established national ambient air quality standards (NAAQS) for major air pollutants from stationary sources, states were given responsibility for implementing air pollution plans within federal guidelines, and both the federal and state governments enforced these plans on industrial polluters.^[2] These NAAQS represented the maximum concentration of common air pollutants permitted and were divided into primary standards protecting human health and secondary standards protecting forests, agricultural products, and buildings.^[3]

Further amendments to the Clean Air occurred in 1977 and 1990. Most significantly, the 1977 amendments formally added the prevention of increased air pollution in areas such as national parks or forests having minimal air pollution concentrations below NAAQS.^[4] The 1990 amendments included provisions that established new emission standards for mobile sources of air pollution, phased out certain ozone depleting chemicals, and created a program of emissions trading for sulfur oxides.^[2] Of particular interest, policymakers implemented the sulfur oxide trading program as a way to reduce command and control regulation of industry. Although the federal government was still actively involved in this program, it represented a move back to indirect federal involvement in clean air policy.

In 1955, federal clean air policy began as a program of grants to states for research and training. However, by 1970 federal clean air policy had evolved to a command and control system attempting to reduce certain types of air pollutants. This evolution of

environmental policy occurred because of increasing public concern over environmental issues and an increasing awareness by the federal government that most states would not implement air pollution standards without compulsion from the federal government.^[2] By 1990, political influences forced parts of this direct federal approach back to a more indirect approach, yet command and control regulations still constituted the majority of federal air pollution policy. A similar progression of federal involvement in environmental policies can also be observed in clean water policy.

CLEAN WATER POLICY

Federal involvement in controlling water pollution technically commenced with the Refuse Act of 1899. In an attempt to ensure navigable waterways, this act prohibited the discharge of waste such as sludge and sawdust into navigable waterways without a federal permit.^[5] However, federal legislation emphasizing the improvement of water quality for purposes other than navigation began with the Water Pollution Control Act of 1948. In keeping with the prevailing view of the time, this act emphasized indirect federal involvement in controlling water pollution. Specifically, the act gave the federal government authority to conduct research on water pollution and to establish a loan program for municipalities wishing to construct municipal sewage treatment facilities.^[5] This act was amended in 1956. The 1956 amendments created a federal grant program for municipalities constructing sewage treatment facilities and established the concept of holding conferences to determine who should clean up interstate water pollution.^[5] However, because these statutes did not implement water quality or effluent standards and because the conference concept proved to be unworkable, these statutes were unsuccessful in controlling water pollution.

The Water Quality Act of 1965 had water quality standards. In this statute, states were required to develop water quality standards for interstate waters within state borders and determine the maximum discharge allowed into these interstate waterways.^[3] However, as in controlling air pollution, states did not consistently implement or enforce these standards. Consequently, direct federal involvement in water pollution policy came in 1972 with the Federal Water Pollution Control Act. This act established federal standards for point source (i.e., industrial or commercial) discharges of conventional water pollutants based on available technology, gave the federal government initial responsibility for enforcing those standards on individual polluters within a certain time period (states

could take over this enforcement duty at a later time), and expanded the federal grant program for municipalities constructing sewage treatment facilities.^[6]

Further changes to clean water policy occurred in 1977 when Congress passed the Clean Water Act. This act delayed some of the deadlines included in the 1972 statute for establishing technology-based effluent standards for individual dischargers and added effluent limits on toxic pollutants.^[7] Changes occurred again when Congress passed the Water Quality act in 1987. This act further postponed deadlines for establishing technology-based effluent standards, required states to implement plans for controlling non-point sources of water pollution such as runoff from agricultural land and urban areas, and reduced the amount of federal aid available to municipalities constructing sewage treatment facilities.^[5]

Over time, federal involvement in water pollution increased because of state inconsistencies in controlling water pollution. However, by the late 1980s the prevailing political environment contributed to a decrease in federal involvement in clean water policy. This decreased federal involvement resulted in a reduced federal grant program for the construction of municipality sewage treatment facilities and the continued delay of establishing technology-based effluent standards for industry and businesses. During this time, policymakers started to advocate approaches to pollution control in which the organization voluntarily assessed environmental impacts. Ironically, policymakers first introduced the concept of organizations assessing environmental impact in the late 1960s with the NEPA.

THE NATIONAL ENVIRONMENTAL POLICY ACT

In contrast to the extended histories surrounding the Clean Air Act and the Clean Water Act, policymakers designed NEPA legislation within one year and passed it into law at the beginning of 1970. The most significant provisions of NEPA required federal agencies to assess the environmental impact of programs before implementation and submit environmental impact statements for programs impacting the environment.^[8] An important provision of NEPA was increased citizen participation. Under the NEPA process, community groups could challenge the adequacy of an agency's initial environmental impact statement.^[9] In response to these challenges, agencies often changed program designs or designed final environmental impact statements reflecting citizen concerns.

While focusing on federal agencies, NEPA introduced the concept of assessing environmental performance at the organizational level. This organizational

idea of assessing environmental performance later became an emphasis of environmental management systems in the 1980s and 1990s. In environmental management systems, organizations voluntarily measure significant environmental impacts due to operations.^[10] Consequently, organizations can use environmental management systems to ensure compliance with current federal regulations regarding pollutant discharges and also to focus on environmental issues not currently regulated such as the sustainability of natural resources.

Environmental management systems allow the organization to focus on improving environmental performance. Although the success of environmental management systems in decreasing environmental impacts has still not entirely been proven, the voluntary standard adopted by organizations using these systems represents a further reduction in federal involvement in environmental policy. As environmental policy progressed through the 1980s and 1990s this trend of reduced federal involvement became increasingly apparent in statutes dealing with clean air and clean water.

CONCLUSION

During the 1960s and 1970s, environmental policy at the federal level shifted from a system of indirect federal management regarding such issues as air and water pollution to a more direct system emphasizing command and control regulation. Generally, these command and control regulations focused on industrial causes to pollution and required the federal government to define pollution standards while state governments designed plans to achieve those standards. The passage of legislation such as the 1970 amendments to the original Clean Air Act and the 1972 Federal Water Pollution Control Act exemplified this command and control process. These acts, along with their numerous amendments through the years, resulted in a dramatic increase in air and water quality.^[11]

While command and control regulations improved air and water quality, other environmental problems such as the depletion of natural resources and non-industrial sources of pollution persisted. To address these types of problems falling outside of the command and control process, environmental policy in the 1980s and 1990s increasingly emphasized strategies of pollution prevention rather than pollution reduction.^[12] Reflecting political concerns over the economic costs of direct federal involvement and a technical challenge of establishing national standards for less common pollutants, these pollution prevention strategies shifted federal environmental policy back to a more indirect

involvement. Typically, this indirect strategy relied on businesses and governmental agencies to engage in cooperative frameworks that often adopted voluntary standards of pollution prevention.^[12] In addition, this strategy also emphasized the notion of businesses adopting environmental management systems, often independent of any governmental incentive, in order to increase efficiency by decreasing negative environmental impacts.^[10]

These strategies of direct and indirect federal involvement currently comprise environmental policy. Command and control regulations implemented in the 1960s and 1970s operate with more voluntary standards emphasized in the 1980s and 1990s in an attempt to improve environmental quality. When operating together, both strategies may significantly improve many aspects of the environment. Nevertheless, the historical pattern of environmental policy development suggests that federal involvement in environmental policy will continue to vary to reflect new political trends and emerging technical challenges.

ACKNOWLEDGMENTS

I would like to thank Dr. Paul Culhane of Northern Illinois University and Dr. Patrick Stewart of Arkansas State University for providing thoughtful guidance on environmental policy.

REFERENCES

1. Portney, P. Air pollution policy. In *Public Policies for Environmental Protection*, 2nd Ed.; Portney, P., Stavins, R., Eds.; Resources for the Future: Washington, 2000; 77–123.
2. Rosenbaum, W. *Environmental Politics and Policy*, 5th Ed.; CQ Press: Washington, 2002.
3. Ringquist, E. *Environmental Protection at the State Level: Politics and Progress in Controlling Pollution*; M.E. Sharpe, Inc.: New York, 1993.
4. Smith, Z. *The Environmental Policy Paradox*; Prentice Hall: Englewood Cliffs, NJ, 1992.
5. Freeman, A. Water pollution policy. In *Public Policies for Environmental Protection*, 2nd Ed.; Portney, P., Stavins, R., Eds.; Resources for the Future: Washington, 2000; 169–213.
6. O'Leary, R.; Durant, R.; Fiorino, D.; Weiland, P. *Managing for the Environment: Understanding the Legal, Organizational, and Policy Challenges*; Jossey-Bass Publishers: San Francisco, 1999.
7. Portney, K. *Controversial Issues in Environmental Policy: Science vs. Economics vs. Politics*; Sage Publications, Inc.: Newbury Park, CA, 1992.

8. Kraft, M.; Vig, N. Environmental policy from the 1970s to the twenty-first century. In *Environmental Policy: New Directions for the Twenty-First Century*, 5th Ed.; Vig, N., Kraft, M., Eds.; CQ Press: Washington, 2003; 1–33.
9. Rosenbaum, W. The bureaucracy and environmental policy. In *Environmental Politics and Policy: Theories and Evidence*, 2nd Ed.; Lester, P., Ed.; Duke University Press: Durham, NC, 1995; 206–241.
10. Coglianese, C.; Nash, J. Environmental management systems and the new policy agenda. In *Regulating from the Inside: Can Environmental Management Systems Achieve Policy Goals?* Coglianese, C., Nash, J., Eds.; Resources for the Future: Washington, 2001; 1–26.
11. Ringquist, E. Evaluating environmental policy outcomes. In *Environmental Politics and Policy: Theories and Evidence*, 2nd Ed.; Lester, P., Ed.; Duke University Press: Durham, NC, 1995; 303–327.
12. Press, D.; Mazmanian, D. Understanding the transition to a sustainable economy. In *Environmental Policy: New Directions for the Twenty-First Century*, 5th Ed.; Vig, N., Kraft, M., Eds.; CQ Press: Washington, 2003; 275–298.

Environmental Pollution Control, State Government Administration of

A. Hunter Bacot

*Department of Political Science, University of North Carolina at Charlotte,
Charlotte, North Carolina, U.S.A.*

INTRODUCTION

Since 1980, state governments have assumed much responsibility for environmental program administration. Administrative responsibilities of states for pollution control policies are crucial to national environmental initiatives to prevent or contain pollution. Therefore, the management of pollution control policies is conducted primarily by states, and environmental affairs continue to be national priorities that rely on states for administration and implementation.

Environmental administration in the United States is generally distinguished according to media—air, water, and waste are the three generally accepted environmental arenas. Concerning pollution control, states enjoy significant latitude for monitoring and enforcement of air and water pollution control policies. National air and water regulations delegate administrative responsibilities to states based on each state's ability to perform these regulatory functions.^a State environmental agencies given responsibility for implementing administrative and enforcement provisions of national air and water policies have primacy [i.e., the state has primary responsibility for environmental management in the state, although the Environmental Protection Agency (EPA) oversees state management]. To date, nearly every state has full or partial authority from the EPA to control air and water pollution within their respective borders.

Three national legislative efforts (and subsequent reauthorizations)—the Clean Air Act (CAA), the Clean Water Act (CWA), and the Safe Drinking Water Act (SDWA)—provide states with much of the responsibility for regulating air and water pollution. This delegation of authority is examined in this article. Major provisions of air and water legislation, the degree to which states assume control of the administration of these legislative acts, and how well states fare in controlling pollution are also reviewed.

DELEGATING REGULATORY AUTHORITY TO STATES FOR POLLUTION CONTROL

States having primacy over environmental policies and programs must meet criteria established by the EPA to manage pollution control within their borders. The EPA has ultimate authority over all pollution control programs in the United States, and thus approves whether states are able to manage their own environmental situations. To acquire primacy status from the EPA, states must adhere to guidelines of the EPA (as stated in acts governing pollution) and receive the agency's approval for managing pollution control programs.

The CAA, CWA, and SDWA provide significant administration and implementation responsibilities to states for air and water pollution control. To demonstrate the capacity to enforce national laws within their borders, states initially submit required biennial reports to the EPA discussing state environmental laws, which cannot be less stringent than the national laws, and how these laws support national environmental legislation. States must demonstrate to the EPA the ability to accept many of the permitting, administrative, and enforcement provisions contained in national laws.

To demonstrate an ability to accept environmental management responsibilities (i.e., to attain primacy), states must submit state implementation plans to the EPA. State implementation plans explain to the EPA how state regulations serve national interests with regard to environmental pollution control. State implementation plans are involved, providing detailed explanations about a state's ability to oversee environmental management programs and policies within its borders. They include information about pollution control legislation, state agency information, due process procedures (hearings, comment periods, appeals, etc.), enforcement capabilities, etc. The EPA must approve the state implementation plan, or it can authorize the state to administer environmental policies on a limited or partial arrangement until approval of the state's state implementation plan is granted. Full authorization or delegation gives states full control of environmental programmatic responsibilities. In the year 2002, most states had primacy for most major environmental programs (see [Table 1](#)).

^aIn addition to the public laws detailing these responsibilities, much of the following discussion draws generally from a host of excellent sources discussing the administration of environmental programs (see Refs. [1–11]).

Table 1 State pollution control status for selected air and water acts

Primacy status	Selected air acts (number of states)					Selected water acts (number of states)		Selected drinking water acts (number of states)	
	NSPS	NES HAPS	PSD	Title V	NSR	NPDES	P	PWSS	WPP
Full program delegation or authorization	40	39	14	17	0	38	31	49	15
Partial program delegation or authorization	8	9	3	0	1	4	0	0	0
Program delegation or authorization is approved, pending, or in interim status	1	2	33	28	43	0	1	0	21
Not delegated or not subject to delegation	1	0	0	5	6	8	18	1	14

Notes: For the CAA: State-delegated responsibilities for the CAA includes state authority to administer the following components of the act, and include the New Source Performance Standards (NSPS; deals with emission standards performance testing and compliance), National Emission Standards for Hazardous Air Pollutants (HAPS), Prevention of Significant Deterioration (PSD; ensures that areas subject to provisions of the National Ambient Air Quality Standards are not declining), Title V Operating Permit authority, New Source Review (ensures continued compliance with air standards according to air standards permitting base on best available control technology). For the CWA: State-delegated administrative responsibilities for the CWA includes state authority to administer the National Pollution Discharge Elimination System (NPDES; governs point source discharges) and pretreatment programs for publicly owned treatment works facilities (P; governs effluents treated at treatment plants before discharge, and directed at non-point source treatment). For the SDWA: Safe Drinking Water Act (SDWA; governs drinking water quality), which is a separate act from the CWA, has delegable components for ensuring water quality for consumption; here we are interested in ensuring safe drinking water and water supply, which are specified in the SDWA as the Public Water System Supervision (PWSS; publicly operated water systems) and the Wellhead Protection Program (WPP; protects public water supply wells from contaminants).

Source: Author, Refs.^[19] and ^[20].

Upon acquiring authorization to administer their own programs, states and the EPA remain partners in pollution control. The EPA supports states through the administration of millions of dollars in grants, and provides additional assistance to states through research, technology, and technical support. If states fail to adhere to pollution control strategies stated in their state implementation plans, the EPA can rescind state authorization for administering environmental programs (which rarely occurs). Thus, although most states currently administer their “own” state environmental programs, continuing administration of national initiatives depends on a state’s administrative capacity, as so deemed by the EPA, to accept this responsibility.

AIR POLLUTION CONTROL REGULATIONS

Provisions in the CAA, which was originally passed in 1972 and subsequently amended and reauthorized, guide cooperation on policies affecting air quality in the United States. The CAA initially regulated air quality standards of pollutants. CAA amendments furthered air regulations to include air toxins, acid rain, and local air quality attainment, which expanded the agency’s responsibility for improving air quality.

The original CAA listed only a few contaminants and established national emissions standards for hazardous air pollutants (NESHAP). The NESHAP increased air pollution regulations by addressing commercial and industrial air pollutants (e.g., asbestos, mercury, benzene).

In 1990, subsequent amendments to the CAA increased the roles of states; provisions of the act mandate state participation in the administration of environmental programs. The amendments also require states to prepare state implementation plans (SIPs), which force states to submit to the EPA an “inventory” of their progress in reducing air pollution. Congress also increased the number of regulated pollutants, which, because most states have attained primacy status, gives states additional regulatory and administrative responsibility for air pollution. The amendments also identify new areas of regulation to control releases contributing to acid rain and ozone depletion (among others), define urban non-attainment areas, and identify a host of contaminants, as well as acceptable emission levels for each contaminant.

Generally, states are given reasonable responsibility for establishing acceptable emission levels (in their SIPs). States must account for their performance by tracking emission levels of chemical sources, as well as other new sources identified through increased technological capacity. New emission regulations

determine areas that have poor air quality, as well as areas that meet ambient air recommendations (which are established using technologically driven analysis based on “best available technology” or other technological criteria). As technological capacity constantly reveals previously undetectable elements, states that meet their specified standards must continually monitor the status of air pollutants, which is noted in biennial pollution reports required by the EPA.

As is evident in testing standards and monitoring responsibilities, states’ assumption of programmatic responsibility for air quality is an arduous assignment. Most states have prospered with program primacy because only a few do not have some program responsibilities delegated (see Table 1). Environmental management of air quality is a continuous, challenging task that is but one of the primary environmental media demanding attention by state environmental agencies.

WATER POLLUTION CONTROL REGULATIONS

An equally demanding area of emphasis for state environmental administrators is surface water quality (e.g., rivers, lakes) and underground water quality (e.g., wellhead protection, aquifers). Like air quality, water quality standards are guided by national legislation—the CWA and its subsequent amendments, and the SDWA—and are generally implemented at the state level. Improved water quality is realized by regulating discharges from industrial, commercial, and municipal (ICM) facilities, as well as non-point sources and stormwater systems.

Water quality pollution control is managed by controlling effluents discharged into regulated waters and protecting drinking water sources. Based on treatment standards and effluent lists, states create ambient water quality standards for ICM facilities that discharge conventional pollutants and non-conventional pollutants into surface waters. Conventional pollutants are discharges that deplete or disturb natural elements and/or the balance of elements, whereas non-conventional pollutants are airborne or leachate pollutants that originate elsewhere and migrate or settle in bodies of water and/or area water supplies.

Similar to most laws regulating pollutants, sections of the CWA provide specific guidelines to the EPA for managing water quality and returning programmatic responsibilities to states. States, with EPA oversight, implement most of the provisions set forth in the CWA. Also specified in the CWA are water quality standards based on the type and use of waters—whether recreational, sporting, or serving as a water

supply, etc.^b States classify bodies of water in accordance with standards required for maintaining the classified use. Common classifications of water types and uses organize waters based on exceedances of the specified minimums for stated categories, according to national definitions, as well as designate certain pristine waters as protected.

Drinking water pollution control is another area of significant responsibility for state agencies, and is governed by the SDWA. Drinking water pollution is caused by the contamination of water supplies. Effluents migrating from offsite locations, whether to an aquifer, to wellheads, or generally into the water table, contaminate water supplies. States protect water supplies by establishing standards (based on national standards), distinguishing among source uses (agriculture, drinking water, etc.), and controlling non-point threats. Determination of water quality regulations becomes states’ responsibility, as water sources, uses, and threats vary widely across states.

The CWA amendments give states responsibility of the permitting process for the National Pollutant Discharge Elimination System (NPDES). The EPA maintained responsibility of pollution discharges until 1987, when responsibility was delegated to states for establishing effluent levels and granting discharge permits. The NPDES regulates direct discharges, or point source discharges into surface waters. National standards guide states in determining allowable discharges and the water quality of water bodies that accept discharges. Facilities discharging effluents into surface waters must have a permit from the state, which grants these facilities permission to discharge effluents into affected surface waters within allowable limits (as established by the EPA).

Over time, states have demonstrated the ability to manage water quality programs. In fact, approximately one-half of the 50 states administer comprehensive water protection programs. Although difficult, states are assuming responsibilities for administering programs to protect and preserve water quality.

STATE PROGRESS IN AIR AND WATER POLLUTION CONTROL

From most perspectives, states have succeeded in controlling air and water pollution in the United States. Yet, despite the successes of environmental

^bCategories are classified based on water quality that supports recreational uses by people (class A); ecological integrity and natural habitation (class B); water supplies for public consumption (class C); and industrial, commercial, and agricultural activities (not intended for consumption) (class D).

management in the United States, as Rosenbaum observed, “the nation’s air and water remain seriously polluted.”^[12] Without legislation and policies directed at controlling air and water pollution, however, the situation would likely be much worse. Thus, overall, states’ efforts to reduce pollution in our air and water systems are deficient, and although there are improvements in curbing air and water pollution, most improvements are based on seriously limited assessment information.

Under the direction and guidance of the EPA, states have shown progress in cleaning air, water, and drinking water supplies since the 1980s. Air pollution policies enforced by states reveal improvement in air quality throughout the United States. This success in managing pollution is achieved despite regulations governing approximately 200 pollutants, which is a dramatic increase in the number of regulated elements from the original legislation enacted in 1970. Of the six criteria pollutants,^c each has decreased from 1989 levels. In fact, four of these six pollutants decreased between 25% and 56% (Ref.^[13]; Table A-9). There are other areas, namely ozone, in which states continue to encounter difficulty in achieving established reduction goals; another area of special importance to states is air quality in rural areas, which is lagging compared with established goals.^[13]

Water quality trends have improved dramatically since 1970, with the enactment of the CWA and subsequent amendments to the act. Improvements in water quality are based on available information, which is woefully inadequate as only approximately one-third of all water bodies have been assessed to determine pollution levels. For example, only 23% of rivers, 42% of lakes, and 32% of estuaries have been assessed to determine surface water quality in the United States.^[14] Of water bodies assessed by states, however, there are indications of improvements in water quality. In fact, “states reported that 65 percent of assessed river and stream miles, 55 percent of assessed lake acres, and 56 percent of assessed estuarine square miles fully support the water quality standards states evaluated.”^[14] As a result, water quality is improving across the states, although this judgment is based on limited information.

Drinking water and groundwater are two other areas involving water quality assessments in states. Although states find overall groundwater quality acceptable, there are measurable problems with contamination from nearby sources, such as underground storage tanks, landfills, septic tanks, etc.^[14] Based on state reports, groundwater quality has improved in

recent years and continues to support major uses of groundwater resources.^[15] As tremendous strides in assessing groundwaters are made, significant deficiencies in coverage remain.^[15] Safe drinking water is probably the success story for states aspiring to improve water quality. Although challenged to provide safe drinking water to all citizens, states—armed with provisions of the SDWA—have improved the quality of drinking water since the 1980s.^[16] As improvements are made in the “testing, treatment, protection, and provision of drinking water to the public,” challenges remain as technology continues to advance the ability to track and detect elements in smaller quantities.^[16]

STATE PROGRESS IN MONITORING AND ENFORCEMENT OF POLLUTION CONTROL POLICIES

Although states appear to be making progress in controlling pollution within their borders, the overall picture of states’ progress in this area is formidable. States made significant progress in controlling pollution on identified sources or assessed sources. However, because identified or assessed sources comprise only a small percentage of pollution threats, much is not assessed or identified by state agencies charged with managing pollution. It is information about “the unknown” that should concern citizens of each state.

According to the EPA’s Office of Inspector General, “air enforcement audits disclosed fundamental weaknesses with state identification and reporting of significant violators of the Clean Air Act.”^[17] Similar findings plague efforts to control water pollution. Another EPA report submitted “that state enforcement programs could be much more effective in deterring non-compliance with discharge permits and, ultimately, improving the quality of the nation’s water.”^[18] States lag significantly in addressing pollution control problems for all contaminant threats to the nation’s air and water. Analysis of state efforts in air and water program enforcement and monitoring suggests states are woefully deficient and, if advances in controlling pollution are to occur, states must improve their efforts, or pollution sources will remain unchecked.

States, therefore, must improve their pollution enforcement and monitoring programs, as well as assessment programs affecting all water bodies and air quality. However, the within-state challenges most states face make improvement extremely difficult and, as a result, unlikely. Most states have improved their pollution control programs dramatically since the 1980s, but these improvements have not kept pace with the increasing program demands required with program primacy. Pollution control for states is proving a

^cCriteria pollutants are carbon monoxide, lead, nitrogen dioxide, ozone, particulate matter, and sulfur dioxide.

Sisyphian task that, despite ample progress, has states struggling to keep up with program requirements.

CONCLUSION

As states continue to cite gains in the quality of air and water by acquiring data necessary to establish current dispositions, more must be known about the true threat of pollution; this is particularly important as information continues to improve knowledge of states' environmental situations. States are continually increasing their capacity to monitor air and water quality, but advances are slow, deliberate, and costly, which imposes considerably on state resources. Accordingly, although states show some success in mitigating pollution, as the analytical capabilities continue to improve with science and technology, states will continually confront new areas of environmental need that commands their attention. More troublesome for states (and citizens) are that many problems remain undetected, and thus may not be recognized until conditions deteriorate to a level unable to support intended uses.^[15] As a result, there can be no celebration of state environmental successes, only brief recognition of these successes. The battle to control air and water pollution is daunting and progress in assessment, monitoring, and enforcement must improve dramatically if pollution is to be curtailed and contained.

REFERENCES

1. Advisory Commission on Intergovernmental Relations. *Intergovernmental Decisionmaking for Environmental Protection and Public Works*; USACIR: Washington, DC, 1992; (A-122).
2. Brownell, F.W. Clean air act. In *Environmental Law Handbook*, 14th Ed.; Sullivan, T.F.P., Ed.; Government Institutes: Rockville, MD, 1997; 72–108.
3. Council of State Governments. *Resource Guide to State Environmental Management*; Council of State Governments: Lexington, KY, 1996.
4. Gallagher, L.M. Clean water act. In *Environmental Law Handbook*, 14th Ed.; Sullivan, T.F.P., Ed.; Government Institutes: Rockville, MD, 1997; 109–160.
5. Halbleib, W.T. Emergency planning and community right-to-know act. In *Environmental Law Handbook*, 14th Ed.; Sullivan, T.F.P., Ed.; Government Institutes: Rockville, MD, 1997; 481–509.
6. Mangun, W.R. Environmental program evaluation in an intergovernmental context. In *Environmental Program Evaluation: A Primer*; Knapp, G.J., Kim, T.J., Eds.; University of Illinois Press: Urbana, IL, 1998.
7. Nardi, K.J. Underground storage tanks. In *Environmental Law Handbook*, 14th Ed.; Sullivan, T.F.P., Ed.; Government Institutes: Rockville, MD, 1997; 360–384.
8. Rabe, B.G. Power to the states: the promise and pitfalls of decentralization. In *Environmental Policies in the 1990s: Reform or Reaction*; Vig, N.J., Kraft, M.E., Eds.; CQ Press: Washington, DC, 1997; 31–52.
9. Scagnelli, J.M. Pollution prevention act. In *Environmental Law Handbook*, 14th Ed.; Sullivan, T.F.P., Ed.; Government Institutes: Rockville, MD, 1997; 510–521.
10. Scheberle, D. *Federalism and Environmental Policy: Trust and the Politics of Implementation*; Georgetown University Press: Washington, DC, 1997.
11. Williams, S.E. Safe drinking water act. In *Environmental Law Handbook*, 14th Ed.; Sullivan, T.F.P., Ed.; Government Institutes: Rockville, MD, 1997; 196–225.
12. Rosenbaum, W.A. *Environmental Politics and Policy*; CQ Press: Washington, DC, 2002.
13. US EPA. Office of Air Quality Natl. Air Qual. Emiss. Trends Rep., 1999; (EPA 454/R-01-004, March).
14. US EPA. Office of Water. *Water Quality Conditions in the United States: A Profile from the 1998 National Water Quality Inventory Report to Congress*; 2000; ES-3(EPA-841-F-00-006, June).
15. US EPA. Office of Water. *Environmental Indicators of Water Quality in the United States*; 1996; (EPA 841-R-96-002, June).
16. US EPA. Office of Water. *25 Years of the Safe Drinking Water Act: History and Trends*; 1999; 34 (EPA 816-R-99-007, December).
17. US EPA. Office of Inspector General. *Audit Report Consolidated Report on OECA's Oversight of Regional and State Air Enforcement Programs*; 1998; i (E1GAE7-03-0045-8100244, September).
18. US EPA. Office of Inspector General. *Audit Report Water Enforcement: State Enforcement of Clean Water Act Dischargers Can Be More Effective*; 2001; i (Report No. 2001-P-00013, August).
19. Environmental Council of the States, <http://www.sso.org/ecos/> (accessed July/August 2001).
20. Local Government Environmental Assistance Network, <http://www.lgean.org/html/stateregs.cfm> (accessed July/August 2001).

Equal Employment Opportunity and Affirmative Action

Norma M. Riccucci

*School of Public Affairs and Administration, Rutgers University, Newark Campus,
Newark, New Jersey, U.S.A.*

INTRODUCTION

Equal employment opportunity (EEO) is largely viewed as a means to prevent discrimination in the workplace. Title VII of the Civil Rights Act of 1964, for example, as amended, is intended to prevent discrimination on the basis of race, color, religion, gender, and national origin in public and private sector workforces. Affirmative action, however, which emerged in response to pervasive employment discrimination, embodies proactive efforts to redress past discriminations and to diversify the workplace in terms of race, ethnicity, gender, age, physical abilities, and so forth. Affirmative action has been viewed as a legal tool to ensure EEO or diversity (see [Table 1](#) for comparison of the concepts).

EMPLOYMENT DISCRIMINATION AND THE EVOLUTION OF EEO

Although there were a handful of early efforts to combat employment discrimination in our society, it was not until the 1960s that this nation, spurred by racial unrest, was galvanized into taking genuine action to end discriminatory employment practices. In 1964, Title VII of the Civil Rights Act, which is the cornerstone of EEO law in the United States, was passed. Title VII was intended to prevent private sector employers, with 15 or more employees, from discriminating on the basis of race, color, religion, gender, and national origin.^[1] Today, there are numerous statutes, executive orders, and policies in place that are intended to combat employment discrimination in the workplace. [Table 2](#) illustrates the major EEO laws and policies that have been advanced since the reconstruction era. As can be seen, Title VII of the Civil Rights Act was extended to public employers with passage of the Equal Employment Opportunity Act in 1972. It is this act, along with provisions of the U.S. Constitution (e.g., the Fourteenth Amendment) and its enforcing legislation (Sections 1981 and 1983 of the Civil Rights Acts of 1866 and 1871, respectively), that is the primary legal mechanism geared to protect women and people of color from employment discrimination in the public sector.^[2]

As with race, ethnicity, and gender, employers are also prohibited from discriminating against persons on numerous other dimensions or characteristics, including age and ability. For example, the ADEA of 1967 makes it illegal for private businesses to refuse to hire, to discharge, or to otherwise discriminate against an individual between the ages of 40 and 70. The act was amended in 1974 to apply to state and local governments. Age discrimination in federal employment is prohibited by executive order, but additional amendments to the ADEA banned forced retirement for federal employees at any age. Enforcement authority over the ADEA was originally vested in the Department of Labor, but was transferred to the Equal Employment Opportunity Commission as part of the federal civil service reform of 1978.

There are also various laws that prohibit discrimination on the basis of a disability. For instance, the Vocational Rehabilitation Act of 1973, as amended, prohibits discrimination against persons with disabilities. These protections were strengthened by Title I of the ADA of 1990, which covers private sector employers as well as state and local government employers with 15 or more employees. The federal workforce, which is not protected by the ADA, continues to be covered by the Rehabilitation Act of 1973. An important aspect of the ADA is that it covers persons who are HIV positive or have AIDS.

FROM EEO TO AFFIRMATIVE ACTION

EEO law is considered to be passive in the sense that it requires employers to refrain from discriminating against protected-class members (i.e., those designated for protection by the specific EEO legislation). Affirmative action, however, requires employers to take positive steps toward employing, promoting, and retaining qualified women, people of color, and other protected-class persons (see description of Executive Order 11246, as amended, in [Table 2](#)). Not only would this help to rectify past and present discrimination, but also would help to create “representative bureaucracies”—that is, government bureaucracies that are demographically representative of the general populations they serve. Because affirmative action is based on proactive

Table 1 Comparing EEO and affirmative action

EEO	Affirmative action
<i>Qualitative/quantitative</i> : emphasis on preventing or ending discrimination	<i>Qualitative/qualitative</i> : emphasis on redressing past discrimination and achieving diverse, representative workforces
<i>Legally driven</i> : Mandated by federal law	<i>Managerially and legally driven</i> : involves voluntarily developed goals and court-ordered programs; common law has defined its legality and constitutionality
<i>Fairness</i> : seeks to end discrimination and create equal opportunities	<i>Remedial and compensatory</i> : specific target groups benefit as past and present wrongs are remedied
<i>Access model</i> : model assumes that protected-class persons will be able to access organizations	<i>Assimilation model</i> : assumes that persons and groups brought into the system will adapt to existing organizational norms; can result in “sink or swim” atmosphere/environment
<i>Level playing field</i> : seeks to ensure equal opportunity and access	<i>Opens doors</i> : seeks to affect hiring and promotion decisions in organizations

(Adapted from Ref.^[2].)

Table 2 Federal EEO law

Law/policy	Provisions/coverage
Civil Rights Act of 1866, Section 1981	Provides that “all persons shall have the Same right . . . to the full and equal benefit of the laws . . . as is enjoyed by white citizens”
Fourteenth Amendment to U.S. Constitution (1868)	Requires all states and their political subdivisions to provide equal protection of the laws to all persons in their jurisdictions
Civil Rights Act of 1871, Section 1983	Prohibits persons acting “under color of any statute, ordinance, regulation, custom or usage . . .” from depriving any citizen or person within the jurisdiction of the United States of any rights, privileges, or immunities secured by the Constitution
Ramspect Act 1940	Prohibits discrimination in federal employment based on race, color, or creed
Executive Order 8802 (1941)	Called for the elimination of discrimination based on race, color, religion, or national origin with the federal service and defense production industries
Civil Rights Act of 1964, Title VII	Prohibits discrimination on the basis of race, color, religion, and national origin
Executive Order 11246, as amended (1965)	Forbids employment discrimination based on race, color, religion, gender, and national origin by federal government and federal contractors and subcontractors and requires the federal government and contractors to engage in affirmative action to hire and promote persons based on these characteristics
Age Discrimination in Employment Act (ADEA), as amended (1967)	Forbids employment discrimination based on age
Equal Employment Opportunity Act of 1972 (amends Title VII of CRA of 1964)	Extends Title VII protection to state, local, and federal government employees and workers in educational institutions
Vocational Rehabilitation Act of 1973	Prohibits federal government and its contractors from discriminating against persons with disabilities
Vietnam Era Veterans’ Readjustment Act of 1974	Requires the federal government and its contractors to promote employment opportunities for Vietnam-era veterans
Americans with Disabilities Act (ADA) of 1990	Forbids private, state, and local government employers from discriminating on the basis of disability
Civil Rights Act of 1991	Overtured several negative U.S. Supreme Court decisions issued in 1989 on EEO and affirmative action and established a Glass Ceiling Commission to study the artificial barriers to the advancement of women and persons of color in the workplace

efforts, it has led to myriad lawsuits challenging its legality and constitutionality. In fact, one of the first legal challenges, *Regents of the University of California v. Bakke* (1978), argued that affirmative action resulted in “reverse discrimination.”^[3] In *Bakke*, the U.S. Supreme Court upheld, for the first time, the legality of an affirmative action program, when it ruled in favor of a flexible admission program of the University’s Medical School, which took race into account as a criterion for admissions.

A Snapshot of Affirmative Action Case Law

Since the U.S. Supreme Court issued its *Bakke* decision in 1978, it has upheld affirmative action in various forms, at least through 1989 (see Table 3). Beginning in that year, the High Court issued several decisions, which proved detrimental to both EEO and affirmative action efforts. In response to these regressive rulings, the U.S. Congress passed the Civil Rights Act of 1991, which reversed these rulings, restoring the use

Table 3 Key U.S. Supreme Court actions or decisions on affirmative action

1978	<i>Regents of the University of California v. Bakke</i> . U.S. Supreme Court upholds the principle of affirmative action but strikes down its operation by the University at California under the Fourteenth Amendment and Title VI of the Civil Rights Act of 1964
1979	<i>United Steelworkers of America v. Weber</i> . U.S. Supreme Court upholds legality of voluntarily developed affirmative action plan under Title VII of Civil Rights Act of 1964
1980	<i>Fullilove v. Klutznick</i> . U.S. Supreme Court upholds constitutionality (under Fifth and Fourteenth Amendments) of federal set-aside programs enacted by the U.S. Congress
1984	<i>Firefighters Local Union and Memphis Fire Department v. Stotts</i> . U.S. Supreme Court upholds, under Title VII of the Civil Rights Act, as amended, the use of a seniority system in layoff decisions, despite its negative impact on affirmative action
1986	<i>Wygant v. Jackson Bd. of Ed.</i> U.S. Supreme Court strikes down, under the Fourteenth Amendment to the U.S. Constitution, the use of affirmative action in layoff decisions
1986	<i>Sheet Metal Workers’ International Association v. EEOC</i> . U.S. Supreme Court upholds, under Title VII and Fifth Amendment to the U.S. Constitution, a court-ordered affirmative action program to remedy past discrimination by a union and apprenticeship committee against people of color
1986	<i>Int’l Assoc. of Firefighters v. City of Cleveland</i> . U.S. Supreme Court upholds, under Title VII, affirmative action consent decree that provided for the use of race-conscious relief in promotion decisions
1987	<i>Johnson v. Transportation Agency, Santa Clara County</i> . U.S. Supreme Court upholds, under Title VII, voluntarily developed affirmative action program intended to correct gender and racial imbalances in traditionally segregated job categories
1987	<i>U.S. v. Paradise</i> . U.S. Supreme Court upholds, under the Fourteenth Amendment to the U.S. Constitution, a court-ordered affirmative action plan aimed at remedying discrimination against African-Americans in hiring and promotion decisions in Alabama Public Safety Department
1989	<i>City of Richmond v. Croson</i> . U.S. Supreme Court strikes down the constitutionality, under the Fourteenth Amendment, of a local government’s set-aside program because it could not satisfy the criteria of the strict scrutiny test
1989	<i>Martin v. Wilks</i> . U.S. Supreme Court allowed White firefighters to challenge, under Title VII, a consent decree, to which they were not a party, years after it had been approved by a lower court
1995	<i>Adarand v. Peña</i> . U.S. Supreme Court rules that the Equal Protection Clause of the Fifth Amendment requires that racial classifications used in federal set-aside programs must undergo strict scrutiny analysis
1996	<i>Hopwood v. State of Texas</i> . U.S. Supreme Court let stand a ruling by the U.S. Court of Appeals for the Fifth Circuit, which struck down the constitutionality of an affirmative action program at the University of Texas Law School
1999	<i>Lesage v. Texas</i> . U.S. Supreme Court throws out a reverse discrimination suit filed under the Equal Protection Clause of the Fourteenth Amendment against the University of Texas’ Department of Education
2003	<i>Grutter v. Bollinger</i> . The U.S. Supreme Court upheld affirmative action at the University of Michigan’s Law School under the Fourteenth Amendment, arguing that there is a compelling state interest in “racial diversity”. Race was one factor considered among many other factors
2003	<i>Gratz v. Bollinger</i> . The U.S. Supreme Court struck down the use of affirmative action in the University of Michigan’s undergraduate program under the Equal Protection Clause of the Fourteenth Amendment and Title VI of the Civil Rights Act. The Court stated that the program was based on a formula or “quota” system, giving extra points for race

of EEO and affirmative action to their pre-1989 legal status (see Table 4). In 2003, the U.S. Supreme Court again issued a landmark ruling in *Grutter v. Bollinger*, which upheld the use of affirmative action.^[4]

The *Grutter v. Bollinger* case involved the constitutionality of an affirmative action program at the University of Michigan's Law School. In a 5-4 ruling, the court majority opined that the racial diversity of a study body can be a sufficiently compelling interest on the part of a state university to warrant its use of racial preference in its admissions decisions. The court ruled that the Equal Protection Clause allows for the "Law School's narrowly tailored use of race in admissions decisions to further a compelling interest in obtaining the educational benefits that flow from a diverse student body" (on page 4 of Ref.^[4]).

In a second case, *Gratz v. Bollinger*, the Supreme Court, in a 6-3 decision, struck down the use of affirmative action in admission decisions at the University of Michigan's undergraduate programs in the College of Literature, Sciences, and Art.^[5] This program awarded 20 points on a scale of 150 for membership in an "underrepresented minority group"—African-Americans, Latinos, and American-Indians—with 100 points guaranteeing admission. In ruling that the

program violated the Equal Protection Clause of the Fourteenth Amendment, the court ruled that the program was "not narrowly tailored to achieve respondents' asserted compelling interest in diversity . . . [and] the policy does not provide . . . individualized consideration" of each characteristic of a particular applicant (on page 246 of Ref.^[5]).

In short, the court would not accept an affirmative action program in the form of numerical ratings. Thus, if affirmative action programs serve a compelling state interest (e.g., to promote diversity) and are sufficiently narrowly tailored (e.g., are not based on point systems) the court is likely to uphold their use.

THE FUTURE OF EEO AND AFFIRMATIVE ACTION

Although it is unlikely that EEO laws would be repealed at this point in time, the degree to which courts are willing to enforce the laws can vary. Indeed, as the U.S. Supreme Court continues to promote its policy of "federalism," whereby the court upholds the sovereignty rights of the states, protecting them from unnecessary or unwarranted intrusion by the

Table 4 The Civil Rights Act of 1991 overturns the U.S. Supreme Court's 1989 decisions

Case	Supreme Court ruling	Civil Rights Act of 1991
<i>Wards cove packing co. v. Atonio</i> (Title VII)	Plaintiffs must show that adverse was intentional, i.e., that the employment practice resulting in disparities was used not out of business necessity, but for illegal reasons	Plaintiffs required to demonstrate adverse impact. Employer must then "demonstrate that the challenged practice is job-related for the position in question"
<i>Patterson v. Mclean Credit Union</i> (42 USC Section 1981)	42 USC Section 1981 of the Civil Rights Act of 1866 is limited to hiring and some promotion decisions but does not extend to harassment on the job, discriminatory firing, or other post-hiring conduct by the employer	42 USC Section 1981 of the Civil Rights Act of 1866 covers all forms of racial bias in employment
<i>Martin v. Wilks</i> (Title VII)	White male firefighters can challenge, without time limitations, affirmative action consent decrees settling employment discrimination dispute, even if they were not original parties to the consent decree	The act prohibits challenges to consent decrees by individuals who had reasonable opportunities to object to the decrees
<i>Lorance v. At&t</i> (Title VII)	Women challenging the legality of a collective bargaining seniority system must file suit within the first 300 days of the system's adoption	Employees may challenge a seniority system when it affects them, as well as when the system is adopted
<i>Price Waterhouse v. Hopkins</i> (Title VII)	In "mixed-motive" cases (i.e., those where lawful and unlawful factors motivated the employment decision), the employer can avoid liability by demonstrating that the same action would have been taken without the discriminatory motive. The burden of proof rests with the employer	Any intentional discrimination is unlawful, even if the same action would have been taken in the absence of the discriminatory motive

(Adapted from Civil Rights Act of 1991. Text and analysis. *Employment guide*, Special supplement; Bureau of National Affairs: Washington, DC, November 11, 1991; Vol. 6, No. 23, 5.)

federal government, severe restrictions have been placed on state workers' ability to sue their state in federal court for federal EEO violations. In two recent cases, *Kimel v. Florida Board of Regents* and *Garrett v. the University of Alabama*, the U.S. Supreme Court, for the first time, held that the ADEA and the ADA, respectively, were inapplicable to the states.^[6,7] In effect, state workers can file suit against state governments, but only if the state agrees to be sued and only under the state's EEO law. These laws tend to be weaker, less effective, and narrower in scope. For example, state disability laws tend to be much less rigorous than the ADA around reasonable accommodations.

The future of affirmative action also rests with the courts. However, it is important to note that given the demographic shifts in the United States as well as public employers' continued efforts to develop diversity programs, public sector workforces are becoming more diverse. Thus, despite court rulings, employers

seem unwilling to abandon their efforts to diversify their workplaces.

REFERENCES

1. Rosenbloom, D.H. *Federal Equal Employment Opportunity*; Praeger Publishers: New York, 1977.
2. Riccucci, N.M. *Managing Diversity in Public Sector Workforces*; Westview Press: Boulder, 2002.
3. *Regents of the University of California v. Bakke*. 438 U.S. 265, 1978.
4. *Grutter v. Bollinger*. 539 U.S. 306, 2003.
5. *Gratz v. Bollinger*. 539 U.S. 244, 2003.
6. *Kimel v. Florida Board of Regents*. 528 U.S. 62, 2000.
7. *Garrett v. the University of Alabama*. 193 F.3d 1214, 11th Cir., 1999; cert. granted; *University of Alabama at Birmingham Bd. of Trustees v. Garrett*. 529 U.S. 1065, 2000; *Board of Trustees of the University of Alabama v. Garrett*. 531 US 356, 2001.

Ethics and Administrative Reform

Jeroen Maesschalck

*Instituut voor de overheid, Katholieke Universiteit Leuven,
Leuven, Belgium*

INTRODUCTION

Administrative or civil service reform has impacted the public sectors of many countries throughout the world since the 1980s. The reforms have included several strategies (maintain, modernize, marketize, minimize.^[1]), but most prominent were the changes that Hood assembled under the conceptual umbrella “New Public Management” (NPM). This includes the introduction of performance management systems, more responsibility and accountability for public managers, more competition in the public sector, the introduction of quality management techniques, etc.^[2,3]

The implementation of these reforms in governments throughout the world has sparked concern, both among academics and practitioners, about their impact upon the ethics of public servants. This article identifies four positions in this debate over the impact of the reforms on public servants’ ethics and concludes with the observation that in spite of their seeming rivalry, the empirical claims underlying these positions are in fact complementary.

ETHICAL CONCERNS ABOUT ADMINISTRATIVE REFORM

The apprehension for the potentially negative impact of New Public Management (NPM) reforms on the ethics of public servants has been strengthened by highly publicized scandals in many Organization for Economic Cooperation and Development (OECD) countries. Many of these scandals are about individuals behaving unethically, and often also illegally, and causing their public sector organization financial and other difficulties. A case in point is Robert Citron in Orange County, CA.^[4] Typically, these individuals were initially seen as exemplars or even heroes, who successfully managed to escape bureaucratic constraints in order to produce results and “get things done” until it was discovered that their escaping bureaucratic constraints also included manifest illegal behavior. It is argued that these excesses are caused by NPM reforms because the NPM-type rhetoric and reforms provided these individuals not only with the

opportunity to perform such behavior but also with the moral mindset to justify it.

NPM-type reforms are also alleged to lead to more collective or even systemic unethical behavior. A case in point are “creaming” strategies, a typical perverse effect of performance management systems. Suppose that the public servants who are in charge of vocational training of unemployed people would be assessed (and paid) on the basis of their “success rate,” defined as the number of applicants who are able to obtain a job following the training. It is not unlikely that these public servants would only allow the most job-ready applicants into their training courses. This would leave a considerable category of unemployed people without training, which is unlikely to be an intended effect of this policy.^a This is just one example of the ethically undesirable consequences that are alleged to be fostered by NPM reforms.

FOUR POINTS OF VIEW

The answer of public administration scholars to these scandals and to the more general concerns about the impact of administrative reforms on (un)ethical behavior of public servants is a discussion that can, in terms of Hood and Jackson, be conceived of as a “doctrinal debate.”^[5] Doctrines are “specific maxims about administrative whos, hows, and whats,” they “denote specific ideas about what should be done in administration.” Roughly summarized, four rival positions can be identified in this debate, each proposing their own administrative doctrines to foster ethical behavior in the public sector. The following paragraphs will summarize those positions. Although this overview will focus on the recent debate over the ethical impact of NPM reforms, it is in fact much broader because those against NPM reforms propose their own alternative, which is then in turn criticized by the other participants in the debate. Moreover, three of the four positions are

^aAlthough these creaming strategies are currently very relevant because of the popularity of performance management systems, they were already pointed out by Blau, P.M., 1955. *The dynamics of bureaucracy*. Chicago: University of Chicago Press.

rooted in a broader administrative philosophy, thus making this debate a contemporary version of older debates.^b

New Public Management

The first position is anchored in the NPM administrative philosophy. NPM's heyday started with the seminal 1991 article of Christopher Hood^[2] who conceived of NPM as a set of doctrinal ideas with a certain degree of coherence, (i.e., an administrative philosophy) that has come to be widely accepted. Hood^[3] mentions six doctrines to define the NPM administrative philosophy.

1. "Unbundle" the public service into corporatized units organized by product.
2. More contract-based competitive provision, with internal markets and term contracts.
3. Stress on private sector styles of management practice.
4. Put more emphasis on visible "hands-on" top management
5. Make performance standards and measures explicit, formal, and measurable.
6. Greater emphasis on output controls.

One could add to this the doctrine: "Apply quality management and a customer service focus." This doctrine was not the initial focus of NPM reforms, but has become very important in a "second wave of reforms."^[6]

The NPM proponents' consideration of the ethical consequences of administrative reform is rather modest. Sometimes, NPM proponents emphasize how competition mechanisms can reduce corruption, e.g., by lessening the opportunities of public servants to corruptly charge monopoly rents.^[7] More often, however, they remain silent on the topic of ethics.^[8] The assumption seems to be that ethical behavior will naturally follow from an implementation of NPM reforms. Ethical behavior is then mainly understood as behavior that strengthens "the three E's" (economy, efficiency, and effectiveness).

Traditional Public Administration

The second position is at the other side of the spectrum. It is rooted in "Traditional Public Administration"

(TPA), also referred to as the "old public administration"^[9] or the "orthodox model."^[10] Of course, TPA contains many different ideas and it would be incorrect and simplistic to present them as one monolithic block of doctrines and justifications without any internal debates or changes. Nevertheless, there seems to be broad consensus over a limited number of core doctrines including the following:^c

1. Public administration should be politically neutral.
2. The organizational structure should be a centralized bureaucracy.
3. Programs should be implemented through top-down control mechanisms, limiting discretion as much as possible.
4. Officials should be kept in a procedural straitjacket.

The administrative doctrine of TPA with regard to ethics management prescribes that the organizational processes be organized in such a way that "the individual ethical choice is limited to choosing to follow the rules (the ethical thing to do) or to violate them by commission or omission (unethical acts)."^[11] In the debate over the ethical effects of NPM reforms, the proponents of TPA have "throw[n] up their hands in horror at recent trends, seeing them as marking the destruction of the public service ethos."^[12] In the United Kingdom, Chapman and O'Toole^[13-15] argue for a return to the traditional civil service ethos and to the traditional concept of "public duty." Writing from the U.S. context, Frederickson^[16] also forcefully expresses his concerns about the ethical consequences of NPM innovations such as marketizing, privatizing, or contracting out, and he argues for a "re-regulation." One particularly prominent strand in the TPA tradition emphasizes the legal point of view and criticizes NPM "for its sometimes cavalier treatment of the rule of law, especially its free and easy slogans about eliminating red tape and letting managers manage."^[17]

New Public Service

The third position is derived from what has recently been labeled the "New Public Service" (NPS).^[9] Skidmore^[18] describes a similar approach as the "classical" or "Aristotelian" mode of organizing (referring to its roots in MacIntyre's work), while others^d have

^bAn earlier version of this discussion of the four positions was published by Maesschalck, J, 2004. The impact of new public management reforms on public servants' ethics: Towards a theory. *Public Administration* 82(2):465-489.

^cThe first three doctrines are taken from Denhardt and Denhardt.^[9] The fourth doctrine is taken from Ref.^[2]

^dFor example, Fox, C.J., Miller, H.T., 1996. *Postmodern public administration. Towards discourse*. Thousand Oaks: Sage.

labeled it the “communitarian/citizen alternative.” Denhardt and Denhardt present the NPS approach as a viable third alternative for the observed dichotomy between “the old public administration” and “the New Public Management.” Rather than traditional bureaucracies that are controlled from the top down and largely closed for citizens, Denhardt and Denhardt propose new mechanisms in which “the primary role of the public servant is to help citizens articulate and meet their shared interests rather than to attempt to control or steer society.”^[9] In order to achieve this overall aim, Denhardt and Denhardt propose seven doctrines that embody their administrative philosophy:

1. Serve, rather than steer.
2. The public interest is the aim, not the by-product.
3. Think strategically, act democratically.
4. Serve citizens, not customers.
5. Public servants should be attentive to the law, community values, political norms, professional standards, and citizen interests.
6. Value people, not just productivity.
7. Value citizenship and public service above entrepreneurship.

As for the recent debate over the impact of NPM reforms on ethics, those writing from an NPS point of view largely join the TPA advocates in their negative assessment of the ethical consequences of NPM reforms, but propose a different solution. Specifically, Denhardt and Denhardt argue that public sector organizations should be organized in such a way that public servants are not responsive to “constituents and clients” (TPA), nor to “customers” (NPM), but to “citizens.” Citizens are described as those people who “demonstrate their concern for the larger community, their commitment to matters that go beyond short-term interests and their willingness to assume personal responsibility for what happens in their neighborhoods and the community.”^[9] To make public servants capable of being responsive to these kinds of citizens, the doctrines described above should be the guide. Several other authors have criticized NPM from an NPS point of view. Although he also used TPA arguments to criticize NPM (see above), most of Frederickson’s prescriptive claims (particularly in “The Spirit of Public Administration”) rather correspond to the NPS administrative philosophy. One example of this is his plea for a “combination of patriotism (the love of the regime values) with benevolence (the love of others).”^[19] Gawthrop^[20] fits the NPS philosophy even more neatly. He contrasts the “democratic spirit” with the NPM-type “entrepreneurial spirit” and argues that the former should prevail. He also strongly criticizes

the TPA-type “ethics of compliance” he observed in the U.S. public administration the past five decades for leading to a soulless public administration, obsessed with procedural correctness. He argues that public servants should have a personal responsibility, with “service as the center of value.”

Ethics Management

The fourth position in the debate does not clearly build on a single identifiable administrative philosophy such as the previous three. Its central argument is that public sector ethics management is an important and evolving subfield of public management, which deserves its own set of doctrines, hence the “Ethics Management” approach. Its core doctrine is that an ethics policy should be a combination of doctrines from the three aforementioned approaches, adapted to the specific circumstances of the organization. Admittedly, arguments for a combination of different approaches have been articulated before. Cooper,^[21] for example, proposed such a combined approach in his seminal “The Responsible Administrator.” However, the success of the NPM discourse and the concomitant ethical concerns fostered a recent revival of this approach. The Public Management (PUMA) Department of the OECD^[22,23] has been among its most prominent promoters, together with practice-oriented academics such as Gilman^[24] or Uhr.^[25] These and other recent proponents agree with the NPM advocates that NPM reforms can have many beneficial consequences, but are at the same time conscious of the undesirable effects (particularly unethical behavior). However, in contrast with TPA and NPS, their solution is not to do away with NPM innovations, but to complement them with a well-developed public sector ethics management that is adapted to the reforms. They see two possible approaches to such an ethics policy. The OECD, Gilman, and originally Paine^[26] identify these as the “compliance” and the “integrity” approach, respectively. The distinction goes back to the famous Friedrich-Finer debate over the importance of internal and external controls on public servants, respectively.^e The “compliance” approach to ethics management emphasizes the importance of external controls on the behavior of public servants (e.g., legislation, strict behavioral ethics codes, and extensive control mechanisms) and thus comes very close to the ethics management doctrines of TPA. However, the “Ethics Management” proponents

^eFor a discussion, see Cooper, T.L., *The responsible administrator: An approach to ethics for the administrative role*; San Francisco: Jossey-Bass, 1998; 131–163.

observe that this compliance approach does not fit the NPM-style results-based managerial approach. It thus violates their basic doctrine of a management style adapted to the circumstances. Instead, NPM-type changes need to be complemented by an “integrity” approach to ethics management, which emphasizes internal control: control exercised by the public servant on herself/himself. This integrity approach is “based on aspirations, relies on incentives, and encourages good behavior rather than policing and punishing errors and wrongdoing.”^[23] Such an ethics management style is necessary because the increased discretion that goes with the reduction of hierarchy in NPM-type reforms needs to be paralleled by an ethics management style that supports the public servants in dealing with this discretion, rather than limiting their discretion by developing new rules and control systems.

CONCLUSION

This article presented four prominent positions^f in the debate about the impact of administrative reform on the ethics of public servants. Although the positions in this doctrinal debate seem rival, all authors can easily point at empirical instances that support their claims. For example, while the NPM literature is ripe with success stories of NPM reforms and convincing examples of the undesirable effects of traditional bureaucracies, the TPA proponents can easily point at the scandals that followed NPM reforms and show how these would not have occurred in a more traditional bureaucracy. How should we deal with these diverse claims? If all participants in the debate can provide convincing examples of their position, who should we believe?

To answer this question, it might be useful to distinguish between the normative and the empirical claims made in this literature. The normative claims underlying the three first positions^g are clearly rival. The proponents of these positions evidently disagree in their evaluation of both the desirability of particular types of ethical behavior and the seriousness of types of unethical behavior. In contrast with this rivalry among the normative claims, the empirical claims are rather complementary. While the proponents of a particular approach concentrate on its beneficial effects, the opponents point at its deficiencies. Although this might look rival, such claims can

empirically be complementary: It is perfectly consistent to claim that a particular approach can be beneficial in particular circumstances, but engender undesirable effects in other circumstances. In fact, this is an important maxim in management: Each innovation, however well intended, carries with it the risk of particular excesses. Kathryn Denhardt^[27] formulated this eloquently in the context of ethics management: “Every organizational structure can promote an ethical stance yet at the same time contain the seeds of pathology.”

One prominent strategy to cope with this risk is to ensure that all main perspectives are taken into consideration and structurally anchored in the organization. The permanent tension between the approaches that follows from this then helps to avoid blind spots for the potential negative effects of one particular approach.^h According to this strategy, the decision is not which of the three main administrative philosophies—New Public Management, Traditional Public Administration, and New Public Service—should be applied, but how all three of them should be combined. This is in fact what the Ethics Management approach, the fourth position in the debate, sets about to do. Proponents of this approach accept some NPM reforms, but also require that these are combined with two other approaches: a considerable degree of TPA (the “compliance” approach in the ethics management jargon) to delineate the discretion and ensure that the discretion is not abused and a considerable degree of NPS (the “integrity” approach) which stimulates public servants to use the remaining discretion to “serve citizens, not customers.” Such a complex balancing exercise will not guide us to clear-cut and neat solutions, but it does reduce the risk of administrative reforms leading to ethical disaster.

REFERENCES

1. Pollitt, C.; Bouckaert, G. *Public Management Reform. A Comparative Analysis*; Oxford University Press: Oxford, 2000.
2. Hood, C. A public management for all seasons? *Public Adm.* **1991**, 69 (1), 3–19.
3. Hood, C. *Explaining Economic Policy Reversals*; Open University Press: Buckingham, Philadelphia, 1994.
4. Cohen, S.; Eimicke, W. Is public entrepreneurship ethical? *Public Integr.* **1999**, 1 (1), 54–74.
5. Hood, C.; Jackson, M. *Administrative Argument*; Aldershot: Dartmouth, 1991.
6. Pollitt, C. *Managerialism and the Public Services*; Blackwell: Oxford, 1993.

^fIt is obvious that these are four sets of doctrines and not four groups of authors. Although most authors clearly have one of the four positions as their favorite, several of them endorse doctrines from different philosophies. This goes for at least two of the most prominent authors that were discussed: Cooper and Frederickson.

^gThe Ethics Management position is an exception.

^hFor further development of this argument, see Maesschalck, J., forthcoming. Approaches to ethics management in the public sector: A proposed extension of the compliance–integrity continuum. *Public Integrity*.

7. Klitgaard, R. *Controlling Corruption*; University of California Press: Los Angeles, CA, 1988.
8. Menzel, D.C. The morally mute manager: fact or fiction? *Public Pers. Manage.* **2000**, 28 (4), 515–527.
9. Denhardt, R.B.; Denhardt, J.V. The new public service: serving rather than steering. *Public Adm. Rev.* **2000**, 60 (6), 549–559.
10. Fox, C.J. The use of philosophy in administrative ethics. In *Handbook of Administrative Ethics*; Cooper, T.L., Ed.; Marcel Dekker: New York, 2001; 105–130.
11. Fox, C.J. The use of philosophy in administrative ethics. In *Handbook of Administrative Ethics*; Cooper, T.L., Ed.; Marcel Dekker: New York, 2001; 110 pp.
12. Greenaway, J. Having the bun and the halfpenny: can old public service ethics survive in the new whitehall? *Public Adm.* **1995**, 73, 357–374.
13. Chapman, R.A. Problems of ethics in public sector management. *Public Money Manag.* **1998**, 18 (1), 9–13.
14. Chapman, R.A. Change in the civil service. *Public Adm.* **1994**, 72, 599–610.
15. Chapman, R.A.; O'Toole, B.J. The role of the civil service: a traditional view in a period of change. *Public Policy Adm.* **1995**, 10 (2), 3–20.
16. Frederickson, H.G. Public ethics and the new managerialism. *Public Integr.* **1999**, 1 (3), 265–278.
17. Rohr, J.A. *Civil Servants and their Constitutions*; University Press of Kansas: Kansas, 2002; xi pp. (For examples of this legal criticism on NPM, see, e.g., Cooper, P.J., 2000. *Public law and public administration. third edition*. Itasca, IL: F.E. Peacock.).
18. Skidmore, M.J. Ethics and public service. *Ann. Am. Acad. Polit. Soc. Sci.* **1995**, 537, 25–36.
19. Frederickson, H.G. *The Spirit of Public Administration*; Jossey-Bass: San Francisco, CA, 1997; 202 pp.
20. Gawthrop, L.C. *Public Service and Democracy*; Chatham House: New York, 1998; 80 pp.
21. Cooper, T.L. *The Responsible Administrator: An Approach to Ethics for the Administrative Role*, 1st Ed.; Kennikat: Port Washington, NY, 1982.
22. OECD. *Ethics in the Public Service. Current Issues and Practices*; OECD: Paris, 1996.
23. OECD. *Trust in Government. Ethics Measures in OECD Countries*; OECD: Paris, 2000; 25 pp.
24. Gilman, S.C. Public sector ethics and government reinvention: realigning systems to meet organizational change. *Public Integr.* **1999**, 1 (2), 175–192.
25. Uhr, J. Institutions of integrity. Balancing values and verification in democratic government. *Public Integr.* **1999**, 1 (1), 94–106.
26. Paine, L.S. Managing for organizational integrity. *Harvard Bus. Rev.* **1994**, (2).
27. Denhardt, K.G. Organizational structure as a context for administrative ethics. In *Handbook of Administrative Ethics*; Cooper, T.L., Ed.; Marcel Dekker: New York, 1994; 181 pp.

Ethics and Information and Communication Technology

Brian Negin

Israel Central Bureau of Statistics, Jerusalem, Israel

INTRODUCTION

Information and communication technology, in particular the Internet, has benefited global society by enabling unfettered dissemination of, and access to, information on a scale previously unknown to mankind. At the same time, this technology has created challenges to the enforcement of existing regional norms, pitting the value of freedom of speech (and of information) against other equally important values, such as the protection of human dignity.

THE IMPORTANCE OF INFORMATION AND COMMUNICATION IN SOCIETY

Human life and society depend on the ability to communicate information. While this seemingly self-evident truth applies to human development over tens of thousands of years, the recognition of the importance of information to human society is only a relatively recent one, rooted in the Enlightenment. The Enlightenment signified a societal change of orientation from natural duties to natural rights.^[1] One of the natural rights that has figured prominently in the development of Western society has been the *freedom of speech*, necessary for the actualization of democratic society. As recognized by the Supreme Court of the United States, *freedom of speech* includes the freedom to both access and disseminate information.^[2] An alternative approach expressed in the draft *Declaration of Principles* of the World Summit on the Information Society views the freedom to access and disseminate information not as a corollary of the *freedom of speech*, but as an independent right.^[3]

LIMITATIONS ON FREEDOM OF INFORMATION AND COMMUNICATION

Human society depends on the ability of human beings to communicate and also on the existence of norms that establish the borders of permitted behavior. Among these norms are those that define the permissible *content* of speech, as well as those that define permissible *access* to and *use* of that content. In a modern society, one may distinguish between such norms that

are based solely on social or cultural pressures (*extra-legal norms*) and those that are expressed in law (*legal norms*).

Extralegal normative control of behavior, including speech, generally occurs within social and professional networks that require *correct* behavior from their members or as a result of dedicated pressure groups promoting an agenda, such as those pressing for the removal of *controversial* reading material from school curriculums.^[4] Legal control of the content of speech protects societal values considered at least equally important as the freedom of speech. Criminal laws, for example, prohibit sedition, incitement to violence, incitement to racism, and the dissemination of obscene material. Civil laws prohibit the dissemination of defamation, the invasion of privacy, the violation of trade secrets, and the infringement of copyright.

Thus we see that while *freedom of speech*, including freedom of access to, and the dissemination of, information, is a fundamental value of modern civilization, it is not an absolute and unlimited freedom. Restraints on freedom of speech usually reflect a fine balance among competing societal values as developed over time within evolving social and technological contexts.

THE PROMISE AND THE CHALLENGE OF THE INTERNET

The Internet has created an exceptional platform for the dissemination of, and the access to, information. As reflected in the World Summit on Information Technology, one of the primary global issues today is how to provide full and uncensored Internet access to the world's population, including those nations that either lack a sufficient technological infrastructure or that choose, for political reasons, to limit access. It is commonly believed that access to information over the Internet, including the ability to communicate through it, is an enabling factor in the positive development of all people around the world. However, this utopian spin on the Internet tends to overlook social, ethical, and legal problems that the new technology has created. As scholars have noted, new technology can bring good to society, but can also bear unexpected consequences.^[5]

One of the unexpected consequences of the Internet is that its technology has created challenges to the enforcement of accepted societal norms, legal and extralegal, that define the permissible content of speech and the permissible access to, and use of, information. The Internet also has created new difficulties regarding competing regional or national norms.

These challenges highlight the ongoing struggle of society to define and enforce normative values necessary for human survival. This struggle is apparent in Section 9 of the current draft Declaration of Principles of the World Summit on the Information Society, titled *Ethical Dimensions of the Information Society*. Article 52 in this Section states:^[6]

The Information Society should be subject to universally held cultural and ethical values such as truth, justice, solidarity, tolerance, human dignity, shared responsibility, transparency and accountability, and without prejudice to the moral, social and religious values of all societies. All actors in the Information Society should seek to promote the common good, protect privacy, and to prevent abusive uses of ICTs. The freedom of use of ICTs should not undermine the human dignity, human rights and fundamental freedoms of others, including personal privacy, matters of faith and other personal beliefs. These values are particularly relevant when commercial activities are conducted through networks.

Following are three topics that illustrate some of the ethical issues consequential to the development of information and communication technology and particularly the Internet. These are online defamation, protecting minors from sexually explicit content on the Internet, and online privacy and confidentiality.

ONLINE DEFAMATION

The prohibition against defamation is at least as old as the Bible. The prohibition in the Ninth Commandment, against bearing false witness, has been interpreted as also prohibiting gossip and slander.^[7] In modern terms, the prohibition can be understood as protecting the value of human dignity, central to the modern worldview evolved in the Enlightenment. As a basic human value, its pedigree predates *freedom of speech*. At the same time, the content and enforcement of defamation law must be balanced against the value of freedom of speech.

A person can be held legally responsible for defaming another person. When defamatory speech is disseminated online, it must normally reside, if only temporarily, on the computer system of a service provider in order for it to be accessed. Barring the existence of specific legislation to the contrary, the

common default legal position in most Western countries would hold service providers liable for not removing such illegal content if duly requested to do so, analogous to the legal liability of distributors for the dissemination of third-party defamation in the physical world. This was the situation according to U.S. case law until 1995,^[8] as well as the position expressly taken by Germany in its Act on the Utilization of Teleservices.^[9]

However, in 1997, a U.S. Court of Appeals upheld a lower court decision immunizing service providers from all liability for *third-party* illegal content, including immunity from the equivalent of *distributor liability* in the physical world for not removing defamatory content when asked to do so.^[10] The court was concerned that recognizing service provider distributor liability would chill free speech on the Internet by encouraging an onslaught of demands by purportedly injured parties for the removal of allegedly defamatory material from service providers' computer systems. Service providers would either comply, thus implicating themselves in prior restraint of free speech, or not comply and face enormous legal fees to defend themselves while passing on the additional costs to users in the way of higher fees.

The preponderance of U.S. case law, following the above precedent, today holds that an injured party to online defamation can only hold the creator of the defamatory content liable to pay compensation for damage.^[11] The result is that the substantive law of defamation remains the same, while the ability to enforce it in the online environment has been reduced. This change reflects a weakening of the norm protecting human dignity while strengthening a competing societal norm, the freedom of speech. In this sense, the technology of the Internet has wrought a substantive change in the balance of societal values, at least in the United States, in the context of defamation law.

PROTECTING MINORS FROM SEXUALLY EXPLICIT CONTENT ON THE INTERNET

Freedom of speech includes the freedom to disseminate sexually explicit material, as long as it does not violate local community standards. However, even sexually explicit material that is legally available to adults might be harmful to vulnerable and innocent minors. The technology of the Internet has made access to sexually explicit content very easy, even for minors. In addition, a mistyped web address can inadvertently bring up a pornographic Web site, even to the most experienced adult. How can children be protected against the harmful effects of such material online?

While some argue that the best protection for minors against harmful content on the Internet is

education and parental guidance,^[12,13] the U.S. government has decided that this is a matter for legislative intervention. Two federal attempts at criminalizing the dissemination to minors, over the Internet, of constitutionally protected material that, for minors, would be considered *harmful*^[14,15] have been judged to be unconstitutional because they were overbroad and limited adults' rights to otherwise legally protected speech.^[16,17] A third legislative attempt to protect minors from such material, the Children's Internet Protection Act (CIPA),^[18] has been upheld by the U.S. Supreme Court.^[19]

Under CIPA, a public library may not receive federal assistance for the provision of Internet access, unless it installs software to block obscene or pornographic images and to prevent minors from accessing material deemed harmful to them. The legislation addresses, among other things, the problem that library patrons of all ages, including minors, regularly use public library Internet computers to access not only information, but also online pornography. In addition, library staff or patrons may at times be exposed to pornographic images left displayed on computer terminals or printed on library printers.

The Supreme Court held that public libraries, by necessity, must make value judgments about what material will be included in their collections. Public libraries may also make value judgments regarding the type of content they will make available to the public over their Internet computers. Just as most libraries exclude pornography from their print collection, they may exclude it from Internet access on library computers by using filtering applications—even if those applications are imperfect and might *overblock* access to legal and non-harmful information while *underblocking* access to some illegal or harmful sites. Therefore public library use of Internet filtering software does not violate adult patrons' First Amendment rights.

Notwithstanding the legality of the use of filtering software in public libraries, discussion continues as to the efficacy of the use of such software and the alternatives to it. In the settlement of a workplace harassment case brought by Minneapolis librarians regarding patron use of public library Internet computers to access pornography, it is reported that library officials will consider, in addition to filtering software, changes in the printing of Internet material, increased sanctions for those who violate library Internet policy, and consultation with staff about placement of terminals.^[20,21]

The Children's Internet Protection Act also requires installation and use by schools of filtering or blocking software on Internet computers to be eligible to receive or retain universal service assistance. The Electronic Frontier Foundation, in its study on Internet blocking in public schools, criticizes the efficacy of the use of such software.^[22] It concludes that because filtering

or blocking software both *overblocks* permissible material and *underblocks* harmful material, it cannot sufficiently protect children while at the same time, it damages educational opportunities for students to access information.

PRIVACY AND CONFIDENTIALITY ON THE INTERNET

The technology of the Internet enables the collection of information about users of services such as the World Wide Web, even without the individual user being aware of such. This information might include data on the user himself (name, E-mail address, and perhaps more) and the web pages visited on any given site. Sometimes the information on the user is anonymous, in the sense that Internet technology enables, through the use of *cookies*, the tracking of a particular computer's navigation through a given Web site. The computer can be identified [by *cookie*^[23] or Internet Protocol (IP) address], but unless personal information has been provided by the user of the computer, it is difficult to attribute web surfing habits to a particular, identified person.

However, the tracking of identified users on the Internet is becoming easier as more and more Web sites require *registration* to access the information on the site. The *New York Times* Web site is an example.^[24] To receive free access to the online newspaper, one must first register and provide the following personal information: E-mail address, country, zip code, age, sex, household income, industry, job title, job function, and subscription status to the *New York Times*. The use of this information is in accordance with a privacy policy^[25] that describes what personal information is collected, how usage is tracked on the site (*cookies* and IP address), and what use will be made of the information collected (statistical analysis and banner advertising, E-mail if you elect it as an option during registration, and sharing information with partners on an opt-in basis only). This example demonstrates the striking of a balance between commercial enterprise and the protection of privacy on a contractual basis. To receive a commercial commodity for free, a person gives up a bit of his or her own privacy under the terms of an agreement that delineates the permitted use of the information provided.

Non-contractual limitations on the use of personal information also exist, mandating maintaining the confidentiality of third-party private information stored on computer systems. For example, the U.S. government must abide by the Privacy Act of 1974 that prohibits use of personal information by the federal agencies for purposes other than those for which it was provided.^[26] While U.S. legislation applies only

to the federal government, the European Union Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data applies to everyone.^[27] It also requires that personal information be used only for the purpose for which it was provided. The directive has been implemented by the member states in their own internal legislation.^[28]

Setting the boundaries of permitted use of private information in a digital environment requires taking steps to protect that information from unauthorized access or use. The person, business, organization, or government holding the information must establish administrative policies regarding the use of the information that conform with the privacy norms binding it. In addition, it must also take computer security measures to prevent unauthorized access to that information both from within and from without. Steps from within include limiting access to such information only to those who have been granted access privileges through user name, password, and access rights administration. The protection from outside access via public communication networks, including the Internet, will usually include, in addition to the above, firewalls and other software and hardware that can make unauthorized access by outsiders very difficult. These computer security measures are normally grounded in an organization's policy on computer security intended to set the guidelines not only for the protection against unauthorized access to sensitive information residing on the computer system, but also to protect the system itself against being compromised as a result of external attack, natural disasters, or simple equipment failure. For example, Appendix III of OMB Circular A-130 on the management of federal information resources requires that federal "[a]gencies shall implement and maintain a program to assure that adequate security is provided for all agency information collected, processed, transmitted, stored, or disseminated in general support systems and major applications."^[29,30]

CONCLUSION

The ease of use of the Internet as a means of accessing and disseminating information, its global reach, and the power of its technologies create challenges in enforcing established societal values or balancing among competing values. Freedom of speech enjoys a technological advantage in this environment, as does the ability to gather and process information about individuals. As a result, other values such as the protections against defamation, against harm to vulnerable minors, and against the invasion of privacy may suffer. Human society has always had to find the proper balance

between these and other competing values, especially in the context of new information technologies.

This process is an ongoing one. Society continuously examines and reexamines its resolution of these conflicts even in the absence of new technologies, learning from experience as well as responding to external forces, such as the terror attacks of September 11. In this context, the conflicts discussed above in this entry continue to be scrutinized by the public, the courts, and the legislator and may remain unresolved for some time to come. At the same time, information and communication technologies shall continue to develop, providing new challenges to realizing their potential while maintaining the ethical balances necessary to promote the common good.

REFERENCES

1. Strauss, L. *Natural Right and History*; The University of Chicago Press: Chicago, 1953.
2. *United States et al. v. American Library Association, Inc. et al.* United States Supreme Court, 2003; <http://supct.law.cornell.edu/supct/html/02-361.ZS.html> (accessed September 2003).
3. *Draft Declaration of Principles*; Document WSIS/PC-3/2-E, The World Summit on the Information Society: Geneva, Switzerland, 2003. Section A; http://www.itu.int/dms_pub/itu-s/md/03/wsisp3/doc/S03-WSISPC3-DOC-0002!!PDF-E.pdf (accessed September 2003).
4. Ravitch, D. *The Language Police—How Pressure Groups Restrict What Students Learn*; Alfred A. Knopf: New York, 2003.
5. Dunlop, C.; Kling, R. Introduction—social controversies about computerization. In *Computerization and Controversy: Value Conflicts and Choices*; Dunlop, C., Kling, R., Eds.; Academic Press: Boston, 1991; 1–29.
6. *Draft Declaration of Principles*; Document WSIS/PC-3/2-E, The World Summit on the Information Society: Geneva, Switzerland, 2003. Section 9; http://www.itu.int/dms_pub/itu-s/md/03/wsisp3/doc/S03-WSISPC3-DOC-0002!!PDF-E.pdf (accessed September 2003).
7. Scherman, N. *Stone Edition Chumash: The Torah, Haftaras and Five Megillos with a Commentary From Rabbinic Writings*; Artscroll Mesorah: Brooklyn, 1993; 412 pp.
8. Negin, B. *Illegal Content on the Internet*; Israel Government Internet Committee: Jerusalem, Israel, 1998; <http://www.itpolicy.gov.il/topics/docs/mso2F8.pdf> (accessed September 2003).
9. *Act on the Utilization of Teleservices*. Germany. http://www.iid.de/iukdg/aktuelles/fassung_tdg_eng.pdf (accessed September 2003).
10. *Zeran v. America Online Inc.*, 129 F. 3d 327, United States Court of Appeals, 4th Circuit, 1997; <http://laws.lp.findlaw.com/4th/971523p.html> (accessed September 2003).

11. *Batzel v. Cremers*, United States Court of Appeals, 9th Circuit, 2003; <http://caselaw.lp.findlaw.com/data2/circs/9th/0156380p.pdf> (accessed September 2003).
12. Willard, N. *Choosing Not To Go Down the Not-So-Good Cyberstreets*. <http://responsiblenetizen.org/pdf/nwnas.pdf> (accessed September 2003).
13. *Welcome to ALA's Web Site on the Children's Internet Protection Act (CIPA)*; American Library Association. http://www.ala.org/Content/NavigationMenu/Our_Association/Offices/ALA_Washington/Issues2/Civil_Liberties,_Intellectual_Freedom,_Privacy/CIPA1/CIPA.htm (accessed September 2003).
14. *The Communications Decency Act (CDA)*. Section V of The Telecommunications Act, Pub. L. No. 104-104, 110 Stat. 56 (1996), codified at 47 U.S.C. §223(a) and §223(d).
15. *The Children's Online Protection Act (COPA)*. Pub. L. No. 105-277, 112 Stat. 2681 (1998), codified at 47 U.S.C. §231.
16. *Reno v. ACLU*, 521 U.S. 844, United States Supreme Court, 1997; <http://supct.law.cornell.edu/supct/html/96-511.ZO.html> (accessed September 2003).
17. *ACLU v. Ashcroft*, United States Court of Appeals for the 3rd Circuit, 2003; <http://laws.findlaw.com/3rd/991324.html> (accessed September 2003).
18. *Children's Internet Protection Act (CIPA)*. Pub. L. No. 106-554, 114 Stat. 2763A-335; http://www.ala.org/Content/NavigationMenu/Our_Association/Offices/ALA_Washington/Issues2/Civil_Liberties,_Intellectual_Freedom,_Privacy/CIPA1/cipatext.pdf (accessed September 2003).
19. *United States et al. v. American Library Association, Inc. et al.*, United States Supreme Court, 2003; <http://supct.law.cornell.edu/supct/html/02-361.ZS.html> (accessed September 2003).
20. *Librarians Settle Internet Porn Case*. Brainerddispatch.com, August 16, 2003. http://www.brainerddispatch.com/stories/081603/sne_0816030021.html (accessed September 2003).
21. Volokh, E. *Squeamish Librarians*; 2001. <http://www1.law.ucla.edu/~volokh/harass/library.htm> (accessed September 2003).
22. *Internet Blocking in Public Schools*; Electronic Frontier Foundation, 2003; http://www.eff.org/Censorship/Censorware/net_block_report (accessed September 2003).
23. Whalen, D. *The Unofficial Cookie FAQ*; Cookie Central, 2002; <http://www.cookiecentral.com/faq/> (accessed September 2003).
24. New York Times On The Web; <http://www.nytimes.com> (accessed September 2003).
25. *Privacy Policy*. The New York Time on the Web, 2003; <http://www.nytimes.com/ref/membercenter/help/privacy.html> (accessed September 2003).
26. *The Privacy Act of 1974*. Codified at 5 USC §552a; <http://www.usdoj.gov/04foia/privstat.htm> (accessed September 2003).
27. *Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of such Data*; The European Commission. http://europa.eu.int/comm/internal_market/privacy/law_en.htm (accessed September 2003).
28. *Status of Implementation of Directive 95/46 on the Protection of Individuals with Regard to the Processing of Personal Data*; The European Commission. http://europa.eu.int/comm/internal_market/privacy/law/implementation_en.htm#ukingdom (accessed September 2003).
29. *Management of Federal Information Resources*. Circular No. A-130, Revised, (Transmittal Memorandum No. 4); Office of Management and Budget: Washington, DC; <http://www.whitehouse.gov/omb/circulars/a130/a130trans4.html> (accessed September 2003).
30. *Security of Federal Automated Information Resources*. Appendix III to OMB Circular No. A-130; Office of Management and Budget: Washington, DC; http://www.whitehouse.gov/omb/circulars/a130/a130appendix_iii.html (accessed September 2003).

Ethics and Public Policy

Amy K. Donahue

*Institute of Public and Urban Affairs and Department of Political Science,
University of Connecticut, Storrs, Connecticut, U.S.A.*

INTRODUCTION

Ethical questions surrounding the substance and application of public policy arise daily, as a cursory review of newspapers and news broadcasts readily demonstrates. In a single week in 2001, for example, the questions of whether the government should fund research that uses human genetic material, how to allocate tax refunds, whether to invest in building a national missile defense system, and whether to allow oil exploration in a wildlife refuge all competed for attention as part of the national policy agenda. Such issues are inherently ethical because they turn on normative choices—trade-offs among multiple values, judgments about how to interpret complex data, and decisions that distribute power and resources. In short, interactions between ethics and public policy are at once ubiquitous and complex.

DEFINITIONS

Public policy and ethics are each vast and disparate realms of scholarship and practice that defy concise definition. Nonetheless, some broad characterization of these concepts and the linkages between them is warranted and useful.

Public Policy

Broadly speaking, public policy comprises the development and outcomes of government decisions. It is fundamentally concerned with three kinds of decisions: first, what problems exist in society?; second, when should government intervene to solve them?; third, how should it intervene?; or, as Harold Lasswell said in 1936, public policy is about deciding “who gets what, when, and how.”^[1] As commonly used, the term “public policy” has several variants. It can refer to the outputs or outcomes of government processes—that is, the substance of what government does, rather than the way it makes decisions—noting also that what government intends to do is different from what it actually does. Alternatively, it can connote the general

framework of authoritative rules that guide government activities, including but not limited to legislation. Finally, it may mean a program of action designed to achieve a specified goal or purpose that often involves all levels of government and formal as well as informal actors.^[2]

Public policy fulfills an array of societal objectives. Among the most important functions of policy are the reconciliation of conflicting claims on scarce resources (accomplished by establishing incentives for cooperation or collective action that would be irrational absent government intervention, or by providing direct services or benefits to citizens), the prohibition of behavior unacceptable under reigning societal norms, the protection of individual and group rights, and the promotion of activities that are essential or important to governance. To meet these objectives, policy is created in a vast array of substantive realms that distribute and redistribute resources and regulate behavior.

Public policy is generated through a complex series of negotiated activities collectively known as the policy-making process. This process is how society validates claims about resource and power needs—how it decides which needs are real and legitimate, and what approach it should take to resolving them. This is particularly salient because government policy is generally believed to be about collective interests, but needs are almost always experienced individually. In a sense, policy politics is about defining individual needs collectively, as “public” needs.^[3]

Generally, the policy-making process incorporates policy formulation, implementation, and the subsequent effects of policy. An important point is that the process must also account for the range of factors that affect whether policy decisions and their social effects are consistent with popular preferences. In other words, public policy cannot be separated from the environment in which it is made. There are two important dimensions and effects of the environment to consider. First, the political and ideological context must be acknowledged. These are fundamentally important because they define what forms of policy are acceptable and whether they can garner enough support to be implemented. Second, a whole array of

socioeconomic factors affects whether a policy can achieve its desired outcomes. These contingencies bear on how ethical principles can be defined and applied in the context of public policy making.

Ethics

Ethics receives treatment in several entries in this volume that clarify others of its aspects. With respect to public policy, ethics refers to moral reasoning and behavior by policy makers and other stakeholders in the public policy system. Ethics enters this system because all public policies express value judgments by allocating resources, entitlements, and authorities. As Swain and Duke explained, “ethics concerns things about which individuals have behavioral choices. Those choices have social consequences.”^[4]

The exercise of ethics in policy making proceeds from an array of philosophic traditions—most generally, deontological theories and teleological theories. Generally, deontological approaches are more classical and are concerned with the moral legitimacy of the policy process. These approaches tend to yield rules that prescribe proper behavior for decision makers and are meant to be universally applied. Along these lines, some in the field take a deontological view, under which ethics refers to the system of moral conventions that direct conduct. This perspective gives rise to published codes of ethics, available for an array of professional associations and public agencies, that influence the nature of public policy by guiding the behavior of policy makers.

Teleological approaches, in contrast, are focused on the goodness of consequences, rather than the character of the process that caused them. This view permits decision makers discretion in the application of moral rules to account for contingencies that affect the achievement of desirable outcomes. The literature of public administration commonly exhibits a teleological bent whereby actors weigh policy alternatives according to the net utility (or satisfaction) they generate. This is most evident in the frequent references to benefit–cost analysis as an appropriate rational basis for making normative policy decisions.

In practice, neither deontology nor teleology offers fully apposite guidance under all circumstances, and elements of both typically enter policy-making activities. Policy makers thus may employ broad ethical principles—or even specific rules of conduct—but make exceptions when these principles and rules are inconsistent or policy outcomes can be improved.^[5] The impracticality of rigid conventions has prompted several scholars to characterize the nature of ethics in practice as evolutionary. Here the professional responsibilities of actors in the policy process are defined over

repeated application to generate what Cooper termed an “operational ethic.”^[6] Similarly, Tong saw public actors as developing (or in need of developing) a personal moral perspective that upholds a chosen set of ideals^[7]—a compass, rather than a rule book.

ETHICS IN THE POLICY-MAKING PROCESS

Most scholars agree that public policy making involves an iterative staged process, and most agree what the stages generally are, although how exactly they are delineated varies somewhat. Each stage poses a forum for value-based debate and thus an opportunity for ethical dilemmas to arise. Generally, the main stages and the ethical considerations each raises are (see Fig. 1):

1. *Issue identification:* Public officials and non-governmental actors care about a broad range of matters. During issue identification, this list is narrowed to an agenda, or the set of concerns policy makers seek to address. In other words, issue identification is the process by which policy makers recognize specific problems, decide that government can and should address them, and initiate the development of focused solutions. Ethics bears on issue identification because alternative sets of values offer different conceptualizations of the public interest and thus suggest different public policy objectives. Moreover, how policy issues are characterized at this stage of the policy process later shapes the nature of solutions sought and the commensurate distribution of resources that results from execution of these solutions. Further, the particular values on which a policy agenda is based prompt the use of certain criteria for assessing the value of benefits and costs that will accrue from a given policy over other possible criteria.
2. *Policy creation:* This phase starts with policy analysis, a mechanism by which the range of possible alternative solutions to a problem is identified and analyzed in terms of their monetary and non-monetary benefits and costs. Then, an alternative is chosen and a policy is drafted that specifies what objectives are sought and how things will be done to reach them. Political considerations often modify the proposal to make it acceptable (known as the process of legitimation). The policy is then adopted, most commonly as law, but executive orders, rules, regulations, and court orders are also possible. Ethics enters the public policy process at this stage because all policy fundamentally involves the distribution of resources, which are finite.

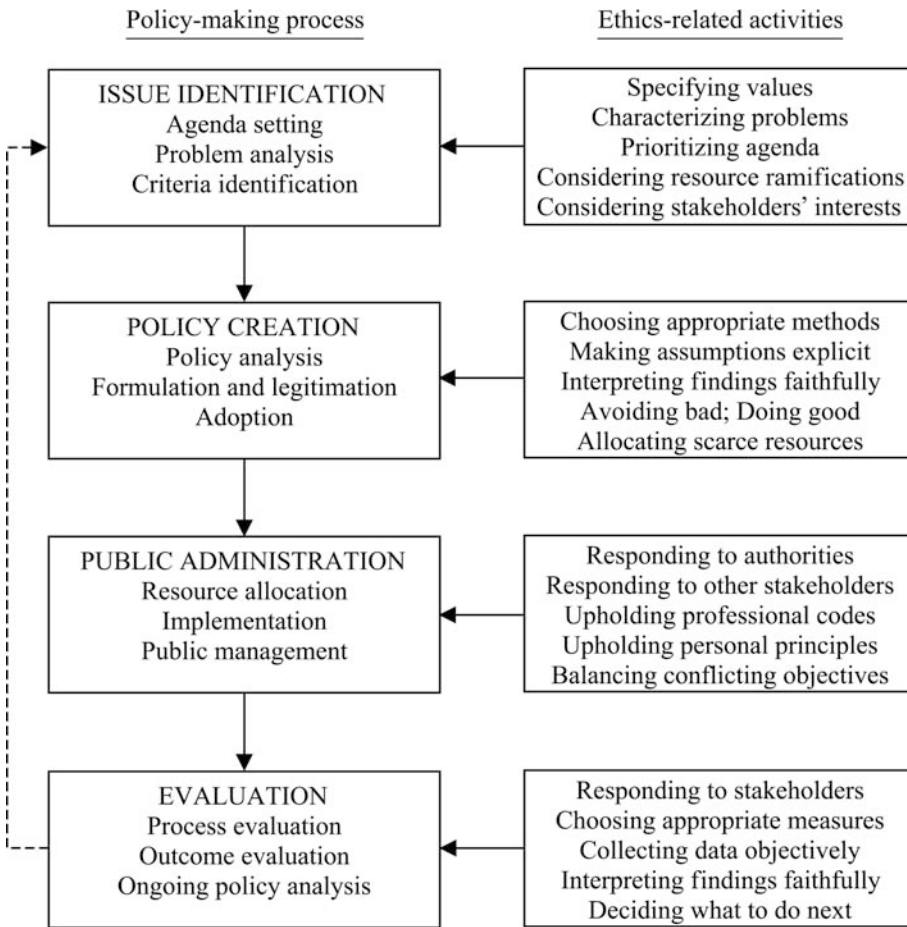


Fig. 1 Relationship of ethics to the policy-making process.

This scarcity forces policy makers to make choices that trade off between what Rohr characterizes as the two key activities of ethics: preventing bad and promoting good.^[8] Ethics is also an important dimension of analysis where the choice of research methods, and the interpretation of data and findings, involve a series of judgments that profoundly influence how policy makers understand the ramifications of each policy alternative.

3. *Public administration*: Once adopted, the public bureaucracy usually takes over to implement the chosen policy—to put it into action. Implementation may proceed according to a rigid program that explicitly and rationally lays out procedures to forestall cooptation of the process, or it may be freely structured to permit policy improvements and more flexible response to contextual realities.^[2] Regardless of how carefully or loosely defined the process, implementation typically involves a long series of activities and a multitude of actors; thus, there are many points at which there may be pressure to modify the policy from its original design. As a result, public administrators are repeatedly

confronted with ethical dilemmas in the course of implementation. This may occur for a host of reasons, including that the stated policy goals are vague, obscured, or forgotten; stakeholders seek to substitute their own objectives or to intervene to halt or alter implementation; unforeseen technical difficulties arise that require adjustment; new information about policy effects arises as implementation progresses; multiple internal and external authorities have conflicting expectations; or there are overlapping jurisdictions and authorities that must be resolved.

4. *Evaluation*: Once implemented, it remains to decide if the policy worked—in other words, to determine the extent to which it produced its intended consequences, and whether there were any unintended ones. Frequently (even usually), policies fall short of their objectives, often because of ambiguous goals, ignorance about how to reach those goals, adversarial political conditions, or poor measures of success. Thus, evaluation also helps to uncover why policies do not wholly meet the agendas of those who designed them. There are hundreds

of types of evaluation that essentially fall into two big categories: process evaluation looks at the extent to which implementation comports with prescribed guidelines, and outcome evaluation looks at the extent to which a policy achieved a positive change in the problem being addressed. Ethical concerns arise during evaluation because evaluation fundamentally depends on measurement, a tricky business that involves resolving what measures should be used, whether it is possible to measure everything that matters, and the extent to which observed effects were within the government’s control.

ETHICS AND POLICY ANALYSIS

The bearing of ethics on analysis deserves specific comment because the central objective of policy analysis is to improve public policy, an inherently normative target. Policy analysis directly intervenes in the policy-making process at several junctures. It is the set of formal methods by which public problems, their various alternative policy solutions, and the outcomes of chosen solutions are understood. Often these methods attempt objectivity—that is, they involve systematically gathering data, employing reliable methods to analyzing it, and choosing the best possible option according to specified criteria. Yet, the reality that emotions, values, beliefs, capabilities, and other mitigating contingencies enter into any diagnostic process means that policy analytical endeavors have embedded non-rational components that challenge their ostensible objectivity.

Over time, the field of policy analysis has undergone dramatic transformation to contend with the role of values in social scientific study. Policy analysis first emerged as a self-conscious discipline in response to the demands of the Planning, Programming, and Budgeting System in the Department of Defense after Word War II. These origins engendered a rationalist analytical approach that presumed that logical reasoning and scientific methodologies could yield

optimal decisions. The field matured as its role expanded throughout the branches and levels of government, as well as beyond the public sector, so that policy analysts were asked to respond to a diverse set of demands and decision makers across society.^[9] With such diversity came recognition of the role of values as an important influence on the policy analytical and policy-making process. As Dunn noted, “The same policy-relevant information may be interpreted in markedly different ways, depending on the assumptions contained in the frame of reference, theory, or ideology of policy analysts and other policy stakeholders.”^[10]

Inevitably, disagreements about values arise in the policy process. Some argue that because values claims cannot be demonstrated empirically, and merely communicate preferences, they cannot enter rational debate. Others, however, hold that policy debates and analysis can and should be founded on explicit statements of ethical rules and moral principles, although debates about the legitimacy of value choices can only be resolved politically. Sound contemporary policy analysis thus attempts to use evidence and logic to choose the most favorable alternatives, but recognizes that public policy decisions cannot be divorced from the ideologies, politics, values, or environment in which these decisions are set, or from the moral perspective of those who make them. Thus, policy analysis and ethics should go hand in hand. As Brown explained, “ethics helps us identify assumptions we hold in common with others, and discard non-crucial assumptions at variance with those of others. It therefore helps us identify areas of consensus.”^[11] We can therefore identify where empirical research will not improve a given decision process because remaining disputes are over values.^[5]

ETHICAL BEHAVIOR BY POLICY MAKERS

Finally, a word about the behavior of those who practice policy making. One reason that ethics is central to the public policy process is that citizens expect decision

Table 1 Internet resources

Organization	Web site
American Society for Public Administration Section on Ethics	www.niu.edu/~tp0dcm1/aspa/ethicsec/index.htm
<i>Journal of Power and Ethics: An Interdisciplinary Review</i>	www.spaef.com/JPE_PUB/index.htm
Public Integrity	www.niu.edu/~tp0dcm1/PI/index.htm
International Institute for Public Ethics	www.iipe.org
Association for Practical and Professional Ethics	www.php.indiana.edu/~appe/home.html
Ethics Updates	www.ethics.acusd.edu/index.html
U.S. Office of Government Ethics	www.usoge.gov/

makers who are in the privileged position to analyze, make, and implement policy to “do the right thing.” In other words, as Tong explained in her discussion of ethical behavior by policy experts, with privilege comes a requirement for trustworthiness.^[7] Even analysts, who may not wield final decision-making authority, have power derived from their presumed expertise and thus have some responsibility for policy consequences.

This requirement can be onerous when a public official is asked to serve numerous constituents with varying needs, consider the implications of policy for third parties (present now and in future generations), evaluate data and findings susceptible to alternative interpretations, and balance conflicting values. On top of these demands, Weimer and Vining argued that policy makers and analysts must also uphold their own moral principles—as they explain it, individuals must be true to their own personal conception of “the good society.”^[12] In short, the broad and common teleological perspective in public policy requires public officials to weigh these demands simultaneously, and to manage continuously outcomes, circumstances, and values throughout the policy process. This multiplicity of constraints on behavior creates what Frankena called a “conflict of duties.”^[13]

Many organizations and associations have developed codes of ethics designed to prescribe behavior in ways that assist actors in managing conflicts of values, objectives, and interests (Table 1). Such sets of rules are difficult to apply universally and are subject to individual interpretation, but they can provide guidelines for how to handle common ethical dilemmas. They also serve as the formal representation of consensus within a particular agency or field about the goals or ideals to which members aspire and the standards of conduct that all agree constitute professional behavior. Moreover, as Patton and Sawicki argued, the exercise of developing ethical codes has utility because it can help to clarify moral principles and shed light on ethical dilemmas, and because it

helps to keep policy makers conscious of the ethical choices at hand.^[5]

REFERENCES

1. Lasswell, H.D. *Politics: Who Gets What, When, How*; McGraw-Hill Book Company, Inc.: New York, 1936.
2. Gerston, L.N. *Public Policy Making: Process and Principles*; M. E. Sharpe, Inc.: Armonk, NY, 1997; 6–8, 95–119.
3. Stone, D. *Policy Paradox: The Art of Political Decision Making*; W. W. Norton and Company: New York, 1997; 99–103.
4. Swain, J.W.; Duke, M.L. Recommendations for research on ethics in public policy from a public administration perspective: barking dogs and more. *Int. J. Public Adm.* **2001**, *24* (1), 126.
5. Patton, C.V.; Sawicki, D.S. *Basic Methods of Policy Analysis and Planning*, 2nd Ed.; Prentice Hall: Englewood Cliffs, NJ, 1993; 37, 39–42.
6. Cooper, T.L. *The Responsible Administrator: An Approach to Ethics for the Administrative Role*; Associated Faculty Press, Inc.: Port Washington, NY, 1982; 12 pp.
7. Tong, R. *Ethics in Policy Analysis*; Prentice Hall: Englewood Cliffs, NJ, 1986; 81–105.
8. Rohr, J.A. *Ethics for Bureaucrats*; Marcel Dekker, Inc.: New York, 1978; 6–7.
9. Radin, B.A. *Beyond Machiavelli: Policy Analysis Comes of Age*; Georgetown University Press: Washington, DC, 2000; 9–54.
10. Dunn, W.N. *Public Policy Analysis: An Introduction*, 2nd Ed.; Prentice Hall: Upper Saddle River, NJ, 1994; 126 pp.
11. Brown, P.G. Ethics and education for the public service in a liberal state. *J. Policy Anal. Manage.* **1986**, *6* (1), 62.
12. Weimer, D.L.; Vining, A.R. *Policy Analysis: Concepts and Practice*, 3rd Ed.; Prentice Hall: Upper Saddle River, NJ, 1999; 45–47.
13. Frankena, W.K. *Ethics*; Prentice Hall: Englewood Cliffs, NJ, 1963; 2 pp.

Ethics in Organizations, Implementation of

Evan M. Berman

*Public Administration Institute, Louisiana State University,
Baton Rouge, Louisiana, U.S.A.*

INTRODUCTION

Organizations undertake many activities that aim to increase the ethical conduct of employees and to ensure that ethical principles are reflected in their programs and policies. Although no organization can prevent isolated incidents of unethical conduct, much can be done to systematically improve ethical conduct, thereby increasing stakeholder trust and, in some instances, protecting organizations from legal or moral culpability. The following discusses major elements of an ethics implementation strategy.

RECENT DEVELOPMENT OF ETHICS IMPLEMENTATION EFFORTS

Concern for ethics in public administration greatly increased during the post-Watergate period in the late 1970s. Although initially focused on individual ethics (specifically, criminal acts), efforts soon led to the passage of new or more stringent ethics laws. Many of these new laws focused on conflict of interest situations that involve the acceptance of gifts, employment (limiting “revolving door” and lobbying practices for those who recently left a government office), and financial disclosure. The federal government and many states established or increased their government ethics office to oversee the implementation of ethics laws.^[1]

During the late 1970s and early 1980s, many observers also called for an aspirational ethics strategy that aimed to promote public-sector values. This aspirational strategy complements the legal strategy that focuses on minimizing wrongdoing. During the late 1980s, many organizations began to develop or emphasize codes of ethics that identified their aspirational values. Typically, codes of ethics discuss organizational values related to democratic governance, the public interest, personal and professional integrity, having respect for others and a constructive attitude, transparency and inclusiveness in decision making, and effective and efficient services.

Despite the growing importance of ethics in the 1980s, few organizations had a well-defined ethics implementation strategy in place. Merely having a code of ethics does not guarantee ethical conduct.

During the 1990s, codes of conduct were developed that detailed specific conduct that should or should not occur. Training was increasingly used to communicate these standards to employees and to help them address a broad range of common but ethically troubling workplace situations. Facilitation was also used to help groups of employees develop their own codes of ethics and conduct. In some organizations, ombudsmen were also used. Today, these tools are beginning to be integrated with other, established practices, such as hiring and leadership, to provide a systematic ethics implementation strategy that involves assessment, training, and reinforcement.

ETHICS ASSESSMENT

If the purpose of ethics implementation is to increase ethical conduct, or support existing ethical conduct, then managers need to know the current state of ethics in their organizations. An ethics climate survey is a systematic, written survey that probes existing employee knowledge of ethics standards, as well as the presence of conduct that may be ethically troublesome. Such a survey supplements managers’ subjective assessment that may be based on their intuition and the unethical acts of a few. Typically, all employees or a sample of employees are asked to voluntarily participate in an ethics climate survey. At this time, climate surveys are diagnostic, only; they have not been psychometrically validated against a broad sample of organizations. The results of ethics climate surveys can also be used as an element of training (below), when results are discussed with work groups.

For example, employees are asked to evaluate the following questions on a seven-point Likert scale that ranges from Strongly Agree–Strongly Disagree (Ref.^[2]). This instrument contains 43 items):

“Employees maintain the same ethical standard even when no one is watching.”

“My supervisor encourages me to act in an ethical manner.”

“My department has a defined standard of integrity.”

“You can rely on the accuracy of the organization’s information about what will or will not happen.”

“Members of my department have not misused their positions to influence the hiring of their friends or relatives in government.”

IMPLEMENTING ETHICS THROUGH LEADERSHIP

Many studies support that top leaders are examples of moral conduct for others in organizations. When top leaders are seen to espouse the highest values of virtue, and are strong advocates that these values are adopted by others and incorporated into decisions and organizational systems (such as human resource management), then ethical conduct is readily promoted. Conversely, the popular expression “rot starts at the top” identifies the corrupting impact of top leaders on organizational culture when ethics is not espoused in word and deed. When top managers are perceived as getting away with or condoning wrongdoing, then calls for increased ethical conduct by others ring hollow and will have little impact.

It is currently believed that top leaders should develop ethics priorities for their organization. These include a variety of standards and specific actions, such as regarding client relations, program development, accounting, financial disclosure, personal responsibility, and so on. The above ethics climate survey can assist top managers in deciding which ethics items should be priorities. When top managers make ethics standards important priorities for lower managers, a cascading effect occurs when lower managers advocate the same standards in their units, setting examples for their employees and supervisors. This effect is reinforced by informal and formal structures of accountability; for example, managers may be asked to ensure that specific ethical conduct is discussed with employees, or that certain ethical standards be reflected in program decisions.

For example, one such way in which ethics is promoted is by making ethics a criterion in hiring and promotion. The background check of candidates includes questions about their ethics judgment and examples of exemplary ethical behavior and ethical misconduct, if any. During job interviews, candidates are given scenarios for analysis that involve matters of ethics, such as potential conflicts of interest (for example, accepting minor gifts). They may also be asked about their support for ethics training and other implementation activities and their experience with making ethics a priority in their units. Such ethical standards and practices in hiring increase the likelihood of obtaining candidates with ethical behavior. Disseminating these standards sends a signal to existing employees who hope for promotion that ethics is an important priority for their organization.^[3]

ETHICS TRAINING

In recent years, many professional associations published new ethics training materials for their members. These materials included seminar training manuals, ethics readers, and ethics self-tests in paper and electronic formats (for example, CD-ROM). Ethics training is one of several ways through which top leaders indicate their commitment to ethics. The purpose of many training packages is to increase ethical behavior through new insight and understanding. Ethics training is typically conducted in small groups, which promote insight through feedback and participation by all. Ground rules typically include that everyone should participate, and that comments stay in the room, which promotes confidentiality.^[2]

Ethics training efforts typically begin with a definition of ethics as a standard of conduct. A code of ethics and conduct captures the important standards of an organization. Then, a distinction between ethical and legal conduct is made. For example, leaving work a few minutes early each day is legal but unethical. The importance of ethical conduct is linked to heightened trust, morale, and productivity. Cases are widely used to highlight the complexity of many ethical dilemmas, to discuss proper ways of dealing with ethical complexity, and to heighten awareness for unethical temptations and various warning signs (for example, “no one will ever know...” or “I deserve a little extra...”).

For example, trainees are asked to discuss the following situation. A very competent employee of the agency has quietly been using agency resources, including office supplies, copies, phone, vehicles, and health supplies, to operate a small, private business. Over a period of several years, this has resulted in the misuse of several thousands of dollars in agency resources. Is there an ethical problem? What should you do?^[4] This case is often adapted in the following way: What if a competent worker has used office supplies for incidental, personal use. Is there an ethical problem? What should you do?

Ethics training not only teaches employees to identify and deal with ethically compromising situations, it also provides them with a general framework for dealing with ethical issues. Typically, such frameworks begin with fact gathering, preparing a statement of ethical principles, preparing a statement of legal issues (if any), identifying consequences, examining options, and analyzing of options.

REINFORCEMENT

Ethics is reinforced by providing feedback to employees about their conduct. Feedback can be used

to reinforce the ethics priorities of organizations. Theories of motivation suggest that the impact of feedback on behavior is most effective when feedback is perceived as being swift, certain, and significant. Two types of feedback are distinguished: feedback that aims to eliminate (discourage) unethical behavior, and feedback that aims to encourage ethical behavior.

Many public organizations have officials and bodies whose job includes investigating wrongdoing. Federal agencies often have Inspectors General (IG) who provide a broad range of audits, include those relating to fraud and abuse. IGs operate independently of agency executives. Many states have Ethics Commissions whose job is to oversee the implementation of ethics laws and to investigate allegations of ethical wrongdoing. The U.S. Office of Government Ethics oversees the implementation of ethics laws in the federal government and provides regulatory and compliance information to agencies; many large federal agencies have their own ethics offices. Collectively, IGs, Ethics Commissions, and Government Ethics Offices might be thought as “ethics police.” In addition, some agencies have ombudsmen who also oversee ethics complaints, and many professional associations provide ethics advice and sometimes investigate ethics complaints against their members.

Consensus exists that ethics policing is only somewhat effective in reinforcing ethics. Although the work of IGs and Ethics Commissions on individual cases typically is thorough in its fact finding and unbiased in its conclusions, ethics policing is neither swift nor certain. Many complaints take years to investigate. The number of complaints, and those that are investigated, are quite small. Investigations are often limited to allegations of legal wrongdoing. Thus, the presence of an ethics police does not appear to be much of a deterrent force for those inclined to commit unethical acts. Rather, the rigor with which individual organizations deal with unethical acts may be a greater deterrent. When managers pursue allegations and take a proactive stance in their investigations, the effect on conduct may be more certain.

Similarly, managers can reinforce positive ethical behavior by acknowledging and rewarding it. This can be done formally and informally. In some organizations, performance appraisals now include a category that allows managers to reward such behavior. Ethical actions can also be the basis for evaluating employee performance awards. In addition to these formal sanctions, managers can also informally acknowledge ethical conduct to individuals. Recognition is likely to be more effective when it is timely, certain, and significant.

IMPLEMENTING ETHICS IN SPECIAL SITUATIONS

Many work situations involve special attention to ethics, a few of which are discussed below. Readers interested in ethics applications in their areas can contact their professional associations which sometimes have publications or training materials relevant to ethics. They can also obtain information from the references below and from Internet sites that deal with ethics.

Law enforcement, for example, involves concerns for the rights of alleged offenders and their victims. This has received heightened interest in the wake of publicity related to wrongdoing. Those in police custody deserve to be treated with respect and dignity, regardless of race and alleged crime. Law enforcement also raises concern about bribery and criminal wrongdoing by officers. Prolonged exposure to criminal activity can have a corrupting impact on those who come in contact with it. Thus, ethics implementation in law enforcement requires the development of ethical standards that deal with these problems, and implementation through training, policing, reinforcement, and top leadership.^[5]

Health care settings require consideration for patient rights. Patients may be quite vulnerable financially, emotionally, and physically when they seek treatment. Ethical treatment of patients requires full and timely disclosure of diagnosis and medical options. Doctors have an obligation to avail patients of their options, including second opinions. Then, nursing of patients requires attention to various side effects of treatments, including emotional consequences.^[6]

City management has its own unique challenges. It is increasingly common for public organizations to employ both spouses, which may be necessary when recruiting scarce talent. This raises issues regarding supervision and promotion, especially when one of the spouses is the city manager. City management also involves ethical challenges of dealing with elected officials, such as requests for personal services. In this regard, ethics training of elected officials is now increasingly mandatory for newly elected officials.^[3]

Finally, in some developing countries, corruption and graft have become accepted ways of dealing with government officials. This is furthered by extraordinarily low compensation structures for public officials. Ethics implementation in these settings requires not only an increase in ethics policing but also greater transparency in the interactions of employees and citizens. This is achieved through, for example, increased job rotation as well as publication (transparency) of exemptions of delinquent taxes and regulatory requirements that often are the object of graft and corruption.^[7]

INTERNET RESOURCES

A variety of ethics-related web sites are available. A good starting point is the Ethics Section of the American Society for Public Administration at <http://www.aspanet.org> or <http://www.niu.edu/~tp0dcm1/aspa/ethicsec/index.htm>. The Ethics Section Newsletter, *Ethics Today*, regularly publishes a list of ethics-related web sites (see, for example, the Winter 2000 issue, pp. 4–5). Some useful web sites are as follows (all accessed June 8, 2001):

Ethics Officer Association: <http://www.oea.org>.

Ethics Updates (Material for Instructors and Students): <http://ethics.acusd.edu/index.html>.

International City/County Management Association: <http://www.icma.org>.

Federal Inspectors General: <http://ignet.gov>.

Freedom to Care (Ethics in Policing): <http://www.freedomtocare.org/page24.htm>.

Organization for Economic Cooperation and Development, Anti-Corruption Unit: <http://www.oecd.org/daf/nocorruption/index.htm>.

National Institutes of Health: <http://ethics.od.nih.gov/>.

State of Wisconsin Ethics Board: <http://ethics.state.wi.us/>.

U.S. Department of Defense Standards of Conduct Office: http://www.defenselink.mil/dodgc/defense_ethics/index.html.

U.S. Office of Government Ethics: <http://www.usoge.gov>.

REFERENCES

1. Cooper, T., Ed.; *Handbook of Administrative Ethics*; Marcel Dekker, 2001.
2. International City/County Management Association. *Ethics in Action*; ICMA, 1999; 86–88.
3. Berman, E., West, J., Bonczek, S., Eds.; *The Ethics Edge*; ICMA, 1998; 227–236.
4. Mertins, H.; Bruke, F.; Kweit, R.; Pops, G. *Applying Standard and Ethics in the 21st Century*; American Society for Public Administration, 1998.
5. Kleinig, J. *The Ethics of Policing*; Cambridge University Press, 1996.
6. Spencer, E. *Organization Ethics in Health Care*; Oxford University Press, 2000.
7. Organization for Economic Cooperation and Development. *No Longer Business As Usual: Fighting Bribery and Corruption*; OECD, 2000.

Ethics in Public Administration, Teaching of

Donald Menzel

Division of Public Administration, Northern Illinois University, DeKalb, Illinois, U.S.A.

INTRODUCTION

The teaching of ethics in public administration is a multifaceted and often controversial enterprise. It is multifaceted because the field of public administration ranges broadly within and across organizations, nations, and cultures. It is often controversial because there is little agreement on what to teach and how to teach ethics. Indeed, some persons believe that ethics cannot be taught to others in a traditional class or course context. Rather, the best that can be hoped for is to teach *about* ethics. Still, there is widespread view among practicing public administrators and educators in the United States and abroad that an ethical public service is essential to a well-functioning democracy. According to this view, the teaching of ethics to men and women who occupy positions of public trust should and must be pursued, regardless of the arguable outcomes.

ETHICS AND PUBLIC OFFICE HOLDING

“When a man assumes a public trust, he should consider himself a public property,” President Thomas Jefferson admonished. Yet, there are legions of stories in various countries where these words have fallen on deaf ears. In Jefferson’s age, it was often believed that men of character were the essential foundation for building a democratic and just society. Today, it has become abundantly clear that, although character qualities are important among the men and women who conduct the nation’s public affairs, they have not proven to be reliable qualities.

President John F. Kennedy, in a message to Congress on April 27, 1961, also recognized the importance of recruiting men and women of sound character to public office when he said, “the ultimate answer to ethical problems in government is honest people in a good ethical environment.” Furthermore, we “must develop in all government employees an increasing sensitivity to the ethical and moral conditions imposed by public service.” In other words, recruiting honest people to public service is highly desirable, but is not enough. Thus, we have here a statement about teaching ethics—it should increase a person’s “sensitivity to the ethical and moral conditions imposed by public

service.” This is not so arguable, and it is doable; that is, it does not get mired in the quandary of teaching character traits that could presumably change one’s character.

SENSITIVITY AND AWARENESS TEACHING

A sensitivity and awareness approach to the teaching of ethics in public administration has followed two primary paths. The first path is legalistic and is often reflected in state-level ethics commissions that provide advice and instruction to state and local public employees and elected officials on the “dos and don’ts” of state ethics laws. Such laws typically emphasize conflicts of interest, financial disclosure, and confidentiality of information. This “how to stay out of trouble” approach is also embraced by the U.S. Office of Government Ethics. The bad news about this approach is that it is often reduced to the level of “if it’s not illegal, it’s okay!” This approach, as John Rohr aptly reminded us in *Ethics for Bureaucrats*,^[1] is the “low road” to public service ethics.

The second path is semilegalistic in its focus on professional codes of ethics or agency rules of acceptable behavior. At the professional association level, for example, nearly every public service association has adopted a code of ethics that its members are expected to support. Two professional associations are illustrative in this regard—the American Society for Public Administration (ASPA) and the International City/County Management Association (ICMA). The ASPA is a 10,000-member organization consisting primarily of educators, students, and public employees drawn from local, state, and federal agencies, as well as members of non-profit associations. The ASPA code, which was adopted in 1984, identifies five key principles: 1) serve the public interest; 2) respect the Constitution and the law; 3) demonstrate personal integrity; 4) promote ethical organizations; and 5) strive for professional excellence. Members who violate the Code can be expelled from ASPA, but there has been a noticeable absence of such actions taken (see <http://www.aspanet.org/>).

The ICMA is a 7500-member organization consisting primarily of practicing public managers in cities and counties in the United States and abroad (e.g.,

Ireland, Australia). The ICMA code dates to 1924, and provides specific guidance on acceptable and unacceptable behaviors of local government managers. For example, it is deemed unethical for a city manager to leave the management post with less than 2 years' service, unless there are extenuating circumstances such as severe personal problems. It is also viewed as unethical for a city manager to publicly endorse a commercial product that a vendor might sell to a city. The ICMA actively enforces the code with a half-dozen or more members sanctioned nearly every year for violations (see <http://www.icma.org/go.cfm>).

Agency rules that define acceptable behaviors and practices of public employees are common place among local governments in the United States. In Kansas City, Missouri, for example, the *Ethics Handbook* issued to all city employees asserts that it is unacceptable behavior to spend several hours a week of city time to download Internet information on a relative's medical condition (see <http://www.kcmo.org/index.htm>). In California, many local governments place restrictions on their employees receiving gifts. In San Diego (<http://www.sannet.gov/>), for example, employees "shall not accept gifts, gratuities, or favors of any kind which might reasonably be interpreted as an attempt to influence their actions with respect to city business." The Town of Los Gatos' (<http://www.los-gatos.org/>) administrative regulations are even more restrictive. For example, one rule states that "no employee shall accept money or other consideration or favor from anyone other than the town for any reason."^[2]

The teaching of ethics based on codes or administrative rules of behavior such as those above stress the contents of the codes or the rules themselves, which unfortunately can sometimes become ends themselves. The teaching of codes and rules are often conducted

by training personnel within a governmental agency, management consultants, and college and university instructors, especially those in graduate degree granting programs that prepare men and women for public service careers.

TEACHING MORAL REASONING

Another approach to the teaching of ethics in public administration is moral reasoning. This approach, which is prominently advocated by Terry L. Cooper in successive editions of *The Responsible Administrator*,^[3] presumes that one can learn to reason through a difficult moral or ethical dilemma. Stated differently, learning how to be ethical in public service is just that—a learning process that, when reality sets in, can be applied with desirable outcomes.

The reasoning process set forth by Cooper puts the accent on decision-making through ethical reflection based on the interplay of moral rules, ethical principles, self-appraisal, and justification. His ethical decision-making model is shown in Fig. 1. At the heart of this exercise is what Cooper calls exercising one's moral imagination to sort through right or wrong decision outcomes.

Another proponent of teaching ethical decision-making is Carol W. Lewis who, in *The Ethics Challenge in Public Service*,^[4] presents the reader with a problem-solving guide. This guide engages the learner with real and hypothetical decision-making scenarios, self-assessment tools, and questions that stimulate ethical reflection. It contends that neither the "low" road of compliance nor the "high" road of integrity are realistic guides for navigating the often stormy political and bureaucratic environments of public service.

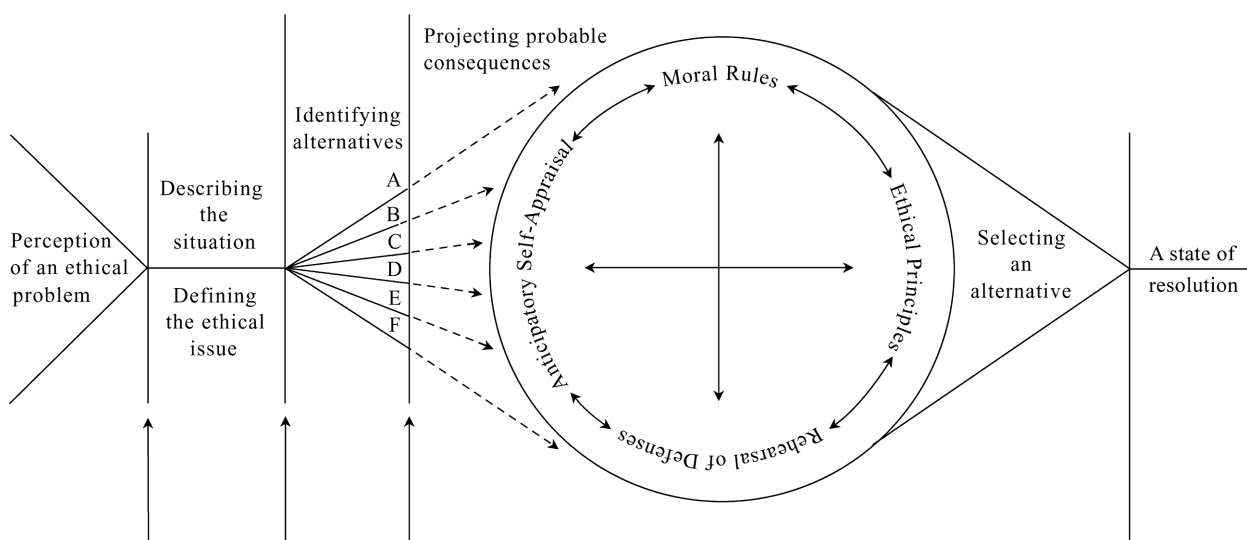


Fig. 1 Ethical decision-making model.

Rather, it is necessary to develop a two-pronged, systematic approach that incorporates both the path of compliance with formal standards and the path of individual integrity. She labeled this approach the “fusion route” to meeting the ethics challenge in public service.

How does one learn to engage in moral reasoning? Related, how can ethical decision-making be taught? The answer here is practice, practice, practice. That is, one can learn how to reason and make ethical decisions by practicing; the learner can engage himself with decision dilemmas and work through them. A teacher of ethics can use decision-making scenarios and small-group processes to help the learner practice ethical decision-making and acquire skill in doing so. This approach has much in common with virtue ethics espoused by Aristotle in the Age of Antiquity. Aristotle believed that one could only acquire a virtue by engaging in virtuous acts. But, he was wise to add, one does not acquire a virtue by engaging in foolhardy acts. Jumping into a lion’s cage to acquire the virtue of courage is not what he had in mind!

LEADERSHIP AND EXEMPLAR TEACHING

A third approach to the teaching of ethics in public administration centers on leadership and exemplars in public service. This approach has had a time-honored tradition in the U.S. military academies and is increasingly reflected in the curricula of professional public administration graduate schools that award the Masters’ of Public Administration degree. Some schools, such as the LBJ School of Public Affairs at the University of Texas, have established (2001) a Center for Ethical Leadership, which they believe will attract men and women who have a strong desire for public service leadership responsibilities (see <http://www.utexas.edu/lbj/research/leadership/>).

The study of leadership, of course, is rather wide reaching, and includes business, industry, political,

and educational sectors. Interestingly, however, the study and teaching of administrative leadership has been problematic as Larry Terry notes in *Leadership of Public Bureaucracies*.^[5] Several factors have contributed to this situation—the complexity of modern public organizations including the growing interdependency of private and public sector organizations, the antibureaucratic ethos that permeates American politics, and the challenge of distinguishing administrative leadership from political leadership.

Nonetheless, there is increasing attention given to this important subject and in linking it to the behavior of exemplars such as those found in Terry Cooper and Dale Wright’s edited volume *Exemplary Public Administrators*.^[6] Their book examines the life and career histories of nearly a dozen public administrators who have successfully resolved challenging ethical dilemmas. Students and teachers who use this material are able to discuss the life histories and experiences of these exemplars, relate them to their own life experiences, and therefore engage in ethical self-reflection.

In summary, Table 1 illustrates the key features of the three approaches to the teaching of ethics discussed above.

TEACHING FUTURE PUBLIC SERVANTS

The teaching of ethics in public administration described above, and other approaches detailed in James S. Bowman and Donald C. Menzel’s book *Teaching Ethics and Values in Public Administration Programs*,^[7] are promising beginnings. That more effort and results are needed is without dispute. Perhaps most revealing are the findings reported by Paul C. Light in his book *The New Public Service*.^[8] Light conducted a study of the graduates of the nation’s leading public policy and administration programs (e.g., Syracuse, Kansas, University of Southern California, University of Michigan, Harvard). These graduates, regardless of current sector of employment

Table 1 Approaches to teaching of ethics in public administration

	Sensitivity and awareness	Moral reasoning	Leadership
Focus	Moral choice among right vs. wrong behavior	Deliberation, self-examination, decision-making	Moral choice among right vs. wrong and right vs. right behavior
Substance	Rules, standards of conduct, conflicts of interest	Dilemmas, trade-offs, experience, learning	Observation and modeling
Purpose	Behavioral and legal compliance, obedience	Personal integrity, responsibility	Norm setting and following behavior
Illustrations	U.S. Office of Government Ethics and professional association codes	Cooper, Lewis	U.S. Military Academies—Oliver North’s congressional testimony

Source: Adapted from Ref.^[4].

(government, non-profit, private), placed “maintaining ethical standards” at the top of the list of skills considered important for success in their current job. At the same time, when asked if their school was helpful in teaching skills that would enable one to maintain ethical standards, most rated their education as insufficient. In fact, Light reported that this gap between how helpful a school is in teaching ethics and how important ethics is to one’s job success was the largest of all skills listed, including such important skills as “budgeting and public finance,” “doing policy analysis,” “managing motivation and change,” and “managing conflict.”

In 1989, the National Schools of Public Affairs and Administration (<http://www.naspaa.org/>), through its Commission on Peer Review and Accreditation, incorporated language in its accreditation standards (4.21) to the effect that graduates should be able “to act ethically.” Standard 4.21 has encouraged schools to put into place ethics courses or otherwise demonstrate that they are teaching ethics across the curriculum. This tightening up of the standards is promising, but evidence regarding the outcome is skimpy. A survey conducted in 2001 by the author of more than 80 of the 143 accredited programs in the United States found that many ($n = 32$) schools have increased the emphasis placed on ethics in their curriculum, but the majority ($n = 44$) have not. Some schools claim they have not increased their emphasis because they are already giving considerable attention to ethics—a claim that is difficult to validate. A sizeable majority of schools ($n = 59$) believed that Standard 4.21 is sufficient, although some express concern about its implementation and enforcement by the Commission on Peer Review and Accreditation. Only a handful of schools claim that they have developed outcome measures of their graduates ability “to act ethically.”

In conclusion, the teaching of ethics in public administration is a diverse, dynamic, and challenging enterprise that could be described as a “cottage” industry. However, there is evidence that this industry will grow in the decade ahead in both the United States and internationally. A chapter in an edited volume by Terry L. Cooper, *Handbook of Administrative Ethics*,^[9] tracked the emergence of administrative ethics as a field of study in the United States since the 1950s and leaves little doubt that more, not less

attention will be devoted to the teaching of ethics in public administration.

Internationally, there is also considerable movement as emerging Third World democracies struggle for economic and political independence, and as nations like Russia attempt to shake loose from the vice of corruption and embrace the rule of law. The United Nations (UN) has also increased its efforts to lend a helping hand. *Public Service in Transition*,^[10] a report prepared by the UN Division of Public Economics and Public Administration, emphasized “the critical importance of probity and integrity” of governments worldwide to conduct the public’s business. The UN has also developed an impressive web site called UNPAN (<http://www.unpan.org/>), which provides valuable advisory and training resources that can be drawn on for the teaching of ethics in public administration (see also <http://www.unpan.org/training-professionalism.asp>). The teaching of ethics in public administration is a robust and future-shaping enterprise.

REFERENCES

1. Rohr, J. *Ethics for Bureaucrats*; Marcel Dekker, Inc.: New York, 1978; 352 pp.
2. Simmons, C.W.; Roland, H.; Kelly-DeWitt, J. *Local Government Ethics Ordinances in California*; California State Library, California Research Bureau: Sacramento, CA, March, 1998.
3. Cooper, T.L. *The Responsible Administrator*, 4th Ed.; Jossey-Bass Publishers: San Francisco, 1998; 278 pp.
4. Lewis, C.W. *The Ethics Challenge in Public Service*; Jossey-Bass Publishers: San Francisco, 1991; 228 pp.
5. Terry, L.D. *Leadership of Public Bureaucracies*; Sage Publications, Inc.: Thousand Oaks, CA, 1995; 207 pp.
6. Cooper, T.L., Wright, N.D., Eds.; *Exemplary Public Administrators*; Jossey-Bass Publishers: San Francisco, 1992; 352 pp.
7. Bowman, J., Menzel, D., Eds.; *Teaching Ethics and Values in Public Administration Programs*; State University of New York Press: Albany, NY, 1998; 352 pp.
8. Light, P.C. *The New Public Service*; The Brookings Institution: Washington, DC, 1999.
9. Cooper, T.L., Ed.; *Handbook of Administrative Ethics* 2nd Ed.; Marcel Dekker, Inc.: New York, 2001; 763 pp.
10. United Nations, Division for Public Economics and Public Administration. *Public Service in Transition*; United Nations: New York, 1999; 164 pp.

Ethics, Feminist Perspective on

Dana Burr Bradley

*Department of Political Science and Gerontology, University of North Carolina at Charlotte,
Charlotte, North Carolina, U.S.A.*

INTRODUCTION

Since at least the 1980s, feminist philosophers and ethicists have been debating new ground in ethical theory. Feminist ethics criticizes the normative analysis of issues and concepts concerning right action, the qualities of humanitarian good, and social justice. Many different versions of feminist ethics exist, and eight major strains of feminist theory (liberal, Marxist, radical, psychoanalytical, socialist, ecological, phenomenological, and postmodern)—each with its own epistemology, ontology, and ethical implications—have been identified.^[1]

DOMINANT THEMES OF FEMINIST ETHICS

Feminist ethics criticizes the gender blindness and biases that characterize much of traditional ethical theory. For example, traditional ethics has tended to assume a public-private dichotomy, according to the primary characteristics considered to be part of the “moral point of view” of universality and impartiality, that corresponds implicitly to institutions and activities dominated by men. This has led to the questioning by feminist philosophers of the opposition between reason and emotion presumed by much traditional ethical theory. In response, the field has offered important accounts of the role of emotions in moral reason.

Feminist ethics has developed new theories and concepts that are more gender sensitive, and has worked to conceptualize issues of right action, social justice, and the human good from out of the more gendered experience of diverse groups of women. Each major strain of thought—liberal, Marxist, radical, psychoanalytical, socialist, and postmodern—attempts to address the causes and solutions its framework identifies for the subordination of women.

Liberal Feminist Thought

Liberal feminist theory traces its roots to Wollstonecraft’s *A Vindication of the Rights of Woman*^[2] and in Mill’s *The Subjection of Women*.^[3] In this paradigm, it is cultural constraints that hinder women from competing fully in the public world.

Traditional society excludes women from law, politics, medicine, and so forth because they are believed to be less capable. Therefore, liberal feminists endeavor to change this situation by insisting that women be accorded the same educational and occupational opportunities as men.

Marxist Feminist Thought

Marxist feminist thought promulgates that it is impossible to secure genuine equal opportunity within a class-based society. Private ownership of the means of production is responsible for an inequalitarian class system. A communist system must replace a capitalist system to ensure that no one would be economically subordinate to anyone else. Thus, men and women would be economic equals.

Radical Feminist Thought

Radical feminist thought, like Marxist feminist thought, stands in contrast to liberal feminist paradigms. Radical feminists believe that the patriarchal system so oppresses women that it cannot be reformed. Therefore, the legal, economic, political, social (the family), and cultural (church) institutions associated with a patriarchal system must be eliminated. In contrast to Marxists feminists, radical feminists insist that men, not capitalism, are women’s primary oppressors.

Psychoanalytical Feminist Thought

Psychoanalytical feminists also focus on gender and sexuality, but in a different way than that of radical feminists. This theoretical orientation seeks to explain differences in male and female sexuality by using Freudian-based analyses of infant or childhood experiences. Psychoanalytical feminists seek to rewrite the Oedipal drama by promoting dual parenting and dual career by heterosexual couples. This paradigm views the raising of children by both men and women as essential to promoting the universality of human values. Following this precept, children would no longer view the values of authority and autonomy as male the values of love and dependence as strictly female.

Socialist and Postmodern Feminist Thought

The liberal, Marxist, radical, and psychoanalytical perspectives are not the only ones offering paradigms for feminist ethics. The socialist feminist perspective tries to weave a coherent ideology from each of the other major perspectives. Juliet Mitchell^[4] argued that four structures overdetermine women's existence: production, reproduction, sexuality and the socialization of children. Woman's status and function in all these components must change if true equality is to be realized by women. Alison Jaggar^[5] built upon this perspective by using the concept of alienation to explain how, under a capitalist system, everything and everyone can be both a source of a woman's integration and disintegration as a person. Postmodern feminists have in turn challenged these efforts as "phallogocentric" thought and dismissive of the true diversity (from a class, culture, and race perspective) of women.

ETHICS OF CARE

The most famous of the feminist paradigm is the ethics of care first proposed by Carol Gilligan^[6]—and pursued and refined by other philosophers^[7]—in part as a response to the androcentric ethics of justice model that they view as the traditional approach to framing ethical issues. It is most closely allied with the liberal perspective, drawing upon the importance of culture in framing the experience of woman.

Noting that many feelings are as widely held as valid as rationality, the feminist care framework calls for using the women's perspective in approaching ethical issues by incorporating intuition and emotion along with reason. Feminist moral theory stresses the value of incorporating context, practical considerations, and relationship into one's ethical relationship.^[8] In feminist ethics, intimate contexts are seen as important as more global ones, because universal principles do not in actuality apply to all circumstances or individuals. Gilligan posited the mother-child relationship as the model for ethical decision making. Key elements include an emphasis on self-sacrifice for the good of the other, a concern for the well-being of the whole person, and a desire to nurture individual development and independence.

Although Gilligan's argument for gender differences received widespread attention, the assertion of systematic sex differences has not held in subsequent research.^[9,10] However, the ethics-of-care paradigm offers a contrast to the bureaucratic approach to social policy and public administration. In the latter, focus is on rules, control, formal lines of authority, and ensuring relative uniformity in service delivery. A feminist ethics-of-care paradigm promotes attention to the

individual, as well as to family, context, consideration of the uniqueness of the particular individual, and seeking solutions that fit specific circumstances rather than applying standards.

Historically, the success of bureaucratic organizations has been evaluated in terms of their precision, continuity, and technical utility.^[11] With its historical emphasis on the abstract agent, public administration leaves little room for emotion, ambiguity, and social relationships.^[12] Thus, organizational goals emphasize efficiency and effectiveness, and put less weight on context and relationships.

The traditional approach to public administration and public policy emphasizes a values-neutral perspective, and reliance upon structures and rules. This means that the application of established principles increases the likelihood that decision making is clear and applied in an unbiased manner. The downside to this approach is that strict adherence to rules in decision making may not fully account for the complexity of the problem and tends to discount the needs of marginalized populations.

The feminist ethics-of-care approach provides a different perspective by emphasizing the inclusion of many voices in the decision process. By valuing relationships, the feminist ethics of care connects the decision maker and those affected by the decisions. This leads to a viewing of more complex and contextual factors that are then explicitly incorporated into the decision process. Most important, it stresses the culture of caring along a dynamic continuum.^[13]

Feminist care ethics is not a panacea. Because this paradigm encourages the development of caring relationships within the formulation and administration of policy, there is a demonstrable problem with ensuring equality of attention, service, and care. This issue leads to an important ethical question regarding the delivery of services to individuals with established relationships to decision makers over those who are perhaps less well known, but have more pressing needs.

An additional disadvantage to the feminist care approach is that, with its emphasis on meeting individualized needs, it can theoretically be highly labor intensive or time consuming. Issues relating to accountability can also be a drawback. Most programs and services have developed recognizable benchmarks, related to measuring the program's structures or policies. A feminist care approach would challenge this assessment strategy, relying more upon an individualized approach to both service delivery and quality measurement.

Some who have assessed feminist care ethics have argued that, although it may be a powerful ethical approach for one-to-one or community-level issues, it is a less useful guide for public administration or public policy due to its emphasis on the personal and

individual contexts and its concern with individual relationships.^[14]

Liedtka^[15] described a caring organization as one in which networks of ongoing personal relationships are formed by “connected selves.” Derived from concepts in feminist ethics, she suggested caring organizations are ones that focus on persons, are ends in themselves instead of a means to profits, are essentially personal, and are growth enhancing for the members by developing people to their full capacities. Liedtka, and Dobson and White,^[16] suggested that organizations that exhibit these caring tendencies will have better economic performance than organizations based on other foundations. Dobson and White specifically suggested that a caring organization will be a place where trust will flourish because trust is required for personal relationships: “One needs to be trusting if one sees oneself as interdependent and connected.”^[16]

The feminist ethics paradigm, with its focus on contextual sensitivity, flexibility, client autonomy, and individuality, adds important elements to debates in public administration or public policy. Health care, in particular, with its current emphasis on productivity, concern for market dynamics, and a disconnect between decision making and patient input, would benefit from incorporating a feminist ethics perspective.

CONCLUSION

Bringing feminist ethical theory to reflect on specific social policies and on the development of public administration theories is productive for both theory and practice by raising important questions for each. Do feminist ethical critiques and positive concepts suggest directions for policy other than those currently on the public agenda? A feminist approach to policy and administration asks about the differences between men and women that “so-called neutral” policy evaluation has failed to highlight. Normative approaches to the gendering of policy often reveal the need for more contextual applications of principles and methods for understanding the plural identities that characterize most people’s lives. Such normative approaches also tend to show that simple alternative between

formalized equality and special treatment, and/or allowing or forbidding actions often do not adequately address feminist policy issues.

REFERENCES

1. Tuana, N., Tong, R., Eds.; *Feminism and Philosophy*; Westview Press: Boulder, CO, 1995.
2. Wollstonecraft, M. *A Vindication of the Rights of Women*; Poston, C.H., Ed.; W. W. Norton: New York, 1975; reprint.
3. Mill, J.S. The subjection of women. In *Essays on Sex Equality*; Rossi, A.S., Ed.; University of Chicago Press: Chicago, 1970; 184–185.
4. Mitchell, J. *Psychoanalysis and Feminism*; Vintage Books: New York, 1975.
5. Jaggar, A.M. *Feminist Politics and Human Nature*; Rowman and Allanheld: Totowa, NJ, 1983.
6. Gilligan, C. *In a Different Voice: Psychological Theory and Women’s Development*; Harvard University Press: Cambridge, 1982.
7. Hoddings, N. *Caring: A Feminine Approach to Ethics and Moral Education*; University of California Press: Berkeley, CA, 1984.
8. Held, V. *Feminist Morality: Transforming Culture, Society and Politics*; University of Chicago Press: Chicago, IL, 1983.
9. Jurkiewicz, C.L.; Massey, T.K., Jr. The influence of ethical reasoning on leader effectiveness: an empirical study of non-profit executives. *Non-profit Manag. Leadersh.* **1998**, 9 (2), 173–186.
10. Walker, L. Sex differences in moral reasoning. In *Handbook of Moral Behavior and Development, Vol. 2: Research*; Jurtines, W., Gewirtz, J., Eds.; Erlbaum: Hillsdale, NY, 1991.
11. Dill, A. Defining needs, defining systems: a critical analysis. *Gerontologist* **1993**, 33, 453–460.
12. Stivers, C. Toward a feminist perspective in public administration. *Women Polit.* **1990**, 10 (4), 49–55.
13. Tronto, J.C. *Moral Boundaries: A Political Argument for an Ethic of Care*; Routeledge: New York, 1993.
14. Carse, L.A.; Nelson, H. Rehabilitation care. *Kennedy Inst. Ethics J.* **1996**, 6 (1), 19–35.
15. Liedtka, J.M. Feminist morality and competitive reality: a role of an ethic of care? *Bus. Ethics Q.* **1996**, 6 (2), 182–194.
16. Dobson, J.; White, J. Toward the feminine firm. *Bus. Ethics Q.* **1995**, 5 (3), 463–478.

Ethics, Governance Structures

Carole L. Jurkiewicz

*Public Administration Institute, Louisiana State University,
Baton Rouge, Louisiana, U.S.A.*

INTRODUCTION

Organizational theorists, researchers, and practitioners have increasingly recognized the importance of ensuring behavioral compliance with ethical standards through a governance structure.^[1] A governance structure refers to an ethics infrastructure that contains the tools, systems, and conditions for motivating and enforcing high standards of conduct.^[2] These can take many forms, but all include systematized policies and procedures to impartially monitor compliance with an ethical code, ethical mission, or other form of articulated standards of conduct. Generally, the governance structure operates as a separate entity with an organizational hierarchy that protects it and the people working on behalf of it from repercussions of those it may necessarily investigate and discipline. The scope of its oversight can be organizationwide or focused on specific departments. Within its purview, it has the authority to create and enforce policies regarding the conduct of individuals from the highest to lowest levels of authority and responsibility. Effective governance structures are believed to reduce the number of ethical violations,^[3] strengthen organizational culture^[4] and managerial accountability,^[5] and enhance public trust in government.^[6]

The framework of governance structures assumes that a broad-based approach is most effective in ensuring compliance behaviors;^[7] policies and processes are thus targeted simultaneously to people, structure, and technology. Although the concept is rooted in the arguments of Friedrich^[8] and Finer,^[9] the theoretical foundations of governance structures can be traced to the early literature on organizational design^[10] and change.^[11] This systems approach was first alluded to as a method for moderating ethical behavior in public administration as early as 1955.^[12] In 1978, efforts to elaborate upon and refine governance structures garnered institutional legitimacy with the establishment of both the U.S. Office of Government Ethics (see <http://www.usoge.gov>) and the Office of Inspectors General (see <http://www.ignet.gov/>). Calls for substantive innovation in governance structures continue, notably with Truelson's argument for institutional controls;^[1] the emphasis on establishing appropriate ethical climates^[4,13] and systematized values;^[14] and

the introduction of transparency systems.^[15] One of the most active areas of implementation and innovation is currently in municipal governments in Europe and Africa (see <http://www.oecd.org/>).

FRAMEWORK OF ETHICS GOVERNANCE STRUCTURES

The necessary underpinning of governance structures is to ensure that the ethical standards are clearly communicated and understood.^[16,17] This includes dissemination of materials, training, updating information as appropriate, and creating accountability systems. Efforts are made to reinforce the importance of this message to the same degree in all levels of the organization.^[18] In addition to articulating the standards, specifying the repercussions of violating these standards in terms of organizational policies and legal frameworks is essential. Although laws are inflexible and generally fall short of specifying desired behaviors, they are important safety nets in governing structures.^[2] Accountability mechanisms should provide adequate control, while allowing for reasonable flexibility. Ethical wrongdoing is frequently contributed to by too many and by rules that are too complex and difficult to comprehend and follow, even for those whose intentions are to do so.^[2] Employees should understand what is expected of them in terms of ethical behavior, in terms of the scope of their decisional boundaries, and in terms of consequences if those standards are violated. Disciplinary sanctions typically range from warnings and reprimands to financial penalties and dismissal. Disciplinary measures are expected to be timely, fairly enforced, and open to public scrutiny. It is sometimes advisable to involve external institutions in cases involving individuals or issues that may be particularly intractable.

It is expected that the ethical standards communicated should be demonstrated and reinforced by officials in both their day-to-day activities, the reputation they strive to uphold, and the policies they support on behalf of the organization. Implicit and explicit commitment to ethics should be evidenced in both the spirit and the letter of their management philosophies.^[3] Particular attention is generally given to issues relating

to the use of official information and public resources, receiving gifts or benefits, and interactions crossing the public-private sector boundaries. More detailed oversight is generally given to individuals in positions that are historically more susceptible to corruption, such as procurement and human resources.^[19] Such oversight may include regulations and guidelines specific to these functions, stricter controls, and regular redeployment. Managers' decision-making processes, financial interests, and potential conflicts of interest should be transparent, and available for public scrutiny facilitated by democratic processes and with legislative oversight.^[20] This pattern of openness, referred to as transparency systems,^[15] becomes increasingly important as one ascends the hierarchy. The importance of disclosure systems, as well as a free and independent media, are essential elements in the governance structure.

Mechanisms for detection and independent investigation of wrongdoing need to be reliable, consistently employed, and subject to the same ethical standards to which the employees themselves are being held.^[21] In addition, citizens and employees should be made aware of their rights and obligations if they are a witness to wrongdoing or if they are being pressured to participate in behaviors that violate the ethical standard. Generally, the obligation involves formal legislation or organizational rules that require reporting of misconduct.^[2] Tangential concerns involve ensuring confidential methods of communication are in place, and it is clear to whom the employee should turn in such a circumstance. Safeguards to protect the whistleblower from negative repercussions from revealing such wrongdoing are of course integral to such efforts.^[22]

FUTURE DIRECTIONS IN GOVERNANCE STRUCTURES

Integration of governance structures with other organizational functions and externally with professional associations represent an area of growing interest. An overview of these various initiatives provides insight into the extent to which governance structures are integral to successful management of ethics in public administration, although no organization currently evidences all these in union.

Anticipating and avoiding ethical problems by inclusion of governance structures in organizational and departmental strategic planning sessions, change initiatives, and reform efforts^[19] is an initiative gaining widespread support. Voice is thus given to considerations of the ethical consequences of policy or priority changes as part of the planning process. This integrates governance structures in the early stages of programs and can save resources in preventing unethical

outcomes through careful planning. Shifting legislation, increasing public demands and lobbying activities, shifts in public-private sectors relationships,^[23] and changes in procurement procedures are some examples of conditions that warrant ethical scrutiny in the planning phases.^[24]

The structure of the organization or department itself can be conducive to ethical or unethical behavior and as such is a focus of governance moving toward the future.^[7] Reporting relationships, accessibility to people and information, confidentiality issues, resource distribution, privacy, and special needs are all areas affected by the physical structure. Approaching the layout and organization of a department in the planning of additions or remodeling, and routinely assessing how the environment affects the communication patterns and flow of work through the system can provide keys to areas of needed intervention.^[25] Instances of wrongdoing are analyzed in terms of where the individuals who violated the standards were physically located. Perhaps more or less privacy is needed in that area, or more secure locks or other arrangements need to be provided. Organizational structure is viewed as a preventive measure of governance structures.

Creating human resource policies and procedures that reinforce and reward ethical behavior is another new area of focus in the governance process. Communicating expectations for ethical performance in the job description, recruiting literature, the interview process, and new employee information packets is emphasized. Incorporating these behavioral expectations in measurable form in employee performance reviews and in exit interviews^[26] is equally important in reinforcing the type of ethical behavior expected in the workplace. Individuals selected for rewards or promotions are those who represent in practice and philosophy the ethical standards supported by the organization.^[26] Offering organizational rewards for outstanding skill in managing the ethical aspects of public service is another method of reinforcing desired behaviors. At the core of effective preventive elements of governance structures is to create outcomes for being ethical that outweigh those for being unethical.^[27] Of course in organizations where corruption is rampant, enforcement is more important in the short run, but in organizations where instances of misconduct are minimal, prevention is a more valuable method of governance. Prevention purportedly is less expensive in the long term, has a more positive impact on the culture of public service, and reduces cynicism toward corrupt public service by a civil society.^[28]

More broadly within the human resource function, attention to confidentiality of records, guarded access to employee and insurance information files, equanimity of treatment in evaluating applications and training managers on the ethics of interviewing and

hiring protocol are given equal attention. The way in which an employee is recruited, hired, and socialized into an organization or department establishes a self-governing standard that can supplement the institutions' ethical governance measures.^[29] Monitoring of electronic and voice communications is an increasingly common form of governance activity.^[30,31] Employees are generally made aware of this policy and practice upon employment, and are asked to sign a statement acknowledging familiarity with this policy. It is also important to review personnel policies and procedures annually, and to report the assessment's findings publicly as part of the transparency structure.^[32] The effort spent in this regard represents the balance between prevention and enforcement; sound governance policies include both.

An independent ethics forum is another form of effective governance. Such forums are generally informal luncheon sessions that meet regularly on a weekly or biweekly basis, depending on interest. Invitations to attend are distributed via an intranet messaging system or through memos. Attendance is voluntary, and a culture of open discussion is established whereby ideas or concerns related to ethical issues faced on the job or in theory are put on the table for discussion. All attendees have an equal voice in discussing the issue, generally with an unbiased facilitator present. No records are kept of the meeting and those attending are asked to keep all conversations confidential, a request enforced through organizational sanctions. Attendance by individuals up and down the hierarchy is encouraged, and having representation from the highest levels of management offers validation and perceived importance to the group. This also demonstrates that ethical issues occur at all levels in various forms, and knowing that individuals are not isolated in such occurrences has been shown to facilitate their seeking external advice and supporting transparency systems. A culture of openness is established and safeguarded. Such forums allow for individuals to practice the language of ethical discussion, become aware of the complex issues that usually surround a concern, and provide practice in resolving and addressing ethical dilemmas that carries over into daily practice. These interactions tend to instill confidence in attendees that they can deal with a seemingly insurmountable ethical challenge and that there are objective methods in doing so. These forums are not intended to absolve managers, departments, or organizations from ethical oversight, but rather to act in concordance in support of their efforts.

Phone and electronic hotlines for whistleblowers are a relatively new development in supporting governance measures. These are generally monitored by an outside agency or organization that does not trace the source of the information and passes along the information with all identifying traces removed. If a reward system

is in place to encourage whistleblowing, this agency generally provides a code number to the informant. The information along with the code number is reported to the organization. If, after investigating the issue and determining it was a legitimate concern, the reward is made available at a bank where the informant need only present the preassigned code, with no other identifying information. Notice of the award is posted in company newsletters or on web sites, and the informant usually has 6 to 12 months to claim it. The outside agency chosen for this initiative must be monitored for strict confidentiality compliance procedures. There is some debate over the efficacy of such hotlines,^[33] and results are inconclusive.

Professional associations or unions are another element in establishing governance structures for ethical behavior. The ethical codes they develop and support, the importance given a standing ethics committee, and the strength with which they communicate and enforce professional ethical standards can work in concert with intraorganizational systems governing ethical behavior. However, tensions can result over a conflict between association codes of ethics and governmental codes of ethics. This is often the case with bar associations' codes regarding confidentiality, although attorneys working for the federal government are obligated to protect government interests over such individual obligations. Effective role modeling can evolve from these opportunities to share ideas in professional forums. Training by ethics professionals external to the organization frequently occurs at such meetings and can be a valuable aspect of membership supported by the institution. Individuals can be asked to share these materials upon their return, becoming in turn recognized for their enhanced expertise in a specific area.

Internal controls such as annual or semiannual reviews leading to recommendations for improving management processes or procedures can directly inform the organization and general public of steps to strengthen governance structures. The goal of governance structures is to keep public servants accountable for their actions and to remove opportunities for ethical compromises to occur. Independent scrutiny such as legislative reviews, regular management audits, inspectors general or ombudsman investigations, and specific judicial and ethics reviews are increasingly popular methods of oversight.^[28]

Governance structures themselves are most often subject to independent scrutiny as a monitor of their due process and compliance with the same ethical standards they are charged to enforce. Bias, regardless of whether intentional, can be revealed by overt reviews for fairness and objectivity, as well as audits of the process and procedures for handling complaints. External reviews not only instill confidence in the governance

structure by those subject to them, but they also encourage double-loop learning on the part of those charged with ethical oversight. As a consequence, they are more likely to review their own behavior and the patterns in that behavior and adjust accordingly, knowing that authoritative others will be viewing their actions from that framework.^[34,35]

REFERENCES

- Truelson, J.A. New strategies for institutional controls. In *Ethics Frontiers in Public Management*; Bowman, J.S., Ed.; Jossey-Bass: San Francisco, CA, 1991; 225–242.
- Organisation of Economic Co-operation and Development. *Ethics and Corruption*; 2001. <http://www.oecd.org/puma/ethics/infras.htm>.
- Jurkiewicz, C.L. The trouble with ethics: results from a national survey of healthcare executives. *HEC Forum* **2000**, *12* (02), 101–123.
- Menzel, D.C. The ethics factor in local government: an empirical analysis. In *Ethics and Public Administration*; Frederickson, H.G., Ed.; M.E. Sharpe: Armonk, NY, 1993; 191–204.
- Foster, G.D. Legalism, moralism and the bureaucratic mentality. *Public Pers. Manage.* **1981**, *10*, 93–97.
- Garment, S. *Scandal: The Culture of Mistrust in American Politics*; Time Books: New York, 1991.
- Cooper, T. *The Responsible Administrator*; Jossey-Bass: San Francisco, CA, 1998; 4 pp.
- Friedrich, C.J. Public policy and the nature of administrative responsibility. *Public Policy* **1940**, *1*, 3–24.
- Finer, H. Administrative responsibility in democratic government. *PAR* **1941**, *1* (3), 335–350.
- Weber, M. *The Theory of Social and Economic Organizations*; Free Press: New York, 1947.
- Lewin, K. *Field Theory in Social Science*; Harper & Row: New York, 1951.
- Wood, R.C. Ethics in government as a problem in executive management. *PAR* **1955**, *15* (1), 1–7.
- Bruce, W.M. Ethics education in municipal government: it does make a difference. In *Teaching Ethics and Values in Public Administration Programs*; Bowman, J., Menzel, D., Eds.; State University of New York Press: New York, 1998; 231–249.
- Van Wart, M. *Changing Public Sector Values*; Garland Reference Library of Social Science, Garland: New York, 1998; Vol. 1045.
- Gilman, S. Realignment and public sector ethics: the neglected management problem in the new public administration. Organisation of Economic Co-operation and Development, 2001. <http://www.oecd.org/puma/ethics/symposium/gilman.htm>.
- Hejka-Ekins, A. Teaching ethics in public administration. *Public Adm. Rev.* **1988**, *48*, 885–891 (September/October).
- Darley, J.M. How organizations socialize individuals into evil-doing. In *Codes of Conduct*; Messick, D.M., Tenbrunsel, A.E., Eds.; Russell Sage Foundation: New York, 1996.
- Bruhn, J.G. Being good and doing good: the culture of professionalism in the health professions. *Health Care Manag.* **2001**, *19*, 47–58.
- Lewis, C.W. *The Ethics Challenge in Public Service*; The Jossey-Bass Public Administration Series; Jossey-Bass: San Francisco, CA, 1991.
- Hays, S.W.; Gleissner, R.R. Codes of ethics in state government: a nationwide survey. *Public Pers. Manage.* **1981**, *10*, 48–58.
- Adams, J.S. Inequity in social exchange. In *Advances in Experimental Social Psychology*; Berkowitz, L., Ed.; Academic Press: New York, 1965; 2 pp.
- Bowman, J.S. Whistle-blowing in the public service: an overview of the issues. *Rev. Public Pers. Adm.* **1980**, *1* (1), 15–27.
- Goodman, J.B.; Loveman, G.W. Does privatization serve the public interest? *Harvard Bus. Rev.* **1991**, *69*, 28.
- Jurkiewicz, C.L.; Thompson, C.R. Conflicts of interest: organizational vs. individual ethics in healthcare administration. *J. Health Human Serv. Adm.* **2000**, *23* (1), 100–123.
- Daft, R.L. *Organization Theory and Design*; South-Western College Publishing: Cincinnati, OH, 1998; 6 pp.
- Jurkiewicz, C.L.; Giacalone, R.A. Healthcare administrative ethics in a global context. *J. Public Aff. Inf.* **2001**, *5* (3), 1–18.
- Klingner, D.E.; Nalbandian, J. *Public Personnel Management*; Prentice Hall: Upper Saddle River, NJ, 1998.
- Organisation of Economic Co-operation and Development. *Building Public Trust: Ethics Measures in OECD Countries*; 2000, PUMA Policy Brief No. 7 <http://www.oecd.org/puma/>.
- Chao, G.T.; O'Leary-Kelly, A.M.; Wolf, S.; Klein, H.J.; Gardner, P.D. Organizational socialization: its content and consequences. *J. Appl. Psychol.* **Oct.**, **1994**, *79* (5), 730–744.
- Prince, M. Employers should establish clear rules on e-mail. *Bus. Insur.* **2001**, *35* (19), 25.
- Scott, A. No privacy in the workplace. *Intern. Aud.* **2001**, *58* (3), 15–16.
- Carter, M. *Complete Guide to Ethics Management: An Ethics Toolkit for Managers*; Wallace and Pikel: London, 1999.
- Moody, M.S.; Serepca, B. Inspector general hotlines: have they been good? at what cost? what is lost? *J. Public Inquiry* **1998**, *4* (1), 9–12.
- Argyris, C.; Schoen, D.A. *Organizational Learning II*; Addison-Wesley: Reading, MA, 1996.
- Jurkiewicz, C.L.; Knouse, S.B.; Giacalone, R.A. When an employee leaves: the effectiveness of clinician exit interviews and surveys. *Clin. Leadership Manage. Rev.* **2001**, *15* (2), 97–100.

Evolutionary Theory and Public Policy

Steven A. Peterson

School of Public Affairs, Pennsylvania State University at Harrisburg,
Middletown, Pennsylvania, U.S.A.

INTRODUCTION

For some 25 centuries, the “nature of human nature” has been the central issue in Western political thought. Fully aware of that issue’s crucial importance, many of our greatest political philosophers made it their explicit point of departure. Although not always as readily recognized, much the same has been true in the formulation of public policy. With few exceptions, the policies adopted in a given polity reflect and flow from the views about human nature held by those charged with setting those policies. The reason is quite simple. Almost invariably, the measures proposed to deal with social, political, and economic problems stem from some (often unarticulated) conception of human nature and a corresponding set of assumptions about the likely consequences of governmental efforts to encourage, redirect, limit, or even prohibit a given type of social, political, or economic behavior.

For at least the past 50 years, the so-called standard social science model (SSSM) has dominated the behavioral sciences and directly and indirectly influenced public policy.^[1] The SSSM insists that for all practical purposes, human nature—and thus human behavior—is shaped by *culture*. The SSSM rests on three cardinal tenets—two of them explicit, the third usually implicit. These are: 1) that humans have no innate behavioral tendencies; and 2) that, consequently, human nature is solely the product of learning and socialization (in short, of “nurture”); from which it follows 3) that human nature (and consequently human behavior) is essentially quite malleable.^[2]

Long the reigning paradigm, SSSM has recently been challenged by an evolutionary model (EM) that challenges the tenets on which the SSSM is based. The EM agrees that culture *is* important—but so are the genetically transmitted behavioral tendencies that our species has evolved over millions of years (“nature,” of course). Given this reality, *both* nature *and* nurture must be taken into account if we are to understand why humans act as they do. The idea that our evolutionary past may sometimes significantly influence how we act today leads to a distinct conception of the wellsprings of human behavior and of human nature itself; in turn, this often leads to different conclusions about the issues that governments should address and of the policies best suited to deal with

those issues. Two key questions emerge from the preceding general comments: First, how does the EM paradigm differ from that of the SSSM in its understanding of human nature? Second, what are the implications of these differences for the formulation of public policy? Extant literature suggests the role of biological factors in public administration^[3,4] and policy.^[5–7] This article builds on such prior research.

KEY DIFFERENCES IN UNDERSTANDING HUMAN NATURE

The two models differ in several respects in their respective approaches to human behavior and nature. Among these are: 1) the role of culture in shaping behavior; 2) genetic factors as behavioral influences; and 3) the relevance of primatological data. Not surprisingly, these conflicting views eventuate in a basic parting of the ways with regard to the malleability of human nature—and of human behavior.

We will very briefly examine these in the order listed.

The Role of Culture in Shaping Behavior

The importance of culture as a factor in shaping human behavior is probably the most commonly misunderstood aspect of the EM. As noted above, the SSSM holds that for all practical purposes, human behavior is the product of culture; the EM differs significantly in that it looks as well at the extent to which behavior might also be influenced by our evolutionary legacy as social primates. But far from denigrating culture, the EM stresses that ours is the only species that has evolved the capacity for developing anything more than a very rudimentary culture. Even more to the point, it emphasizes that *Homo sapiens* is the only species capable of creating and, in many circumstances, then acting in accordance with beliefs that actually *run counter* to some very powerful innate behavioral tendencies.

The point need not be belabored. One model holds that to understand human social behavior, we should look only at culture; the other contends that we should—in fact, must—also consider the possibility

that evolution has endowed us, *as it has all other species*, with innate propensities which often influence that behavior.

Genetic Factors as Behavioral Influences

In light of what has already been said, the discussion here can be quite brief. Beyond doubt, culture plays a very important role in shaping how we act—but so do genetically transmitted behavioral tendencies evolved over literally millions of years. Some of these are probably unique to our species—language, culture (in any “rich” sense), and, surely unique, often a willingness to sacrifice ourselves in behalf of some religious or secular belief.

However, there are other proclivities that we share with most of our fellow primates. Among these are hierarchical social structures characterized by dominance above and submission below, status seeking, aggression, selfishness, an aversion to celibacy and an often-reluctant and unreliable commitment to monogamy,^[8] xenophobia, ethnocentrism, and nepotistic favoritism. These behaviors have become part of our genetic legacy because they have served on balance, to maximize the “inclusive fitness” of our ancestral fellow speciates—and the fellow speciates from which they, in even earlier times, evolved.

Use of Primatological Data

The divergence between the two models on this point is eminently reasonable. If, on the one hand, human behavior is shaped almost entirely by culture and socialization, information about other species, no matter how closely related biologically, serves no useful purpose since it does little or nothing to further our understanding of human behavior. On the other hand, if our genetic legacy meaningfully influences how we act, then the study of closely related species (and especially the apes) that share much of that legacy becomes important to consider.^[9,10] Boyd and Silk^[11] make clear the value of understanding primate behavior for understanding human behavior. In short, primate data may well have relevance for human policy choices.

Reflections

The difference here between the two approaches is not black or white, all or none. Within each camp, there are gradations of belief and, no doubt, considerable overlap of opinion. The fundamental disagreement between the two models is on the factors that shape, if they do not actually determine, human nature. Given their divergent positions on this issue, other differences are almost inevitable. For the SSSM, human nature

and thus human behavior is essentially the product of culture—and it is to culture alone (i.e., nurture) that the behavioral sciences should attend. For the EM, both culture and genetically transmitted behavioral tendencies (i.e., nature) are operative and, consequently, both need fall within our purview.

They differ on the nature of human nature. With its emphasis on culture, the SSSM comes close to the Lockean *tabula rasa*^[12] with humans as blank slates upon which culture and experience inscribe their dictates. For the EM, evolution has endowed our species, as it has all other forms of life, with behavioral proclivities (ranging from aggressiveness to xenophobia); culture may strengthen these proclivities; on occasion it may override them; but in either event they significantly influence our actions.

Closely related to this, of course, is the question of human malleability. Those who believe in the SSSM are much more optimistic about the possibility of altering human behavior than are advocates of the EM who, as did Edmund Burke, give greater weight to the inherent limitations of human nature and of the human mind.

As might be expected, these conflicting conceptions of human nature and human behavior yield dissimilar approaches to social, economic, and political problems. The SSSM, more sanguine about the prospects of appreciably modifying human behavior and our capacity to direct and control the course of events, is more inclined to think in terms of solutions; the EM, in contrast, tends to see problems more in terms of constrained options, difficult trade-offs, and uncertain outcomes. It would be remarkable if the policies advocated by the devotees of one of these models were not frequently at odds with those favored by partisans of the other.

POLICY IMPLICATIONS

One can argue that an EM yields a much more realistic understanding of human behavior than does the SSSM—although this latter model has long dominated our approach to political, social, and economic problems.^[17] We explore this issue through illustrative case studies.

Prostitution Laws

One case study of the different take on policy from an EM as opposed to SSSM perspective is prostitution laws. McGuire and Gruter^[12] note the following:

1. Males will continue to seek sexual partners, including prostitutes, as a result of evolved predispositions.

2. Females will exploit such males to receive resources, probably as a result of evolved predispositions.
3. Laws prohibiting prostitution are likely to fail, given that such prohibitions conflict with evolved predispositions.

As a result of these conclusions, the authors suggest that a form of legalized prostitution makes the most sense, using the EM. The SSSM assumes that aversive legal sanctions would deter people from engaging in prostitution or soliciting the services of prostitutes. However, McGuire and Gruter contend that this is unlikely to succeed. Better to use our understanding of human nature and create laws that meet societal goals with some reasonable degree of success. They say (p. 38):

Given these points, limited legalization seems the best evolutionary-based response. Limited legalization (e.g., registration of sex workers, regular health check-ups) has the social advantage of reducing the transmission of diseases, perhaps reducing criminal involvement, and bringing sex workers into the work force...

Thus, here is a case study that has EM proposals moving in quite a different direction than standard SSSM proposals as adopted in the United States.

Drug Laws

Blank^[13] has advanced an argument that brain chemistry raises questions about the SSSM logic underlying aspects of our drug laws.

1. Addictive behavior regarding drugs is easily explained by brain neurochemistry, including the use of alcohol and nicotine.
2. The findings, largely on drug addiction, may be generalizable to other addictive behavior, such as gambling, aggressive behavior, and the like.
3. Standard assumptions that willpower is enough to control addictive craving are probably inaccurate.

Blank, at one point, notes the following policy implication (p. 68):

This evidence also has implications for drug policy that makes distinctions among potentially addictive substances. Arguments for legalization of marijuana become more difficult to accept given this evidence....

Moreover, society's treatment of nicotine and alcohol may have to be modified if consistency is sought.

Laws based on assumptions of rational behavior by drug users, i.e., punishment will deter the activity, may well be mistaken, once the EM is taken into account. Given the brain's neurochemistry, sheer dependence on rational choice deterrence theory, as per the SSSM, may not be good policy.

Criminal Justice Policy

Ellis^[14] has explored the relevance of brain functioning for criminal behavior and criminal justice policy. His basic contentions:

1. Male sex hormones (androgens) affect the probability of victimizing behavior. Higher levels of androgen go with greater aggressiveness, impulsiveness, and potentially antisocial victimizing behavior.
2. Most people, with normal cortical control of many of their impulses, learn how to express their competitive/victimizing urges in ways that are not apt to elicit response from the criminal justice system; those who cannot control their impulses, on the contrary, are more likely to attract the attention of the criminal justice system as they "act out."

Policy implications? For those who cannot control the impulses generated by androgens, prescription drugs ought to be used to reduce the harmful, anti-social effects of androgens. That is, drug therapy would then reduce levels of victimizing behavior. In part, this logic emerges from a rejection of the SSSM view that deterrence can come from harsh and timely punishment. If Ellis' theory is correct, this SSSM assumption is not appropriate if we are to extirpate the victimizing behavior from individuals prone to commit such actions.

CODA

An EM, then, may yield different understandings of human behavior than does the SSSM—although this latter model has long dominated our approach to political, social, and economic problems. Of course, humans are eminently cultural beings as well as the products of natural selection. Indeed, culture is a product of evolution. This adds a complex layer on to our understanding of evolutionary-constrained policy choices, since we know that culture can override

biological impulses (as witness mass suicides, such as at Jonestown and Masada). Nonetheless, it seems clear that human evolutionary history needs to be accounted for in assessing policy choices.

REFERENCES

1. Lawrence, L.E.; Huntington, S.H. *Culture Matters*; Basic Books: New York, 2000.
2. Pinker, S. *The Blank Slate*; Viking: New York, 2002.
3. Losco, J. Biology and public administration. In *Biopolitics and the Mainstream: Contributions of Biology to Political Science*; Somit, A., Peterson, S.A., Eds.; JAI Press: Greenwich, CT, 2003.
4. White, E., Losco, J., Eds.; *Biology and Bureaucracy*; University Press of America: Lanham, MD, 1986.
5. Caldwell, L.K. Biopolitics, science, ethics, and public policy. *Yale Rev.* **1964**, *56*, 1–16.
6. Caldwell, L.K. Research in biopolicy. In *Biopolitics and the Mainstream: Contributions of Biology to Political Science*; Somit, A., Peterson, S.A., Eds.; JAI Press: Greenwich, CT, 2003.
7. Somit, A., Peterson, S.A., Eds.; *Human Nature and Public Policy: An Evolutionary Approach*; Palgrave MacMillan: New York, 2003.
8. Barash, D.; Lipton, J. *The Myth of Monogamy*; W.H. Freeman: San Francisco.
9. Stanford, C. *Significant Others: The Ape-Human Continuum and the Quest for Human Nature*; Basic Books: New York, 2001.
10. de Waal, F. *Chimpanzee Politics*; Johns Hopkins University Press: Baltimore, 1998 (Revised edition).
11. Boyd, R.; Silk, J.B. *Humans Evolved*; W.W. Norton: New York, 1997.
12. McGuire, M.; Gruter, M. Prostitution—an evolutionary perspective. In *Human Nature and Public Policy: An Evolutionary Perspective*; Somit, A., Peterson, S.A., Eds.; Palgrave MacMillan: New York, 2003.
13. Blank, T.H. Neuroscience, free will, and individual responsibility: implications for addictive behavior. In *Human Nature and Public Policy: An Evolutionary Approach*; Somit, A., Peterson, S.A., Eds.; Palgrave MacMillan: New York, 2003.
14. Ellis, L. Biosocial theorizing and criminal justice policy. In *Human Nature and Public Policy: An Evolutionary Approach*; Somit, A., Peterson, S.A., Eds.; Palgrave MacMillan: New York, 2003.

Executive Budgeting

Brendan F. Burke

*Department of Political Science, Bridgewater State College,
Bridgewater, Massachusetts, U.S.A.*

INTRODUCTION

Executive budgeting resides within a tension opposite budgetary institutions and strategies that emphasize a decentralized, legislative focus. As part of a continuum of budgetary authority, an executive budget scheme centralizes agency guidance, information gathering, and decision-making authority to one actor at the height of the organizational hierarchy. Budgeting always involves the reconciliation of diverse needs and desires with scarce resources into a single plan; the range of possibilities for budget process design parallels the grant of authority to the varied interests and their sponsors or advocates. In an authoritarian system, it is possible that the executive, monarch, dictator, or other single leader aggrandizes all power over competing interests in the budget. The American experience at all levels of government has involved give-and-take over unified executive authority and dispersed, pluralistic designs.

This article describes the original thought of the framers of the U.S. Constitution over centralized vs. dispersed budgetary authority, as well as practical developments that determined the contemporary place of executive budgeting at the subnational levels of government, in turn. The order of treatment is chronological. The Progressive reform wave in the early 20th century involved changes in executive fiscal authority initially at the local government level (around 1910). The states have modernized their administrative capacity especially since the 1950s. Budget reform in favor of the executive came in alterations to the governor's ability to prepare and present a budget, as well as in their veto powers to oppose unwanted legislative priorities. Some of the states have chosen to centralize budgetary authority in the executive's hands, while others retained a decentralized budgetary system.

HAMILTON'S VISION OF EXECUTIVE BUDGETARY POWER

The American Revolution was fought to bring the American people from a dictatorial system of resource acquisition and allocation, whereby the colonists had

little to no determining role in the scope of the Crown's taxation or in the use of fiscal resources. The state constitutions written between 1776 and 1780 renounced central authority, limiting the abilities of the governors to a custodial role except in times of war. The difficulties of coordinating state action under the Articles of Confederation led to the Constitutional Convention's focus on a balance between some centralization of executive authority, while maintaining broad representation in the national legislature. Article I, Section 8 clearly stipulated the initial dominant role for Congress in budgetary affairs, as it would raise "Taxes, Duties, Imposts, and Excises," hold responsibility for debt obligations, "coin Money," arm the military, and build post offices and roads. The president's powers included the commanding of the military, establishment of treaties, and appointment and reporting of executive agencies.

James Madison's and Alexander Hamilton's writings in the "Federalist Papers" embody the contrasting thought between shared and centralized or executive powers in all areas, including fiscal affairs.^[1] In Federalist #30, Hamilton discusses the need for national taxation and resources especially in times of war, and in #31 expands upon the national government's need for taxation beyond the states' abilities, to support other shared public priorities: "A government ought to contain in itself every power requisite to the full accomplishment of the objects committed to its care, free from every other control but a regard to the public good . . ." (193). In Federalist #70, Hamilton specifies the need for a "vigorous executive" who does not compete with equal authorities to determine the means of achieving the public good. Competing leaders will "lessen the respectability, weaken the authority, and distract the plans and operations of those whom they divide." Accountability, the assignment of blame and punishment when mistakes are made, is rendered incomplete with plural authorities.

At least at the time of the ratification of the Constitution, the support for Hamilton's thinking was limited to endorsement of a powerful central government, but not of a powerful single executive. Madison's arguments presented between Federalist #47 and #51 appear to have prevailed in 1787. In Federalist #47, Madison advocates for a division of

executive, legislative, and judicial powers to counteract the potential for tyranny. In a budgetary sense, this would mean that the legislature would determine spending plans, and the president would execute the spending. Neither party should prevail over the other, as “ambition must be made to counteract ambition.” (Federalist #51) Under the system of separation of institutional powers, Madison’s necessary safeguards are in place. But at the same time, unified accountability for spending is lost, as many individuals are responsible for the codification of expenditure needs and final approval of the budget.

THE RISE OF PUBLIC ADMINISTRATION AND EXECUTIVE AUTHORITY

Between the late 18th century and the beginning of the 20th century, there was little need for change in budgeting at the national level. The size and scope of government and its fiscal resources were limited, and the jobs of Congress and the president were only beginning to grow in complexity. At the local level, however, the rise of the political machine and the machine’s corrupt administration prompted a need for centralization and accountability in budgeting.^[2] Academics led by Woodrow Wilson and Frank Goodnow specified a stronger role for executive-focused administration, and organizations like the National Municipal League and New York City’s Bureau of Municipal Research proposed new institutional designs in favor of executive authority.

The Progressive movement put forward strains of thinking to combat corruption and inefficiency in American society, one focused on the advancement of citizen support and participation, and the other on the establishment of business-like principles in government.^[3] Woodrow Wilson’s article “The Study of Administration” and Frank Goodnow’s book *Politics and Administration: A Study in Government* emphasized the latter theme, advocating for a responsible executive authority. Wilson rendered a comparison between “the confusion, sinecurism, and corruption” in Washington, D.C., and the greater responsiveness of administrative systems in France and Prussia. Our own English roots had established “the art of curbing executive power to the constant neglect of the art of perfecting executive methods.” The balance to be struck would involve the establishment of greater administrative efficiency while still honoring the will of public opinion. Goodnow would formalize the split between a legitimizing force in political boards and bodies, who would then hand the task of carrying out their dictates to the executive. All aspects of government outside the legislative and judicial activity would fall within the administration—a broad

responsibility specified as the “executing of the will of the state.”^[4]

Specific designs for more centralized, executive-driven budgeting would fall to professional think tanks, which created reformed government structures and budget processes. The National Municipal League was formed in 1895 out of 46 “Good Government” clubs and associations to confront machine politics. Soon it unveiled its model city charter, which recommended a strong mayor-type government with a unicameral legislature and expanded executive authority. The New York Bureau of Municipal Research, under William Allen, Henry Bruere, and Frederick Cleveland, designed and advocated for consistent accounting systems as a building block for rational budgeting. The model budget that emerged was based on Frederick Taylor’s “scientific management,” whereby the most efficient allocation of resources would be used in the most effective manner, as derived through rigorous analysis of work situations. At the same time, the budget was seen especially by Allen as a tool for public education.^[5] Harkening back to Hamilton, the reformed budget design would offer greater accountability, as it would come from a single executive, clearer priorities with a more streamlined focus, and a standardized accounting language to enhance clarity within the appropriation figures. The council-manager form of government also emerged from the National Municipal League during the 1910s as a model based on communities like Staunton, Virginia, Sumter, South Carolina, and Dayton, Ohio, all of which experimented with a “business manager” design to run public affairs. The council-manager form embodied Goodnow’s politics-administration dichotomy, whereby the elected board would pass policies and the appointed administrator would hold full authority to staff and fund implementation efforts.

The council-manager form of government would spread rapidly, becoming the chosen institutional structure in almost half of American cities with populations greater than 2500.^[6] The budget process in these municipalities tends to begin with the town or city manager establishing revenue availability and important trends in expenditures, then imparting them to the guidance of service managers and agency heads for the preparation of budget requests. The managers may simply state revenue and expenditure constraints, or may restrict agencies to requests limited by a specific growth rate. Analysis of needs and requests is carried out by the manager’s personal staff, who also derive recommended funding levels for managerial review and approval. Elected officials enter the budget process either with broad input regarding their own priorities at the outset of the budget process, or when the completed budget is transmitted to the elected board for their scrutiny and adoption. Elected officials have the

authority to add and subtract priorities from the manager's budget, but their review comes near adoption deadlines, on an established and carefully crafted product.^[7] The national and state governments also have contemporary versions of executive budgeting, but they are less dominated by the chief administrative or executive officer than are the council-manager form of government.

COMPARING EXECUTIVE-FOCUSED BUDGETING ACROSS JURISDICTIONS: GUBERNATORIAL POWERS

Most of the American states have revised their constitutions extensively through the second half of the 20th century, including in ways that enhanced executive budget authority for some governors. But different states have different levels of tolerance for a stronger executive, and thus constitutional revisions on behalf of gubernatorial budgetary powers form an interesting contrast in three different areas: the governor's ability to reorganize government structure; the governor's ability to craft the budget prior to legislative consideration; and the governor's ability to veto legislative budgetary priorities. The variation may arise because different states have different political cultures, including contrasting levels of acceptance for executive

power or different emphases on broad representation. Almost all states have moved toward enhanced executive budgetary authority, but to varying degrees and at varied paces. A depiction of the range of gubernatorial powers on budgetary matters is provided in Table 1.

The states tended to be slower than municipalities and the national government at modernizing their organization structures. Beginning in 1917, the state of Illinois reshaped its functional alignment, reducing the number of agencies and creating fewer independent authorities. Many states followed suit, but the pace of change was slow. This is important to the governor's budgetary power because the governor does not have the authority to propose expenditures, or constrain them, within the agencies that serve independent boards.^[8] In some states, such as in Massachusetts, the governor can recreate the organization chart based on constitutional authority. The governor can make his or her proposal, which will pass if it sustains a veto from both branches of the legislature. But at the other end of the spectrum, states like South Carolina only allow the governor to submit their reorganization design as a regular legislative bill, subject to amendment and veto by a single house of the legislature. This is not to say that the governors of Massachusetts automatically fare better than the governors in South Carolina. In the 1980s, Governor Carroll Campbell was able to streamline a convoluted and unaccountable system of agencies with the help of his legislature. This reorganization brought

Table 1 Budgetary powers of the governors (highest to lowest power, in the order)

	Score
<i>Governmental reorganisation</i>	
Governor's reorganization power is based in constitution, with legislative veto needed by both houses	5
Based in constitution, with legislative veto needed by only one house	4
Based on statute, with legislative veto needed in both houses	3
Based on statute, with legislative veto needed by only one houses	2
Based on statute or is informal, subject to normal legislative bill consideration	2
<i>Budget development</i>	
Governor has full responsibility, legislature may not increase executive budget	5
Governor has full responsibility, legislature can increase by special majority vote or subject to item veto	4
Governor has full responsibility, legislature has unlimited power to change executive budget	3
Governor shares responsibility, legislature has unlimited power to change executive budget	2
Governor shares responsibility with other elected officials, legislature has unlimited power to change executive budget	1
<i>Veto power</i>	
Governor has an item veto and a special majority vote of the legislature is needed to override a veto (three-fifths of legislators elected or two-thirds of legislators present)	5
Governor has item veto with a majority of the legislators elected needed to override	4
Governor has item veto with only a majority of legislators present needed to override	3
Governor has no item veto, with a special legislative majority needed to override a regular veto	2
Governor has no item veto, only a simple legislative majority needed to override a regular veto	1

Source: "State and Local Government," www.unc.edu/~beyle, accessed 1 May 2005.

many independent commissions under the governor's control, so that they would submit their budget requests through the executive office for review and analysis.^[9] In 2003, Massachusetts Governor Mitt Romney was denied the ability to bring state spending under control (especially in the university system) because of opposition by both houses of the legislature to a major reorganization plan.^[10]

The governors operate under a similar system of separation of powers as is found in the U.S. constitution, and they tend to have the same ability to propose budgets as the president. The governor has the responsibility of writing the initial budget with the help of a staff, and the legislature has unlimited powers to change the governor's proposal. But this has not always been the case. In some states, the governor's power to write the budget was shared with the state treasurer, or even with an elected governor's council. New Hampshire still has the council, elected by district, that reviews gubernatorial appointments, but over time, the state narrowed budgetary responsibility to just the chief executive. In some ways, the New Hampshire (as well as the Vermont) governor is limited in budgetary authority because he or she is only elected to two-year terms; every other budget is overshadowed by the governor's election campaign.

While most states reformed the executive budget proposal process to be equivalent to the president's, four state constitutions grant their governors greater powers of proposal tied to an ability to limit expenditures. For example, the Maryland and West Virginia legislatures can only revise their governors' budgets downward; these executive budgets establish the ceiling on expenditures. Thus in these states the governor cannot guarantee the funding of specific priorities, but he or she alone can establish limits to the scope of government spending. In New York and Nebraska, a special legislative majority is necessary before increasing spending levels beyond the governor's budget. This grants these two governors a greater constitutional authority to constrain budget growth than in 46 states and the national government.^[11]

The veto power adds to the governor's ability to constrain expense (as well as policy excesses). North Carolina was the last state to write the gubernatorial veto power into its constitution, beginning in 1997. Forty-two state governors have the line-item veto power, granting them the ability to delete unwanted priorities from legislative appropriations. Some states, such as Wisconsin, add an additional constraint, whereby the legislature needs a supermajority (three-fifths of legislators, or two-thirds present at a given session) to override the veto. Governor Tommy Thompson, during his 14-year reign in Wisconsin, used the veto 1900 times. Not once were his vetoes overturned by the legislature.^[6] A strong veto is another

constitutional tool that gives the governor the ability to revise the legislature's budget downward, tipping the balance of budget authority toward the executive in most American states.

CONCLUSIONS

The framer's debate during the writing of the U.S. constitution, the Progressive reforms at the local level, and the contrasting powers granted to the states' governors assist in the construction of a multidimensional definition of executive budgeting. The components include:

- Accountability: a single elected or appointed administrator takes responsibility for an overall revenue and expenditure plan.
- Centrality: budget requests filter through a single department or office, such that coordination of related programs is possible and duplication of spending is reduced.
- Control: the single authority can propose and enforce limits on spending.
- Reorganization: disparate independent organizations can be unified under the single budget process.
- Regulation: the single executive can establish consistent procedures for the provision of information underlying requests, standards for review and analysis, and uniform presentation and allocation.
- Analysis: more detailed and rigorous analysis is possible if all requests are submitted to a higher standard of scrutiny; the inequity of intensively auditing some agencies, while ignoring others, is less likely to occur.
- Comparison: a single unified budget process enables decision makers to envision all parts of the organization subject to budget review simultaneously.
- Citizen control: a streamlined budget process makes it easier for citizens to understand their government and its resources, and to give credit or assign blame in the event of governmental successes and failures.

There are also advantages to a pluralistic budget process design, with regard to representation, potentially concerning discourse and transparency, and depending on the definition, possibly with regard to a decentralized vision of accountability. These debatable benefits justify the mixed budgetary authority found in the majority of American governments.

REFERENCES

1. Rossiter, C., Ed.; *The Federalist Papers*; Penguin, Inc.: New York, 1961.

2. Waldo, D. *The Administrative State: A Study of the Political Theory of American Public Administration*; Holmes and Meier: New York, 1948.
3. Stillman, R.J. *The Rise of the City Manager: A Public Professional in Local Government*; University of New Mexico Press: Albuquerque, NM, 1974.
4. Wilson, W. The study of administration. In *Classics of Public Administration*, 4th Ed.; Jay, S., Albert, H., Eds.; Harcourt Brace Publishers: Fort Worth, TX, 1997; Goodnow, F. Politics and administration. In *Classics of Public Administration*, 4th Ed.; Jay, S., Albert, H., Eds.; Harcourt Brace Publishers: Fort Worth, TX, 1997.
5. Kahn, J. *Budgeting Democracy: State Building and Citizenship in America, 1890–1928*; Cornell University Press: Ithaca, NY, 1997.
6. Smith, K.B.; Greenblatt, A.; Buntin, J. *Governing States and Localities*; CQ Press: Washington, DC, 2005.
7. Governmental Finance Officer's Association. www.gfoa.org.
8. Sanford, T. *Storm Over the States*; McGraw-Hill Company: New York, 1967.
9. Durning, D. Governors and Administrative Reform in the 1990's. *State Local Govern. Rev.* **1995**, 27 (1), 36–54.
10. Keogh, R. Meeting him halfway. *Commonwealth* **2005**, Winter, 74–83.
11. Beyle, T. The governors. In *Politics in the American States: A Comparative Analysis*, 8th Ed.; Gray, V., Hansen, R.L., Eds.; CQ Press: Washington, DC, 2004.

Executive Leadership in Local Government

James H. Svara

*Department of Political Science and Public Administration, North Carolina State University,
Raleigh, North Carolina, U.S.A.*

INTRODUCTION

Executive leadership is a complex topic in local government because of uncertainty about the scope of this type of leadership and variation in the officials to which it is assigned. This is especially true in the United States, where different forms of government are used in local governments. In simplest form, the topic encompasses the leadership provided by persons in the highest positions in the governmental structure who have responsibility for discerning the preferences of citizens and mobilizing popular support for policies, for developing proposals for what government will do, and for directing the work of the government. The citizen dimension might be considered beyond the scope of a type of leadership that is associated with activity within the governmental and organizational structure. Connection with the citizenry is, however, essential to democratic governance and creates the foundation for the internal executive functions.

Handling some or all the executive responsibilities falls to mayors (and their equivalents in counties) in all governments and in many governments to appointed chief administrators as well. Although it is natural to associate executive leadership with one person who has responsibilities that range from the electoral to the managerial aspects of leadership, in local government, executive leadership broadly defined is commonly shared more or less widely. In view of this division of leadership functions, the discussion is presented in three sections that deal with mayors who are elected executives, non-executive mayors, and appointed executives. For simplicity, and in recognition that the preponderance of the literature on leadership in local government has focused on cities, terms from city as opposed to county or other types of government are used to refer to the officials.

The division of responsibility depends largely on the form of government and, within mayor-council cities, on whether there is a chief administrator present in the governmental structure.^a Connecting with citizens

and mobilizing popular support for proposals is the responsibility of the mayor in all forms of government. Policy leadership (i.e., determining the purpose and policies of the government) is the responsibility of the mayor with approval of council in mayor-council cities, the mayor with advice of chief administrative officer (CAO) and approval of council in mayor-council-CAO cities, and the mayor, council, and city manager in council-manager cities. Finally, administrative leadership or directing the administrative apparatus and work of the staff is handled by the mayor, mayor and CAO, or the city manager, depending on the form of government.

Most cities have one executive and one mayor, but the two offices are not necessarily combined.^[1] Cities that use the mayor-council form of government vest all or some executive powers in the elected mayor, who is the political head of city government, the driving force in setting policy, and in charge of the administrative organization. The executive mayor with full powers ultimately has authority of hiring and supervising staff, formulating and (after approval by the council) expending the budget, and directing the organization, subject to the limitations set by the city charter. If the city has a chief administrator, some administrative functions are delegated to this official. In council-manager cities, the city manager—an executive appointed by the city council—makes a major contribution to policy making and exercises administrative powers. Managers have extensive contact with citizens, are somewhat attentive to popular preferences, and have subtle influence on public opinion. Still, city managers do not get directly involved in mobilizing public opinion or shaping public support, unless instructed by the council to do so (e.g., seeking to secure votes in a bond referendum approved by the council). Council-manager cities also have a mayor who is the political head of the government and presides over the city council, but usually has no powers other than those available to other members of the council.

ELECTED EXECUTIVE MAYORS

The mayor-council form of government is based on separation of powers with authority divided between

^aLargely excluded from the discussion are weak-mayor council forms in which executive functions are spread across a number of offices and boards in city government and small-town mayor-council governments in which executive functions are shared and there is sharing of tasks between elected officials who do a lot of the work of city government and part-time staff.

the executive and the legislature. The strong mayor–council version of this form has separation of powers between a mayor with extensive powers and integrated administrative control over staff, and the elected legislative body. The lines of authority for all or most departments of city government lead to the mayor's office.

In contrast, the council is confined to a more limited role in strong mayor–council cities. Even in policy making, the council is heavily dependent on the mayor for proposals and information, and it can be checked by the mayor's veto power. The mayor also occupies a favorable position for mobilizing public opinion in support of proposals. The council, however, must approve policies and can override vetoes, so there is the potential for the council to overrule the mayor if there is a large (i.e., veto-proof) like-minded majority on the council in opposition to the mayor, or for deadlock between the executive and legislature if a narrow majority opposes the mayor.

The checks and balances in mayor–council cities affect how officials relate to each other and the freedom of mayoral action. Because the purpose of offsetting powers is to permit one set of officials to hold the other in check, it is common for conflict to arise in the relationship between the mayor and the council (see Ref.^[2]). The conflict may concern policy preferences and priorities, administrative performance, and the extent of independent executive authority assigned to the mayor. Separation of powers can allow the departments of city government to play the mayor and council off against each other and develop their own base of constituency support that produces greater autonomy vis-à-vis the mayor.^[3] Strong mayors have the potential to blend the dimensions of leadership. They can tap organizational resources and mobilize public support to advance their political agenda. The executive can face the challenges of overcoming council resistance or opposition and recalcitrance from city departments, but may be able to use threats or inducements to win support.

In the political science literature, the ideal mayor in mayor–council cities is an innovator (also called an entrepreneur) who provides creative solutions to problems and pyramids resources to increase the ability to build coalitions and gain leverage over other actors.^[4,5] This mayor is effective at both initiation and implementation of policies and programs. Leadership is fashioned from a combination of formal and informal resources. The former are part of the governmental structure and official policies of city government. The latter are derived from personal characteristics or the political process. Formal resources available to mayors remain fairly constant over time, but each incumbent differs in their ability to make the best use of these resources. Still, some important

informal resources, for example, support from a political party, have been declining over time.

The terms “strong” and “weak” mayor refer to the level of authority assigned to the mayor, but there is no “pure” strong or weak mayor form. Strong and weak mayors are arrayed along a continuum. The key formal powers, which may be assigned to the executive or divided among a number of officials, are appointment of department heads, developing the budget, directing the administrative departments, and veto of council actions. The mayor's authority varies widely. Beyond the organizational leadership resources that are derived from appointment of department heads, control of other appointments can augment the mayor's leverage, and patronage appointments can reinforce party organizational cohesion and support. Most mayors, however, do not directly control large numbers of city government jobs. Civil service protection of most positions limits the jobs a mayor can dispense. Court cases since the 1980s have limited the ability of elected officials to remove, promote, or transfer staff for partisan reasons, unless employees are directly involved in providing advice on policy strategies, as opposed to policy implementation.^[6]

A 1991 survey conducted by the International City Management Association measures the variation in mayoral authority within the mayor–council form of government.^[7] In cities over 100,000 in population, two-thirds of the mayors develop the budget, and approximately 90% appoint department heads and have veto power. In cities under 50,000, the likelihood that the mayor possesses these powers decreases with declining population. Mayor–council governments of most moderately large to large cities generally correspond to the strong mayor–council form, whereas medium-size mayor–council cities are divided between strong and weak mayoralties. Small cities with the mayor–council form approximate the weak mayor–council form of government.

The formal resources of the mayor's office provide only part of the explanation of effective mayoral leadership. Even in mayor–council cities, informal resources are extremely important. Mayors such as the first Richard Daley in Chicago^[8] and Richard Lee in New Haven^[9] were noted for their ability to convert a formally weak position into one of strength. Support from a political party or community elites, strong popular backing, and a host of private backers indebted for jobs, favors, contracts, and recognition can give mayors the added political clout that can be helpful in dealings with the council and their own administration to get ideas accepted and acted upon.

Several trends in American politics have reduced these informal resources of the mayors and made it more difficult for mayors to use their influence.

First, the declining strength of political parties and the increasing independence of voting behavior by citizens remove a force that mayors could use to bind together office holders and secure support for the mayor's program. Second, there has been a splintering and thinning of power elites resulting from the breakup of large corporations, takeover of local firms by national and international concerns, increased competition and downsizing, and the move of companies to the suburbs. Third, the proliferation of interest groups and the political organization of neighborhoods have increased the range and diversity of organizations operating in local politics. Fourth, council members are more diverse in their characteristics, more activist in their orientation, and less likely to accept the leadership of the mayor out of deference to the mayor's power or common party loyalty. All these changes in city politics make it more difficult for the mayor to lead effectively. The new openness of the local electoral process and the individualism of persons elected to office diffuse the concentration of power needed for governing the city.^[10]

The mayor's performance is also affected by individual characteristics—experience; personal or occupational financial and staff support; personal attributes, such as charisma, reputation, wisdom, and commitment to the job (time, resources, energy expended); and effectiveness in dealing with the media. These factors determine how well and how fully the mayor fills the position, and how creatively the mayor exploits both formal and informal resources. The charismatic or adroit weak mayor may be able to win substantial support from the community and wring more advantage from limited powers than others have done. Ferman argued that “effective political skills” are the critical factor for strong mayoral leadership.^[11]

Effective leadership by the mayor is critical in the mayor-council city. Without it, the offsetting powers of the mayor and council can produce policy stalemate and administrative departments can resist change. The performance of mayors can be judged by their effectiveness on two dimensions: initiating policies, and getting policies implemented. The various types of mayors can be classified using this approach. If the mayor is a *caretaker* with no goals, the city will drift and be reactive when problems occur. If the mayor is a *reformer* or policy initiator but poor at getting things done or a *broker* who can arrange compromise but has a weak policy agenda, city government will lack a key element of leadership. This form functions best when the mayor is an *innovator* who can help to provide a clear direction for city government and ensure that city departments are focused on accomplishing the goals of elected officials.

Non-Executive Mayors in Council-Manager Cities

The council-manager form is based on the unitary model of organizing government. The council ultimately possesses all governmental authority, and executive authority is assigned to the city manager. There is no separation of powers or checks and balances in the system, although the form provides for specialization of roles. The council and mayor occupy the overtly political roles in government, set policy, and select the city manager who is continuously accountable to the council. The manager provides policy advice and recommendations to the council, and directs the administrative apparatus. Within this broad division of functions, there is considerable sharing, which this form of government promotes. The city manager also provides policy leadership in helping to frame the agenda of concerns that the council considers, and has latitude in the way that policy goals are converted into programs and services. The council, in contrast, has the potential to oversee the administrative performance of the city through appraisal of how the manager is doing specifically (and whether the manager will continue in office) and its ability to secure information about the performance of administrative staff in general.

The mayor is typically the presiding officer of the council and has no formal powers different from those of other council members, except for the veto power in 13% of council-manager cities. Mayors, directly elected in 62% of these cities,^[12] can be an important source of policy guidance and coordination of participants, although they rarely exercise any administrative authority. Thus, mayors in council-manager cities are not executives. They have close interaction, however, with the appointed executive and potentially affect the city manager's performance and influence.^[13] It is useful, therefore, to examine the roles of this kind of mayor as an official who helps to bridge the relationship between the appointed executive and the council.

The ideal mayor in council-manager cities is a facilitator who promotes positive interaction and a high level of communication among officials in city government and with the public, and who also provides guidance in goal setting and policy making.^[14] This type of leadership is well suited to the conditions of the council-manager city in which cooperative relationships among officials are common and the city manager provides support to the elected officials to whom the manager is accountable. Effective leadership by this kind of mayor improves the working relationships among officials, makes the form of government function more smoothly, and increases the involvement of elected officials in setting policy.

Typically, the mayor in the council-manager city is formally the presiding officer and serves as ceremonial head of the government. In addition, mayors may provide coordination and communication that helps to link the representative leadership of the council and the executive leadership of the manager. Finally, mayors may provide policy leadership and guide the work of the council. These roles are mutually reinforcing and are filled concurrently.

Facilitative leadership does not depend on a superior power position. There are resources available in the council-manager form and within the incumbent as a person to develop leadership in the areas of coordination and policy guidance. The strategic location occupied by the mayor in the center of communication channels to the council, the city manager and staff, and the public provides the foundation for effective leadership. Mayors with a clear conception of the job who use personal resources skillfully are more likely to be able to take advantage of this resource.

There is variation in the nature and scope of leadership, depending on how well the roles that make up the office are filled. Those who do not fill even the traditional roles (e.g., being an ineffective presiding officer who allows the council to flounder in meetings) could be called a *caretaker*, whereas those who fill these roles well but attend to no other are *symbolic heads*. Both the *coordinator* and *director* create an atmosphere that promotes cohesion and communication among officials and strengthens the capacity of the council to identify problems and make decisions. The coordinators, however, are not strongly associated with a policy agenda of their own, even though they contribute to fashioning and acting on an agenda as part of the council. The directors are associated with a distinctive policy agenda, although this agenda incorporates to a greater or lesser extent the views of other officials. The coordinators are effective at developing a sense of cohesion and purpose in their cities and at strengthening the policy-making process. They are not themselves, however, active policy initiators. They are more process oriented than policy oriented. The director-type mayors create an agenda in the sense that they take the initiative to put it together and their own ideas are central to it. Other officials and the public recognize this contribution and view the director mayor as a policy initiator.

In sum, although these mayors lack formal powers over other officials, they occupy a strategic location in the communication channels with the council, the manager, and the public. The moderately effective mayor goes beyond ceremonial leadership to provide effective coordination and communication and, thus, affect how the manager as executive is connected and interacts with the city council. The director-type, highly effective mayor also helps to develop a common

set of goals with wide council support that provide a framework within which the manager as executive can operate.

CITY MANAGERS AND ADMINISTRATORS

The city manager is the executive officer with extensive authority for directing staff, formulating and (after approval by council) expending the budget, and controlling operations. The manager is appointed by the council without approval by voters and serves at the pleasure of the council without term. The manager is typically the only staff member hired by the council (in some cities, the city attorney and/or clerk are selected by the council as well); all other employees are hired under the authority of the manager. The norm of the system is for elected officials to respect the insulation of staff from "political" interference.

The structure of council-manager government promotes cooperative relationships among officials. Because the ultimate control over city government lies with the council, there is less likelihood of power struggles between the council and the manager. The council and the manager do not compete for the same "rewards" from public service. Elected officials seek public support and reelection, whereas managers are concerned about how the council assesses their work and their standing and respect in their profession in order to have the option for advancement by moving to another city. Furthermore, city managers have a professional commitment to helping the council accomplish its goals; if they do not, they can be replaced. To be sure, tensions can emerge among elected officials or between them and the staff. The important difference from mayor-council governments is that, with no separation of powers, officials do not have to deal with structural factors that lead to conflict, and positive relationships are common. The approach to leadership taken by the city manager in the council-manager form can be different than that of executive mayors operating in a separation-of-powers setting.

The city manager as executive in council-manager cities is extensively involved in the policy formation aspect of executive leadership and responsible for directing the administrative organization. They are also community leaders. To an extent not found at other levels or in other forms of government, city managers along with other local government-appointed executives, such as school superintendents and directors of public authorities, are both general policy leaders and organizational directors. City managers do not report to an elected executive or go through political appointees in developing policy recommendations for elected officials; they report on performance directly to the governing board. Because of their

close working association with the residents of the communities they serve, they have a special obligation to value community leadership, as well as to preserve and protect the democratic quality of the political process as a whole. Unlike the elected executive, city managers are accountable to the entire governing board and are expected to provide leadership from a professional perspective. This does not mean that the city manager is not involved in politics, but rather that this involvement should be guided by professional considerations and standards.

City managers are potentially involved in a wide range of activities. City managers have commonly been shown to be active contributors to policy formulation, and they are responsible for directing the administrative organization.^[15] In addition, they advise the elected officials on a range of topics. Analysis of the extent to which top administrators are policy innovators, advisers to elected officials, and administrators has revealed consistency and variation.^[16] Most city managers (85%) are active as policy innovators. One-half combine this role with emphasis on administrative functions, and one in eight combine it with high involvement in advising elected officials. Over 20% are highly active across the board. The minority (about one in six) who are not active in policy innovation are not very involved in the other roles either. Thus, appointed executives tend to be actively and more or less broadly engaged in the governmental process.

City council members generally give city managers high ratings on their performance. In the national survey of council members in cities over 25,000 in population mentioned earlier, over 80% of the respondents agreed that the council and city manager have a good working relationship and that the manager does a good or very good job of accomplishing the goals established by the council.^[17] Over 70% feel that the manager is doing at least a good job in providing the council with sufficient alternatives for making policy decisions and sufficient information to assess the effectiveness of programs and services. Furthermore, 83% gave a high rating to the manager for improving efficiency, and 87% gave this rating for maintaining high standards of personal conduct for self and for staff.

In over one-half of the mayor-council cities over 2500 in population, there is a CAO^b or city administrator. The scope of the position and the duties depend on what responsibilities are assigned in the charter or by the officials who appoint the CAO. These usually include authority over implementation of programs,

day-to-day administrative concerns, and budget formulation, as well as playing an advisory role in developing other policy recommendations. It has been common to assume that the CAO is appointed by the mayor, derives their influence from the mayor, and operates within the orbit of the mayor.^[18] In a 2001 survey, in cities with a population of 2500 or more, 44% of the CAOs are appointed by the mayor with the approval of the city council, and another 39% are chosen by the city council.^b In only 16% of the cities is the CAO appointed by the mayor alone, although direct mayoral appointment is found in approximately two-fifths of the cities over 100,000 in population. When appointed by the mayor, the CAO serves at their "pleasure" and turnover is high, especially when a new mayor comes into office. When appointed or approved by the council, the CAO is similar in characteristics to the city manager.

Chief administrative officers are also supposed to bring management expertise to city government and can manifest the same characteristics as city managers. In general, however, the CAO is more likely to be recruited from within the city in which they serve and selected by the mayor for reasons other than professional qualifications, although the CAO may have extensive training as well as experience.^[19] City managers are somewhat more likely to be "careerists" who have served in other cities and aspire to move to "better" positions elsewhere. In a 1997 survey of city administrators, approximately one-fourth of the top administrators in both major forms of government are recruited from within the city, although the percentage of CAOs recruited from the city in which they serve increases with greater city population. If recruited from another local government, city managers are more likely than city administrators to have served as the chief administrator or as the assistant manager (51% vs. 36%) in their previous post. However, 17% of CAOs come from the private sector compared with 8% of the city managers.^[20]

When appointed by the mayor, the CAO is the agent of the mayor and has power proportionate to their responsibility to the mayor. The CAO is valued as the mayor's most active troubleshooter. The scope of the position and the duties depend on what the mayor assigns. The CAO's status is ambiguous because mayors have difficulty giving the CAO sufficient power to bring professionalism to the administration of city government, and CAOs have difficulty winning the confidence of the mayor. Often, mayors and department heads bypass the CAO in their dealings with each other. Thus, the CAO's subordination to the mayor compromises their policy and executive leadership. However, when the CAO is appointed or approved by the council, which is more common in smaller cities, this official follows the lead

^bCalculations by authors of data from *Form of Government Survey 2001*; International City/County Management Association, Washington, DC.

of the entire council. These CAOs are similar to city managers in their values and orientation. Still, CAOs tend to stay in their position a shorter time than city managers. In the 1997 survey of city managers and CAOs, the average number of years in the current positions is 6.9 years for city managers and 6.5 years for CAOs. In cities over 50,000, the tenure drops to 6.4 for council-manager cities and 4.3 years for mayor-council cities. In comparison to managers, CAOs are not as upwardly or geographically mobile.

CONCLUSION

City managers are executives and city administrators are partial executives who bring a wide range of professional considerations to the discussion of city government policies and practices. By raising professional considerations in policy discussions, city managers help to ensure a balanced approach to policy decisions.^c By acting out of commitment to strategic goal setting, ethical standards, and sound management, city managers also promote proactive policy making, as well as fairness and efficiency in the delivery of services and the use of organizational resources. These contributions can be combined with those of the mayor as elected executive or the mayor as shaper of the context for the appointed administrator to create a blend of political and professional leadership in the executive functions in local government.

REFERENCES

1. Svara, J.H. The embattled mayors and local executives. In *American State and Local Politics: Directions for the 21st Century*; Weber, R.E., Brace, P., Eds.; Chatham House: New York, 1999; 139–165. Portions of this article come from this chapter.
2. Svara, J.H. *Official Leadership in the City: Patterns of Conflict and Cooperation*; Oxford University Press: New York, 1990; 97–105; ch. 2.
3. Yates, D. *The Ungovernable City*; MIT Press: Cambridge, 1977.
4. Cunningham, J.V. *Urban Leadership in the Sixties*; Schenkman Publishing Company: Cambridge, 1970.
5. George, A.L. Political leadership in American cities. *Daedalus* **1968**, 97, 1194–1217 (Fall).
6. Bingham, R.D., Ed.; *Managing Local Government*; Sage: Newbury Park, 1991; 79–80.
7. Renner, T.; DeSantis, V. Contemporary patterns and trends in municipal government structures. In *Municipal Year Book 1993*; International City/County Management Association: Washington, 1993; 57–69.
8. Banfield, E.C. *Political Influence*; The Free Press: New York, 1961.
9. Dahl, R.A. *Who Governs?* Yale University Press: New Haven, 1961. Wolfinger, R.E. *The Politics of Progress*; Prentice-Hall: Englewood Cliffs, 1974.
10. Ehrenhalt, A. *The United States of Ambition*; Random House: New York, 1991.
11. Ferman, B. *Governing the Ungovernable City*; Temple University Press: Philadelphia, 1985; 202–205.
12. Adrian, C.R. Forms of city government in American History. In *Municipal Year Book 1982*; International City/County Management Association: Washington, 1982; 10 pp.
13. Mouritzen, P.E.; Svara, J.H. *Leadership at the Apex: Politicians and Administrators in Western Local Governments*; University of Pittsburgh Press: Pittsburgh, 2002; 216–217.
14. Svara, J.H. Associates. *Facilitative Leadership in Local Government*; Jossey-Bass Publishers: San Francisco, 1994.
15. Newell, C.; Ammons, D.N. Role emphasis of city managers and other municipal executives. *Public Adm. Rev.* **1987**, 47, 246–252.
16. Mouritzen, P.E.; Svara, J.H. *Leadership at the Apex: Politicians and Administrators in Western Local Governments*; University of Pittsburgh Press: Pittsburgh, 2002; 112–115.
17. Svara, J.H. *A Survey of America's City Councils*; National League of Cities: Washington, 1991; 81 pp.
18. Hogan, J.B. *The Chief Administrative Officer*; University of Arizona Press: Tucson, 1976.
19. Hogan, J.B. *The Chief Administrative Officer*; University of Arizona Press: Tucson, 1976. Nelson, K.L. Assessing the CAO position in a strong-mayor government. *Natl. Civic Rev.* **2002**, 91 (Spring), 41–54.
20. Svara, J.H. U.S. city managers and administrators in a global perspective. In *Municipal Year Book 1999*; International City/County Management Association: Washington, 1999; 25–33.

^cFor example, should a new recreation center go into a neighborhood in which residents have organized to pressure officials to select their area or into a neighborhood where the need is the greatest?

Federal Subsidies for New Entrepreneurial High-Growth Firms

Terry F. Buss

*Center for International Affairs, National Academy of Public Administration,
Washington, District of Columbia, U.S.A.*

INTRODUCTION

Only 3% to 5% of small firms account for three-fourths of jobs created in the United States.^[1] Not only do these high-growth firms create jobs, but they also are much less likely than other small businesses to fail; create considerably more wealth in the form of profits, sales, and value; pay higher wages and offer greater employee benefits; are much more likely to export products and services; and invest more in research and development (R&D). These firms also outperform the Fortune 500 companies.

Many observers believe that for every one that succeeds, many more potential and emerging high-growth firms that are worthy investment opportunities never start up or die prematurely because capital needs at earliest stages of development are unmet. For economists, capital gaps occur when worthy business opportunities go unfunded even though capital is available in the economy overall. Capital is unavailable, inaccessible, insufficient, costly, or in improper form (e.g., debt versus equity).

Believing that some potential and emerging high-growth firms worth investing in lack needed capital, some observers argue that federal action—potentially including tax incentives, subsidies, loans, equity, spending, regulation, and/or geographic reallocation—is necessary to address private capital market failure. Indeed, existing federal economic development programs are based on this market failure premise.

This article reviews empirical evidence on, and economic theory about, possible capital gaps affecting potential and emerging high-growth firms:

- During long-term, sustained growth in the 1990s, the preponderance of evidence suggests that capital gaps do not seriously affect high-growth firm formation and development.
- Some regions attract less outside capital or produce less capital from local investors because these regions

lack requisite industrial clustering and resource concentration, characteristic of regions infused with early stage investment, and produce fewer or much less attractive investment opportunities. This is not a capital gap because large numbers of worthy investment opportunities do not appear to be left unfunded.

- Some technology-based ventures—because of high R&D and startup costs, high risk, lengthy product development time, scarcity of information about potential markets, and lack of management experience on the startup team—may have difficulty accessing early stage capital. Financing these firms sometimes requires highly sophisticated, specialized, experienced, knowledgeable investors, typically venture capital fund managers, former executives in the industry, or corporations in the field.

WHAT IS A HIGH-GROWTH FIRM?

High-growth firms are those that experience rapid employment growth, usually at least 10% to 20% annually over a 5-year period. High-growth firms do not necessarily create or employ new technologies in starting up. Many well-known high-growth firms—Staples, Office Max, Office Depot, Walmart, Starbucks, Planet Hollywood, and even high-tech companies such as Amazon.com—do not use new technologies. Rather, they employ existing delivery systems in more efficient ways to meet consumer demand. High-growth companies do not need to be manufacturers. Neither are they exclusively high-technology or technology-based firms. Using Dun & Bradstreet national data for 1997, analysts discovered that all net growth in jobs derived from the service sector. In the economy as a whole, only 1 in 50 firms can be classified as high tech. High-growth firms are found in every industrial sector.

Potential high-growth firms are those in the idea, seed, or R&D stage of development. Emerging high-growth firms are those starting up or in early stages of development. Most analysts exclude firms created through mergers and acquisitions from discussions of potential or emerging high-growth firms. These activities combine workforces, rather than creating new jobs

An earlier version of this publication was presented to the U.S. Congress as a report from the Economics Division, Congressional Research Service, U.S. Library of Congress.

from scratch, and in any case, are not considered to be as early stage. Mergers and acquisitions are none the less important because they create increased funding for earliest stage firms. Large high-tech firms may reduce R&D investment, preferring to acquire it from fledgling firms, rather than develop it in-house.

Entrepreneurs, here, are people who start, or are in the process of starting, what they hope will be a high-growth firm, rather than a small business. Entrepreneurs use new methods in production or service delivery, or create new products, processes, or services. They are not small business operators.

HOW ARE HIGH-GROWTH FIRMS CAPITALIZED?

Capital flowing to high-growth firms is segmented into overlapping niches. Financing depends on levels of risk and expected return on investment, whether firms require equity or debt capital, and the role the startup entrepreneur(s) play in management.

Insider investors—the entrepreneur, friends, relatives, and acquaintances—typically are *the* major source of investment at seed, startup, and early stages. Entrepreneurs with capital are willing to take chances on their ideas; others close to them either have confidence in the entrepreneur's ability to pay them back or make their investment profitable, or are willing to gamble on the prospect. At these earliest stages, potential and emerging high-growth firms, generally require modest amounts of capital.

Angel investors are the second most important source of investment at earliest stages. Angel investors are wealthy individuals who invest in new ventures in exchange for equity. Their role in capital markets for some firms at least is best compared to baseball analogy. They function as a farm team—preparing firms for venture capital investment and eventual public ownership. Angels invest an average of \$60,000 to \$85,000 in a firm in exchange for equity. Angels often invest in consortia, averaging \$660,000 per firm. There may be as many as 250,000 angels, investing about \$10 billion in 30,000 firms annually, but estimates vary widely because there is no reliable way to count them.

High net worth individuals are generally much wealthier than angel investors and do not take an active role in growing businesses. They allow either venture capital firms or management to grow businesses. High net worth individuals may invest directly, usually from \$5 million to \$50 million, in private placements in ventures without venture capital intermediaries.

Venture capital flows to seed or startup stage ventures, but only in relatively modest amounts. Venture capital refers to capital invested through venture capital firms (actually managed funds) by large

institutional investors—pension funds, university endowments, or banks; corporations; or high net worth individuals. There is no comprehensive source of data on venture capital fund investments. Rather, individual brokerage houses and research firms track some funds but not others. There is also some overlap in tracking by different sources. Venture capital firms not only experience high returns on investment at later stages, but they also experience comparable, although lower, returns at earlier stages. Many venture capital firms generally balance out their portfolios with early- and later-stage investments. In recent years, some venture capital firms have even begun to specialize in early stages, investing as little as \$5000. Perceptions that venture capital does not flow to seed and R&D stage ventures are unfounded. At issue here is the amount of capital available from venture capitalists and number of ventures capitalized.

Initial public offerings (IPOs) generally do not capitalize potential or emerging high-growth firms. They usually capitalize young, but proven firms, trying to reach a new stage of accelerated growth for which large amounts of capital are required. The average IPO is about \$50 million.

Bank and finance company debt financing does not typically flow to early stage high-growth firms. Risks are too high, investment amounts are too small, and returns are too low. Researchers have recently observed that finance companies appear to be willing to take greater risks on startups than banks. Reasons for this have yet to be established. But even in those relatively rare cases when banks and finance companies do lend to earliest stage firms, capital may not be sufficient, sometimes causing firms to fail. Banks and finance companies do not lend at anywhere near the magnitude of angel or venture capital firms. As such, ventures requiring large amounts, but relying instead on commercial and industrial (C&I) loans, may find themselves undercapitalized and at risk of failure.

Even though commercial loans are rare for earliest stage ventures, entrepreneurs may use personal debt instruments to capitalize businesses indirectly. Second mortgages, home equity loans, borrowing against Keogh retirement accounts or insurance policies, and credit cards are commonly used. Entrepreneurs may also use consumer loans to purchase automobiles, computers, and homes for use in businesses. Entrepreneurs may have access to large amounts of capital not taken into account when only examining private equity or commercial loans.

There are three other sources of capital used by early stage firms: corporate venturing, strategic alliances, and trade credit. Corporate venturing occurs when a parent company spins off an independent firm, with the parent capitalizing the venture. Employees of the parent firm are paid or encouraged to create the

new venture. Strategic alliances are joint ventures or partnerships where one firm, having capital, partners with a startup to exploit a market, manufacture a product, or deliver a service. Strategic partners invest about \$250,000 on average. Trade credit is widely used, but not well documented for potential high-growth firms. Suppliers, distributors, or consumers offer credit, usually as extended payment terms, to new ventures to get them started.

Once firms emerge at early stage, later-stage venture capital, as well as bank and finance company loans, tend to become more readily available. Also at this stage, many firms receive financial help from trade credits. However, insider finance remains the major source of capital for the vast majority of firms. Entrepreneurs invest their own capital or retained earnings when they are able to make sales. In some cases, entrepreneurs may sell a prototype and use the income—retained earnings—to reinvest in a better model. Friends and relatives provide capital, as do angels.

An important point about financing early-stage high-growth firms is that some investors bring not only capital, but also expertise to ventures—they are value-added investors. Angels and venture capitalists typically have extensive experience in markets, industry, and management, allowing them to actively participate not only in capital formation, but also in firm development. Angels mostly invest in entrepreneurs with whom they are comfortable and concern themselves less with markets, whereas venture capital firms often control entrepreneurs and concern themselves more with markets. Having an equity position in the firm empowers them to participate in its management. In the case of angels, most invest in firms near their homes, facilitating their active participation in new ventures. Empirical research confirms importance of this guidance and expertise to entrepreneurs. Corporate spin-offs and strategic alliances also are value-added investments. Banks and finance companies, mutual funds, bond traders, stock brokers, and high net worth investors are only minimally value-added investors, relying on the track record of the firm established over time, disclosure required by regulatory agencies, and/or value of pledged collateral.

ARE THERE CAPITAL GAPS?

The preponderance of empirical research, analysis, and opinion in the 1990s suggests that overall, insider, informal, and formal capital markets function well. Surveys of entrepreneurs in businesses generally, and high-growth firms particularly, reveal that capital formation was not a problem. A review—based on surveys of entrepreneurs, business associations, and bankers, along with analyses of lending data—by the

Board of Governors of the Federal Reserve System of capital availability nationally concluded that capital formation was not a significant problem. Business magazines across the United States generally report that the economy is flooded with capital. Venture capital funds over the past few years have invested record amounts of capital in early stage, including seed and R&D, firms. Because some observers believe significant capital gaps exist, even in the face of the evidence above, it is useful to take a closer look at empirical research undergirding this conclusion.

There may be some confusion about problems entrepreneurs have in raising capital versus the lack of capital available for investment in worthy business opportunities. Entrepreneurs typically spend a great deal of time searching for and then obtaining capital. This poses a problem in that entrepreneurs do not obtain capital right off the bat. However, the grueling process of obtaining capital is necessary to help entrepreneurs test the viability of their ideas in the market. As they get turned down by investors, they learn more and more about their idea: that it likely will be commercially viable or not. This process protects entrepreneurs from wasting their own resources—time and money—and those of investors. A major function of capital markets, then, is not only to provide capital, but also to ferret out investments, worthy and unworthy. The tough process of capital formation is not a capital gap because it does not concern a shortage of capital for worthy ideas.

One reason capital formation may not be a problem at early stages is that, for most ventures, only small amounts of capital are required to develop an idea, start up, and begin production. One-half of *Inc.* magazine's top 500 fastest-growing new small businesses in America were seeded with less than \$50,000, one-fourth with less than \$5000. A Coopers & Lybrand study of fast-growing new firms found that median startup capital was \$82,300. Because the vast majority of high-growth firms do not need massive infusions of capital, issues of formal venture capital availability are important to a relatively small number of firms, a finding supported below.

Evidence concerning impact of earliest-stage venture capital on firm growth, rather than startup, appears mixed. One study found that entrepreneurs who were denied venture capital believed that their businesses grew more slowly as a result. Other studies show that firms with venture capital backing grew faster than those without it. However, surveys of entrepreneurs with venture capital backing report that although they had venture capital, it was not critical in helping their firms to grow. These survey results can be interpreted in several ways. It may be that venture capital fund managers are good at picking the most promising high-growth potential firms, with the result

that they grow faster than non-venture-backed firms, which may be less desirable choices. This is consistent with the fact that venture capital firms spend a great deal of resource evaluating prospective deals before they invest. Another explanation may be that venture capital is not so important to growth as the value-added investment and management control that—through equity—venture capitalists bring to emerging high-growth firms. Entrepreneurs in venture capital-backed firms clearly do not distinguish between the finance and knowledge that venture capitalists bring to new enterprises. Human capital may be more important than finance.

Empirical evidence cited above on startup capital and earliest-stage venture capital only concerns successful, proven firms. What can be said about firms that were intended to be high-growth ventures, but never started up because of insufficient capital? Virtually nothing can be said, at least directly. It is easy to identify entrepreneurs with ideas who failed early on for want of capital, but impossible to determine whether their ventures would have attained high growth or whether they would have been viable. If firms initially denied capital eventually start up and succeed, then one might argue that they are not victims of capital gaps by definition. Analysis here is concerned with entrepreneurs who never start.

One way of getting at the issue of unmet capital needs versus non-capital-related failure, at least indirectly, is to look at how existing high-growth firms are capitalized at earliest stages. Numerous studies show that the vast majority of entrepreneurs obtained insider or informal capital, and, to some extent, angel investment, but typically not venture capital funds or other forms of private equity. Because of their age and size, they also did not access public equity or debt. Therefore, if existing high-growth firms did not typically rely on sources other than insider and perhaps some angel capital, then it is unlikely that there is a large, unmet need for formal capital investment at earliest stages. Entrepreneurial failure to start up in most cases must relate to non-capital issues, such as poor business ideas or poor management skills or capacity.

It is also important to reemphasize that venture capital represents a small proportion of earliest-stage investment, compared with insider and angel capital. Yet, venture capital seems to receive much more attention than insider and angel capital investment. Venture capital investment has become highly visible in the press, whereas insider and angel finance remain a private activity. Shortage of venture capital and capital gaps are not equivalent issues.

Another line of argument in the literature is whether the very existence of federal and state venture capital programs suggests prevalence of capital gaps for potential or emerging high-growth firms. It is argued that entrepreneurs accessing these federal and state

programs have been turned down by private investors and have insufficient insider or informal capital to start up. This seems to show that successful businesses funded by publicly subsidized programs represent investment opportunities missed by private capital markets. The Small Business Administration's Small Business Investment Company (SBIC), for example, funded household names in the computer industry in their formative stages—Intel, Cray, and Apple. In short, existence of public programs reinforces the perception of capital gaps at earliest stages.

However, many economists offer other possible interpretations. Public programs do not generally capitalize businesses at seed or startup stages, and only a few address earliest stages.^a Audits by the Inspector General's Office, for example, showed that SBICs judiciously avoid earliest-stage financing even when subsidized to take on increased risk.^[2] In this respect,

^aConsider three very different federal programs as examples. SBICs are widely replicated in OECD countries and represent the largest venture capital program in the United States (see Ref.^[2]). Ostensibly, the program's intent is to encourage investments in pioneer firms, not just traditional ventures, in filling capital gaps left by private capital markets. Yet, in a report on SBIC best practices by SBA's Office of the Inspector General, the OIG found that the most successful—that is, profitable, and not at risk of failure—SBICs invested in later stage companies, especially leveraged buyouts, and avoided seed or start-ups altogether. Further, successful SBICs only invested in firms with an established market niche. Investments made in early-stage companies with speculative markets had relatively low success rates. Therefore, even though SBICs, as financial intermediaries, are backed by government guarantees to encourage debt and equity investment in riskier ventures, SBIC managers still tend to make investment decisions as do private venture capitalists. When they act differently, they risk going out of business, even with government backing.

In the high-tech field consider the federal Advanced Technology Program (ATP) (see Ref.^[3]). Advanced Technology Program provides funding, matched by private investment to companies or firms in consortia of universities, businesses, and government laboratories for the development of generic technologies with broad application. Advanced Technology Program funds high-risk development *past the basic research stage*, but not yet ready for commercialization. In addition, ATP also specifies as a qualification: the past performance of the company or joint venture members in carrying out similar kinds of efforts successfully, including technology applications (consideration of this factor in the case of a startup company or new joint venture takes into account the past performance of key people in carrying out similar kinds of efforts). Only 37% of awardees of ATP grants were small businesses (see Ref.^[4]). A study by GAO found that public funding crowded out private investment in R&D, contrary to the legislative intent of the program (see Ref.^[5]). See also Refs.^[6] and ^[7] for positive and negative assessments of federal technology-based programs for high-growth firms.

The SBA 7(a) loan guarantee program. 7(a) is a loan guarantee program for institutional lenders extending credit to new (no older than 6 months) and existing small businesses, which were turned down for credit using normal business channels (see Ref.^[8]). Earliest stage ventures are clearly not represented in the 7(a) program. Only about one-fourth of 7(a) loans even go to new businesses. There is no requirement that borrowers have high-growth potential ventures or innovative products, technologies, or processes.

publicly subsidized programs are like formal capital markets. Most important, economists argue that many publicly subsidized firms could have accessed private capital (even though they were turned down by private investors or lenders), but were attracted to lower capital costs afforded by public programs.^b In its study of Advanced Technology Program, for example, the General Accounting Office (GAO) found that this federal program was for many firms the first choice of funding, rather than an alternative source once having been turned down by private capital.^[8] To the extent that this is so, publicly subsidized capital programs displace or crowd out private funds that would have been invested were it not for subsidies.^[10]

Although federal programs do not appear to reach earliest stage high-growth ventures, they do, however, reach ventures pursuing earliest stage technologies, processes, or products. This distinction is often overlooked or ignored in evaluating federal program outcomes.

Researchers also find that many state-funded venture capital programs across the United States have been unsuccessful, and many are being defunded.^[11] State-funded programs rarely submit to independent evaluation, but studies generally find job creation claims to be overblown. Many claim greatest success in job retention, raising the question of whether programs are subsidizing jobs that would have been maintained without public funding or keeping alive firms that ought to die or shrink. State venture capital programs, presumably tailored to needs of local business, likely would be more in demand were capital gaps really so prevalent for high-growth firms.^c Some analysts conclude that state venture capital programs exist mostly in response to other states: existence of a program in one state is perceived as a competitive disadvantage by adjacent states, so they too adopt one.^[13] This draws into question whether capital market failure exists. So these programs are created not so much to stimulate new development as they are to channel firms that would have started up anyway into selected locations.

The argument that federal and state programs may not be addressing real capital gaps is supported by empirical research suggesting that venture capital firms, angels, and banks have excess capital to lend, but are not offered viable investment opportunities. Venture capitalists, for example, only fund 3% to 5% of the 200 to 500 (in some cases more) finance proposals most receive annually; this after they spend a great deal of time evaluating the entrepreneur's capacity to grow a company and market opportunity. Economists refer to this as preinvestment due diligence. Venture

capital firms often retain consultants and experts to evaluate products. It is not unusual for a venture capital firm to pay consultants to take a product apart to see how it was made and whether it can be mass produced for profit. Of the 3% to 5% pursued, most will not pan out. Venture capital fund returns, including the more than \$1 billion annually invested in earliest stage firms, are three times higher than publicly traded stock on average, but venture capital returns have three times the standard deviation above or below the average than the average for publicly traded stock. The implication: some venture capitalists receive high returns on investments, whereas many others fail miserably. Venture capitalists also refer viable projects to other venture capital firms when they are unable to fund them, and in numerous cases, venture capitalists counsel entrepreneurs in how to make their business plans marketable. Many venture capital firms, not interested in an idea themselves, will help an entrepreneur to prepare a business plan for another venture capital firm's consideration, sometimes in exchange for equity. Venture capital firm practices suggest that entrepreneurial proposals receive wide consideration in the market.

The notion that good ideas are not plentiful is borne out in studies about why venture capitalists and angels, for that matter, reject investment proposals from would-be high-growth entrepreneurs. The two major reasons for rejection are 1) there is no demonstrated market for the product, and 2) the entrepreneur has insufficient management capacity to grow the firm. A common problem in potential and emerging high-growth, technology-based firms is that they were started by scientists or technicians who lack management, marketing, and entrepreneurial skills.

A final line of argument concerns study methodology. Most researchers examine whether capital formation was a problem for entrepreneurs starting high-growth firms. Study results mostly show that for entrepreneurs capital formation is difficult, and rightly so, as observed above. However, studies fail to ask whether the entrepreneur merely desired more capital or really needed it, and whether the entrepreneur eventually raised money in another way or found ways around the difficulty. When these additional factors are explored, capital gaps among entrepreneurs seem to be more perceptual than real.

In sum, empirical evidence suggests that capital gaps for potential and emerging high-growth firms are not a significant problem in today's economy, at least not significant enough to warrant federal intervention.

ARE THERE REGIONAL DISPARITIES?

Empirical studies suggesting regional capital gaps abound. Regions, states, and rural economies differ

^bResearch on this point is legion (see Refs.^[6,9]).

^cSome analysts take a more positive view of state-funded venture capital programs. See for example, Ref.^[12].

markedly in amount of capital available locally or attracted from outside for investment in earliest-stage potential high-growth firms. Regions differ with respect to rates of startup, employment generation, and industrial sectoral growth. Research focuses either on venture capital investment or prevalence of high net worth individuals, likely because both are easier to track. Angel investment, because it is not publicly monitored or regulated, is virtually impossible to track. Studies look at the amount of venture capital attracted to high-growth regions per firm or per capita, and compare this with the region in question. The difference between what a region has now and could have if it performed like a high-growth counterpart is regarded as the capital gap. Studies of high net worth individuals and angels examine census data showing number of individuals with high incomes in the region and investments made. These are then compared with other regions to discern capital gaps.

Many regional studies finding capital gaps conclude that regions either have local investors—venture capitalists, angels, or high net worth individuals—who are not interested in local investment, or they have insufficient numbers of local investors. In the former, capital flows out of the region and is not replaced by sufficient inflows, in the latter there is a shortage of capital. The same studies tend to rely on studies reviewed in the previous section to support the notion that worthy opportunities for investment are plentiful, but are not capitalized.

Regional studies are problematic because they rely on research and inferences about the prevalence of unfunded worthy investment opportunities seemingly suffering from the same problems already presented in the preceding sections. If the analysis above holds, then it is not likely that there are large numbers of worthy unfunded investments in most regions. Analysts, then, may be misperceiving or overestimating potential high-growth firm opportunities available for investment.

Critics question whether methodologies employed in such regional research actually reveal capital gaps on the investment side. Criticism differs for venture capital, angels, and high net worth investors.

Venture Capital

Concerning venture capitalists, venture capital is attracted to a region from investors across the United States and internationally because ventures in the region present the best investment alternatives—highest return adjusted for risk. Risk is assessed on information that venture capitalists hold about entrepreneurs and that markets entrepreneurs propose to exploit. Venture capitalists have expertise in

industries in which they invest and in the markets these industries serve, and are adroit at evaluating entrepreneur capacity to grow a company. Successful initial venture capital investments attract subsequent investment to a region, a kind of herd effect. What undercapitalized regions lack, then, is not only the best investment opportunities compared with others, but also agglomeration of ventures working in similar industrial sectors, professionals with expertise in these industries—bankers, lawyers, marketing professionals, technicians, scientists, and engineers—educational facilities, especially universities producing research and graduates for industry, and so on. It is unlikely that venture capitalists will extensively invest in regions without this supportive infrastructure.

Consider Route 128 in Boston. This corridor is or was home to hundreds of high-tech companies with household names, Digital Equipment, Lotus Development, Wang, and Data General. It is served by dozens of venture capital funds and some of the largest emerging growth company funds, such as that offered by Fidelity Investments. It boasts of MIT and Harvard, not to mention Brandies, Tufts, Boston University, and Boston College, whose professors, scientists, and students spawn hordes of new high-tech, high-growth companies annually. In addition, there are countless consulting firms and businesses supporting these ventures. Such clustering in turn attracts entrepreneurs and high-tech workers. Only a handful of regions in the United States can boast of this concentration of resources in high-tech. This leads some critics to argue that in addition to lacking viable business ideas, capital does not flow to some regions to potential high-growth firms because they lack the supportive infrastructure, not just money.

Angel Investors

However, not all investment is made by venture capital funds, angels play a substantial role. Concerning angel investors, critics argue that angels, like venture capitalists, are also knowledgeable investors, having experience in industries in which they invest and living near their workplaces so that they can take an active role in operating a company. Angels know about markets, but their main focus for investment is their potential relationship with entrepreneurs. Also, the literature suggests that many angels invest not only to profit, but also as a kind of recreation. In short, they receive both monetary and psychological rewards from investment. If local opportunities for investment do not include both dimensions, then angels may not be interested. Capital gaps can only exist if potential angels elect not to invest in local opportunities that are at least as profitable as others. Because angels usually

invest close to home and take part in businesses, it is unlikely that other regions would attract them as investors who will take an active role in companies. Some studies tend to look at people with the same income or worth as angels as potential investors. But wealth alone is not an indicator of the capacity of individuals to make investments in risky ventures. Again, it is not capital that is at issue, it is expertise and participation that separate angels from comparably wealthy people.

High Net Worth Investors

Critics argue that high net worth investors, like venture capitalists, invest in opportunities representing the best investment adjusted for risk. Capital, then, flows across the United States. Unlike venture capitalists and angels, high net worth investors do not tend to take an active role in assessing investments or operating companies. More important, they invest millions at a time and are generally not interested in directly investing small amounts of capital in emerging businesses. They tend to leave oversight and due diligence to intermediaries—either venture capital fund managers or brokerage firms. So research looking at prevalence of high wealth individuals, critics argue, is misdirected because their money is intermediated by investment firms. Investment firms, then, are the issue, not the wealthy. So, it does not matter whether high net worth investors are themselves present or absent from a region.

Other evaluations suggest that some regional studies may be measuring perceptions of capital gaps that cannot be supported. In a survey of economic development practitioners in Alabama, Maine, Minnesota, and Washington, for example, the GAO found that 66% and 77%, respectively, believed that credit and equity gaps existed to a great or very great extent, for startup businesses, but somewhat less for expanding businesses at 26% and 33%, respectively.^[14] However, when GAO asked practitioners to provide empirical evidence to support their conclusions, none was forthcoming.

In evaluating the regional capital gap literature, it is important to consider evaluations of public programs. One critical study examined firm performance data for participants in the federal Small Business Innovation Research (SBIR) program,^d comparing them with a control group of firms not participating in the

program.^[18] Overall, SBIR participants outperformed non-participants in sales and employment growth. At first blush, it appears that capital gaps were closed by publicly subsidized programs. However, SBIR participants, located in areas with low concentrations of private venture capital and sparse industrial clustering, had less in sales and employment growth compared with SBIR participants in regions with concentrations of venture capital and industry clustering. It might also be the case that SBIR investors do a better job of picking winners in supportive economies than in less supportive economies. Yet, this exception might prove the rule: perhaps SBIR investors in less supportive economies find it difficult to adequately evaluate merging high-growth firms because of the lack of supportive infrastructure.^[17] This also tends to confirm the impression that factors other than capital availability influence business location.

One thing on which nearly all analysts agree is that rural areas have sufficient access to capital generally, but are underserved by equity investors, angels, or venture capital. There are insufficient numbers of entrepreneurs to attract outside investors and relatively few local investors. Consequently, investment opportunities in rural areas are somewhat less likely to come to fruition locally. The perception is that entrepreneurs who take their ideas to cities, where capital is more readily available, face high transaction costs—costs of finding investors or lenders—and high rejection rates because they are not known to investors or lenders.

The rural capital gap for equity investment may be overstated. A study of all Initial Public Offerings (IPOs) from 1970 to 1995, broken down by urban and rural areas showed that 10% of IPOs—measured in amounts invested and numbers of firms—were for rural firms.^[19] In addition, more than one-half of rural IPOs were backed by venture capital, while less than one-half of urban firms had this backing. Rural IPOs performed as well as urban ones, as measured by stock price and return on investment. Venture capital and IPO investment does flow to many rural areas. How can this be reconciled with the fact that rural areas lack investors locally, industrial concentrations, and supportive infrastructure? The answer is that entrepreneurs actually do find investors and support in regional financial centers—Minneapolis, Denver, Houston, and the like—contrary to perceptions. Capital markets function well for emerging high-growth firms as long as they are near urban centers, which most appear to be.^e This research also tends to call into

^dFor an overview of this program, see Refs.^[15,16] The SBIR Program's intent is to stimulate technological innovation, use small businesses to meet federal R&D needs, foster participation by minorities and disadvantaged people, and increase commercialization of innovations. Research on the program finds that it, like ATP discussed above, also crowds out private funding using public subsidies. In addition, the fact that firms tend to receive repeat awards draws into question whether SBIR is reaching startup entrepreneurs. See especially, Ref.^[17].

^eThere are about 2000 rural counties in the United States. The lion's share of these are federal conservation lands, and sparsely populated and remote farming, ranching, and mining regions. Much of the rural population lives in counties adjacent to metropolitan areas. See Ref.^[20].

question prevalence of barriers in linking investors and entrepreneurs in rural areas rather than capital gaps.

ARE THERE TECHNOLOGY-BASED SECTOR DISPARITIES?

Empirical studies on capital gaps for emerging high-growth firms have not definitively resolved the issue of capital availability for technology-based companies, but there is evidence that is suggestive, either from limited empirical investigations, or more often from anecdotal evidence and case studies. In exhaustive, independent reviews of research and venture capital programs in western, industrialized countries by the Bank of England and Organization for Economic Cooperation and Development (OECD)^[21,22]—along with extensive Congressional testimony,^f reports from the 1995 *White House Conference on Small Business*, and others^[24,25]—most observers concluded that technology-based, high-technology, and biotechnology-based (hereafter, referred to as technology-based firms) potential and emerging high-growth firms, although they pass through the same growth stages as other ventures, have more specialized capital formation needs and are more likely to experience capital gaps. Like other firms, technology-based firms depend heavily on insider capital in earliest stages. Exceptions, of course, are spin-offs from existing corporations. But capital needs of technology-based firms tend to be much higher than other businesses starting up. Technology-based companies often require access to expensive equipment and materials during seed and R&D earliest stages. Entrepreneurs in these companies may take years to develop and test a product to see if it even works. Once it works, then more time is required to determine whether it can be manufactured at a reasonable profit. Large amounts of capital over the long-term, required at seed or R&D stages by some ventures, are sometimes not available from insider and informal sources. Angels and venture capitalists—having large amounts of capital, invested at high risk, over an extended period—may be the only source of finance for some ventures. Angels and venture capitalists often invest in rounds—first, second, and third—as entrepreneurs increasingly demonstrate the viability of their ideas. Because this process can take years before investors profit, angel and venture capital is often called “patient capital.” Once they mature, technology-based companies begin to resemble other businesses in their capital needs.

A peculiarity of technology-based firms is that they tend to be started by technicians—engineers,

programmers, or scientists—who have a single product to market, have little or no management experience, and may lack capacity to grow a company. Having a single product or application makes it difficult for investors because all the ventures eggs are in one basket. Most emerging high-growth firms are single-product dominated or use a single application. Dell Computers, for example, the best performing high-tech stock of the 1990s found a competitive advantage in the computer industry: selling directly to customers while avoiding a middleman. Dell Computers did not innovate new technologies across several product lines. This greatly increases investment risk, turning away potential investors. Technicians, having little management experience, further increase risks, if they try to grow the business themselves without seeking professional help. Because technicians may be reluctant to give up firm management, they greatly reduce the pool of investors likely to be interested in their venture.

Even though angels and venture capitalists may be the investors of last resort for technology-based firms, entrepreneurs may not have access to these investors. Angels and venture capitalists tend to specialize in certain technologies where they have expertise. Some good investments may not be of interest to all potential investors. Angels and venture capitalists, with expertise in certain technologies, may not have time to evaluate or participate in the opportunity. Entrepreneurs and investors may be unable to find one another to close a deal. Or, angel and venture capitalists, having to work closely with entrepreneurs they fund, may not like their prospective business partners so they reject a deal.

Although capital formation for technology-based ventures may be subject to the influences above, research that moves beyond this to identify what kinds of firms experience capital gaps, why difficulties occur, and to what extent there are problems is lacking, and ought to be a priority for future study. As it stands now, the doubtful existence of capital gaps generally still seems to apply in the high-tech area.

CONCLUSION

Because most new jobs are created by a small number of emerging high-growth small companies, some observers are calling for government intervention in capital markets to reduce capital gaps, causing emerging high-growth firms to grow too slowly and potential high-growth firms from failing early on.

Empirical evidence suggests that capital markets, including formal and informal investors, work well, and that overall capital gaps are not problematic for potential and emerging high-growth firms. Capital gaps might emerge in the future during severe economic downturns, as they have in the past.

^fSee, for example, Ref.^[23].

Evidence about capital gaps for regions and technology-based firms, is not conclusive. Some regions have greater concentrations of capital for potential and emerging high-growth firms than others, but this may have more to do with the prevalence and quality of investment opportunities and the requisite industrial clustering and financial infrastructure that stimulates and sustains investment, than with capital unavailability. Some technology-based firms appear to experience capital gaps, because of the peculiarities of how these firms typically grow and are financed, but there is little evidence to support or refute the notion that there is a widespread capital gap problem for large numbers of high-tech firms.

It appears that the case for federal intervention to correct capital market failure for emerging and potential high-growth firms has not been made.

REFERENCES

- Birch, D.L. *Hot Industries*; Cognetics, Inc.: Boston, 1998.
- U.S. Delegation to the OECD. *The U.S. Experience. Best Practice Policies for Small and Medium-sized Enterprises*; Organization for Economic Cooperation and Development: Paris, 1997; 45–58.
- U.S. Library of Congress, Congressional Research Service. *Advanced Technology Program*; October 8, 1998; CRS Report 95-36 SPR.
- Technology Administration. *ATP*; National Institute of Standards, April, 1996; 20NIST-ATP-96-2.
- GAO. *Measuring Performance*; General Accounting Office: Washington, DC, April 10, 1997; T-RECD-97-130.
- Branscomb, L., Keller, J.H., Eds.; *Investing in Innovation: Creating a Research and Innovation Policy That Works*; MIT Press: Cambridge, MA, 1998.
- Cohen, L.; Noll, R. *The Technology Pork Barrel*; Brookings Institute: Washington, DC, 1991.
- GAO. *A Comparison of SBA's 7(a) Loans and Borrowers with Other Loans and Borrowers*; GAO, September 1996; RCED-96-222.
- Economic Research Service. *Credit in Rural America*; Agricultural Economic Report; ERS, U.S. Department of Agriculture: Washington, DC, April, 1997; Vol. 749.
- Irwin, D.; Klenow, P.J. *High Tech R&D Subsidies*; National Bureau of Economic Research: Cambridge, 1994#4974. Haynes, G.W. A Comparison of Borrowers with SBA and Other Loan Guarantees, U.S. Small Business Administration, Contract Report, June 1993, Study raises questions on SBIR, *Physiologist*, **1995**, 38, 148. Corporate welfare or funding godsend, *Scientist*, **1997**, 11. See Ref.^[6].
- Eisinger, P. State venture capitalism, state politics, and the world of high-risk investment. *Econ. Dev. Q.* **May 1993**, 7, 131–139. Florida, R.; Smith, D.F. Venture capital, innovation and economic development, *Econ. Dev. Q.*, **November 1990**, 4 (4), 345–360. For an inventory of federal/state programs, see, Coburn, C.; Bergland, D. *A Compendium of State and Federal Cooperative Technology*. Battelle: Columbus, 1995 and for a review see U.S. Library of Congress, Congressional Research Service. *Technology Development: Federal-State Issues*. CRS Report 96-958 SPR, November 22, 1996, by W. Schacht.
- Bygrave, W.D.; Timmons, J.A. *Venture Capital at the Crossroads*; Harvard Business School Press: Boston, MA, 1992.
- Jenkins, J.C.; Leicht, K. Direct action by the subnational state. *Soc. Probl.* **August 1996**, 43 (3), 306–326. Fosler, R.S. State Strategies for Business Assistance. In *Growth Policy in the Age of High Technology*; Schmandt, J., Wilson, R., Eds.; Unwin: Boston, 1990.
- GAO. In *Rural Development: Availability of Capital for Agriculture, Business and Infrastructure*; GAO: Washington, DC, May 1997; RCED-97-109.
- Schach, W.H., Ed. U.S. library of congress, congressional research service. In *SBIR Program*; September 11, 1998; CRS Report 96-402 SPR.
- GAO. In *Federal Research*; GAO: Washington, DC, 1998; RCED-98-132.
- Wallsten, S.J. *Rethinking the Small Business Innovation Program*; Branscomb, L., Keller, J.H., Eds.; MIT Press: Cambridge, MA, 1998.
- Lerner, J. *The Government as Venture Capitalist*; National Bureau of Economic Research: Cambridge, 1996; #5753.
- Brophy, D.J. Developing rural equity capital markets. In *Financing Rural America*; Federal Reserve Bank of Kansas City: Kansas City, April 1997; 159–172.
- Economic research service. In *Understanding Rural America*; Agriculture Information Bulletin, U.S. Department of Agriculture, February 1995; Vol. 710.
- Bank of England, the financing of technology-based small firms. London: Bank of England, 1997.
- Organization for economic cooperation and development (OECD). *Venture capital and innovation*. Paris: OECD, 1997.
- Committee on banking, finance and urban affairs, U.S. house of representatives hearing before the subcommittee on economic growth and credit formation. In *Availability of Financing for High-Technology Companies*; November 16, 1993; 103rd Congress, First Session, Serial no. 103-99.
- Carnegie Commission on Science, Technology and Government. *Science, Technology and the States in America's Third Century*; New York, September, 1992. Westhead, P.; Storey, D.J. Financial constraints on the growth of high technology firms in the United Kingdom. *Appl. Financ. Econ.* **1997**, 7, 197–201. Moore, B. Financial constraints to the Growth and development of small high technology firms. *Financing Small Firms*; Hughes, A., Storey, D.J., Eds.; Routledge: London, 1994. Oakey, R. *High Technology New Firms: Variable Barriers to Growth*; Paul Chapman: London, 1995.
- Majewski, S.E. Using strategic alliance formation. Berkeley, CA: University of California, 1997, working paper.

Federalism in “Homeland Security” and Crisis Management

Alethia H. Cook

David J. Louscher

Department of Political Science, University of Akron, Akron, Ohio, U.S.A.

INTRODUCTION

Responding to a domestic act of terrorism in the United States will be a major challenge for coordinated federal, state, and local response. The word “will” in the preceding sentence was used intentionally instead of “would.” There is general agreement among federal, state, and local governments, as well as the defense community, that the terrorist threat is a question of when, not if, a strike will occur. The Oklahoma City bombing and the World Trade Center bombing brought this harsh reality to the forefront for many in government and academia who were studying terrorism. The rest of the country was made painfully aware of the threat on 11 September 2001 when terrorists hijacked four planes and flew them into the World Trade Center towers, the Pentagon, and a field in Pennsylvania. In spite of these many warnings, the nation is woefully unprepared to meet this critical challenge, particularly if the terrorist act is something other than a conventional explosive.

The concept of federalism, so important in the American political system, itself will be one of the critical challenges. Local authorities who are witnessing the devastation wrought by an act of terrorism in their own backyards will understandably be reluctant to turn over responsibilities for dealing with the crisis to state and federal level entities. Federal organizations are interested in protecting evidence and the crime scene in the interest of catching the perpetrator, as well as assisting the local government with recovering from the event. These potentially conflicting goals may create tensions between the different government agencies at different levels. Response to an intentional biological contamination may be even more complex than for a chemical or conventional strike.

While the nation has been making efforts to improve preparedness, two distinct problems remain that threaten to undermine effective response. The first is that the majority of the effort to improve readiness has been made at the federal level. This is problematic as the response to terrorism, whether it is chemical, biological, nuclear, radiological, or conventional, is necessarily a local event. The police, firemen, emergency management services (EMS), hospital personnel,

and other first responders that will be tasked with dealing with the impact of such an event are the ones that need the resources to improve preparation. The second major problem with preparedness is that most of the preparation efforts have focused on either a chemical, nuclear, or conventional attack. Although these threats are important, they are not the only ones faced. Because the nature of the events are so different, increasing preparedness for chemical response will not help deal with a biological attack. The impact of federalism on these two challenges to preparedness, public administration, and effective response are the focus of this article.

FEDERAL PREPARATION FOR A LOCAL CRISIS

There is no question that the United States has been working to improve its response capabilities. In particular in the aftermath of September 11, extensive efforts to better train, equip, organize, and manage response have been a focus of government. What is in question, however, is whether the resources that are being expended are being used in an efficient and effective manner. As demonstrated on Fig. 1, there is a lot of funding being spent at the federal level, with little money reaching the first responder. There is no centralized coordination for the expenditure of this money or for training and preparation.^[1] This is problematic as there are over 40 government agencies playing an active role in homeland defense.^[2] According to Warren Rudman, “[t]he United States stands in dire need of stronger organizational mechanisms for homeland security. It needs to clarify accountability, responsibility, and authority among the departments and agencies with a role to play in this increasingly critical area.”^[3]

One of the major efforts undertaken by the federal government has been the establishment of 27 National Guard Weapons of Mass Destruction Civil Support Teams. These teams are trained and equipped by the federal government, but will be deployed under the control of governors in the state where the event has occurred. They train regularly and have access

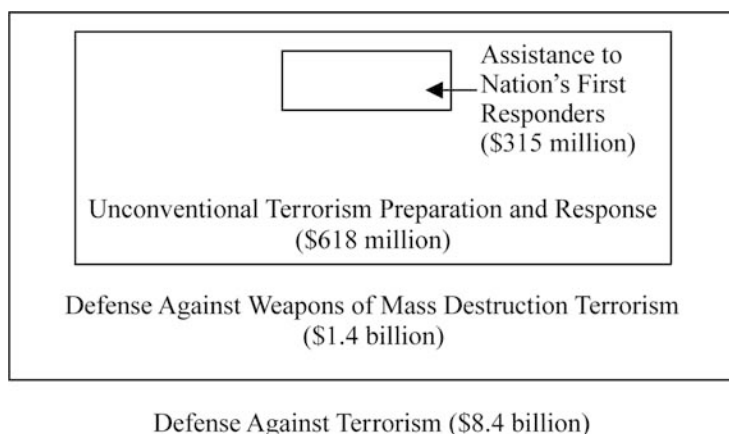


Fig. 1 Breakdown of the federal budget to combat terrorism, 2000. (Adapted from Ref.^[5].)

to state-of-the-art communications equipment.^[4] Although the National Guard teams are a valuable asset, an argument can be made that the funding could have been spent to equip first responders rather than a federal level team that will take time to deploy. According to one critic, “2333 hospitals or fire stations could be out fitted with decontamination capabilities for the cost of standing up one National Guard Civil Support team. If the total 1999 budget for these National Guard teams had been used thusly, 49,800 local facilities could have been armed for decontamination.”^[5]

This situation places the burden on local responders to determine where the federal resources are and how they can get them quickly in case of an emergency. Without this kind of arrangement in place, it does not matter how much the federal government spends on preparedness. If the local-level entities cannot rapidly get equipment, vaccines, and other resources, the terrorist strike will have a much more devastating impact on the local community.

Federal organizations are engaged in outreach activities aimed at informing local responders about the resources available to them and providing training. This task is more complex than it would seem. The event itself will dictate who the responsible local authorities will be. As a result, the federal organizations need to get their information out to all local fire, police, health, and infrastructure officials that may be relevant to a terrorist strike. Because of this complexity, major cities have received the majority of contact with and training from the federal government, leaving smaller communities seriously unprepared.^[6]

An additional problem with this federal training is that once the exercise is over, there is no follow-on training or regular practice of the techniques for effective response. Furthermore, there are no standardized ways of assessing the quality of response.^[5] The implications of these problems are significant. It is unreasonable to expect first responders to learn more

effective response if they are trained only once, are not given adequate feedback on their performance, and do not have the opportunity to learn from their mistakes.

Local authorities clearly need to be better trained to identify and respond to a terrorist event when one occurs. The implications of this federalism problem are significant. Money, training, and cooperation have to be spread among the levels of government if the United States hopes to mount a successful response. The problems of coordinating the levels of government will fall within the tasking of the new Department of Homeland Security. It will be responsible for making the country more prepared to respond to catastrophic terrorism. One of its major challenges will be to find a more effective pattern of funding, training, and organization to overcome the paradox of federal level preparations for local level events. Although there is a definite response role for the federal government, it will not be first on the scene of a terrorist strike, nor is it likely to be the most important element of any response. A careful reassessment of the nation's response spending priorities should be undertaken to correct these problems and put resources in the hands of those tasked with first response. However, before that can be done, government must gain a better understanding of the complexity of response.

FAILURE TO PREPARE FOR BIOLOGICAL EVENTS

As mentioned above, different types of threats confronting the United States require distinctly different types of response. In the event of a chemical or conventional attack, there is an identifiable moment that the attack occurs. If a bomb goes off, responders know when it went off, they mobilize their resources to mount an effective response, arrive on the scene, struggle to mitigate the impact of the attack, and then turn

the victims over to the medical community. The majority of the efforts underway to improve preparedness have focused on this kind of event.^[7]

However, a biological attack will present itself very differently. This was demonstrated in the Anthrax attacks in 2001, when those exposed to the agent became ill over an extended period of time. Biological agents can take anywhere from 2 to 14 days to incubate and become symptomatic. Victims who are exposed to the agent may circulate among the general population during this time without being aware that they have been infected. Depending on the type of agent and the way it is transmitted, those who have been exposed may be spreading the disease within the community. To detect this kind of event, the public health system will have to recognize that an epidemic is underway. The faster the agent used is identified, the lower will be the casualty rate from the event. Each day that the event goes unreported is another day that the victims are infecting others. The problem is exacerbated by the fact that the public health providers must be able to identify the difference between a naturally occurring event and one that is intentional.

Effective response to a biological event, therefore, requires that local-level health providers be aware of the symptoms of the major potential biological warfare agents so that they can identify these diseases quickly (Table 1). They must also understand the ways in which a naturally occurring event is likely to differ from a terrorist strike. Finally, health care providers must be ready and able to report suspicious diseases to a system that can help to identify infection patterns. Although cases of a disease may cluster in a small community if the agent was distributed locally, it may be that the agent is distributed in a way that will initially manifest as a few cases that are widely spread.

Some analysts believe that the government is over-reacting to the threat of biological terrorism. They

argue that biological agents are difficult to grow in quantities necessary to disseminate them widely and that effective delivery of the agent is incredibly difficult.^[8] The anthrax attacks seem to support this argument as the agent was deployed in an inefficient manner and caused relatively limited casualties. However, fanatical terrorists may allow themselves to be infected with an agent, become contagious, and board an international flight; infecting the other passengers on each leg of the flight and effectively spreading the disease around the world.^[7]

This is where the concept of federalism will complicate response. Although diseases will initially be reported the local level, with patients going to their doctors or to an emergency room, identifying the extent of the terrorist act may rely on a national or even international coordination of medical information. The Centers for Disease Control and Prevention (CDC) in the United States issues guidelines to American hospitals for those diseases they must report so that outbreaks can be identified. However, many biological agents have flu-like symptoms and it may not immediately be clear to physicians that they are dealing with something other than the common flu. Furthermore, a widespread epidemic may be difficult to identify as patients arrive at different hospitals but no real trend seems to be present until much later. Every day the spread of the disease goes unrecognized, the potential impact of the event increases exponentially. Communication about disease diagnoses is key.

This poses a significant problem for many local health systems that are consistently underfunded. A survey revealed that many local systems do not have access to e-mail and the Internet, and may even still have rotary phones.^[7] They are not networked in any meaningful way aside from their communications with the CDC. As such, the CDC may be the first to pick up on a local epidemic because hospitals may not be

Table 1 Critical biological agents

First-tier agents	Second-tier agents	Third-tier agents
Anthrax	Brucellosis	Hantaviruses
Botulism	Cholera	Multidrug-resistant tuberculosis
Ebola	<i>E. coli</i>	Nipah virus
Junin (Argentine hemorrhagic fever)	Epsilon toxin	Tickborne hemorrhagic fever
Lassa fever	Equine encephalomyelitis	Tickborne encephalitis viruses
Marburg	Glanders	Yellow fever
Plague	Ricin	
Smallpox	Salmonella	
Tularemia	Shingella	

Note: First-tier agents are easy to employ, spread easily, and have high mortality rates; second-tier agents are moderately easy to employ and have low mortality; and third-tier agents are available, easy to produce and disseminate, and the potential for high mortality.

Source: Adapted from Ref.^[8].

able to effectively communicate about their patients' symptoms. These problems exist in spite of significant efforts on the part of the Department of Health and Human Services and the CDC to help local entities build their response capacities.^[5]

Another major problem that will be confronted by public health care officials is the panic that will likely occur among the local population. Again, a biological event will be different from a chemical or conventional attack. If someone bombs a building or releases a chemical agent in a subway, citizens who were not in those areas can be relatively certain that they were not affected. If a disease begins to spread, people will begin to worry that they have been exposed. Mass numbers of generally healthy individuals, or people with flu-like symptoms, may converge on hospitals and doctors' offices seeking medical attention. The Anthrax attacks provided evidence of this problem, with mailrooms closed around the country and sales of Cipro (the antibiotic for treating Anthrax) being sold on the internet to people who had been nowhere near the known cases of exposure. In any biological attack, local-level officials will need help in containing the panic and dealing with the influx of patients. Communication will also be important in this effort. Notifying the public about the threat, the way in which the disease is transmitted, the symptoms of the illness, and where to seek treatment will be critical to controlling the public's reaction. Calming public fears will be a serious challenge facing local authorities. This is an issue rarely covered by training exercises that are focused on chemical or conventional terrorism.

These problems generated by the federal nature of the American system must be overcome in order to have more effective response capabilities. Policy makers need to rethink the focus of national efforts to increase preparedness so that resources can be expended in ways that assist first responders in

performing their roles. More attention needs to be paid to the unique nature of biological terrorism or a major epidemic so that those kinds of problems can be detected and contained as quickly as possible. Without these changes in U.S. policy, a major terrorist strike will have an unacceptably large impact on the society and may undermine public faith in government.

REFERENCES

1. Neus, E. Bioterrorism: how real is the threat? *USA Today* **1999**, October 27, 1. See also NGB report: No one WMD ready. *Natl. Guard* **1999**, 53 (10), 13–14.
2. Holzer, R. Homeland security bill aims to coordinate agency roles. *Def. News* **2001**, 16 (12), 12.
3. Rudman, R.B. *Statement Before the House Committee on Government Reform*; March 27, 2001.
4. Garamone, J. *Guard Teams to Combat Weapons of Mass Destruction*; www.defenselink.mil/specials/destruction/Garamone.ltr.html (accessed March 29, 2001).
5. Smithson, A.E.; Levy, L. *Ataxia: The Chemical and Biological Terrorism Threat and the U.S. Response*; The Henry L. Stimson Center: Washington, DC, 1999.
6. Stanton, J. U.S. homeland defense policy mired in competing interests. *Natl. Def.* **2001**, 10–13.
7. Nesmith, J. Nation ill-prepared for warfare against germs. *Palm Beach Post* **1998**, August 31, 1A+.
8. Sokolski, H. Rethinking bio-chemical dangers. *Orbis* **2000**, 44 (2), 207–219 and Tucker, J. Chemical and biological terrorism: how real a threat. *Curr. Hist.* **2000**, 147–153.

BIBLIOGRAPHY

1. Kahn, A.S.; Morse, S.; Lillibridge, S. Public-health preparedness for biological terrorism in the USA. *Lancet* **2000**, 356, 1179–1182.

Feminism and Chaos Theory

Mary Ann McClure

*Department of Art, Music and Philosophy, John Jay College of Criminal Justice,
New York, New York, U.S.A.*

INTRODUCTION

Contemporary chaos theory explains phenomena traditional science cannot and reveals hidden order in the apparent random behavior of, among other things, business cycles, epidemics, and insect populations. It also provides a vision of nature that challenges the notion of nature central to traditional science. In its concept of nature and methodology, chaos theory parallels the efforts of contemporary feminists to reshape traditional understanding of scientific inquiry. Both feminist philosophers of science and chaos theory presume a methodology that attends to difference, is non-reductionistic, attributes agency to nature, and conceives of inquiry as fallible.

Traditional science makes a claim to being objective. The scientist who pursues objectivity seeks a context-free knowledge that aspires to be omniscient, fully formalizable, and unbiased and disinterested. What connects these characteristics is the key concept of detachment. Objectivity as detachment means the separation of the knowing subject from the object of knowledge. Objectivity depends upon clear and distinct boundaries between the self and the world. Thus, in the act of trying to gain objective knowledge, the scientific mind separates itself from nature.

THE FEMINIST CRITIQUE OF TRADITIONAL SCIENCE

Feminist epistemologists have examined the concept of scientific objectivity in terms of its claim to value-neutrality and in its association with masculinity, where the “masculine” denotes a cognitive style rather than a biological category. They have traced the historical origins of the concept of objectivity to the emergence of the Newtonian paradigm in the sixteenth and seventeenth centuries, when the new science of that day demanded new concepts of nature and inquiry. Newtonian science was struggling against the older medieval view that held nature to be an alive organism and part of God’s domain. Seeking a new concept of nature, modern science reconceptualized nature as dead and thus passive and inert—like a machine—in a fixed and determinate universe.

Feminist philosophers of science have argued that this conceptual framework of modern science is fundamentally gendered. Attributes such as objectivity, detachment, and abstract thinking that are attributed to science are also stereotypically masculine. The achievement of scientific knowledge, identified as a specifically male way of knowing, demands not only the separation from, but the domination and control of nature and of female human nature in particular. Thus, the feminist project becomes one of rethinking the concept of objectivity and, once again, reimagining our concept of nature.

SCIENCE AS FALLIBLE

If traditional objectivity claims omniscience, feminist philosophers of science, such as Sandra Harding, advocate a “stronger” sense of objectivity in which knowledge is understood to be fallible.^[1] Evelyn Fox Keller and Donna Haraway seek alternative framings of biological and physical theories in which scientific knowledge is modeled on a dialogue that remains open-ended and does not seek closure. For Keller and Haraway, rational knowledge is less a set of facts or theories than a process. This notion of scientific inquiry as a process places emphasis upon on the act of revision, in which we encounter the limits of our control of the natural world. To understand is not merely to know a set of facts about nature but to possess the skill of being open to the experience of limitations.

The move from scientific inquiry having the potential for omniscience to its being understood as inherently limited and fallible is best captured by Keller’s distinction between order and law in nature. A law-like nature fits the demand of traditional objectivity which calls for a total correspondence between theory and reality. Nature that is law-like can be captured best by the language of science—mathematics. In contrast, when Keller understood nature as being orderly, she conceived of the order in nature as being more complex than any law that the human understanding will ever invent. Theory can never provide a complete representation of reality, because nature will always remain more abundant than our representations. Nature is characterized by an a priori complexity that will inevitably confound the capacity of the human inquirer.^[2]

Similarly, chaos theory seeks to find patterns and connections, while jettisoning the requirement that the patterns must be ruled by law-like necessity. Researchers, fascinated by the development of chaotic motion, look for regularities, using computer graphics to identify and explore ordered patterns that would otherwise be buried in reams of computer output. But, there is a difference between searching for patterns and searching for laws. These researchers studying chaotic motion realize that the patterns they discern will not have a unique evolution or act predictably.

Traditional science has been based upon the power to predict and control, but chaos theory asserts that when systems become chaotic, science no longer has this ability. The defining characteristic of chaos theory is a sensitivity to initial conditions, such that for chaotic systems, we will never be able to measure the initial conditions accurately enough to get an accurate prediction. It is inherently impossible, because if the initial configuration of a chaotic system is changed even slightly, its future behavior is altered radically.

It is far better to consider chaos theory as a search for order, a concept broader—as Keller pointed out—than the concept of law. The determinate randomness of chaotic systems can be better understood as conforming to a sense of order rather than law, because we are studying behavior that has intrinsic causes, that is, it is orderly, but remains irregular and unpredictable. Thus, chaos theory by challenging the notion of scientific omniscience validates the insight of feminist epistemologists that scientific knowledge is inherently fallible.

NATURE AS AGENT

The Newtonian vision of the world as a clockwork mechanism implies that physical matter is inert and dead. Feminist epistemologists argue that this metaphor of the world as a machine has functioned as a justification for power over and domination of nature. To move beyond a notion of knowledge as power and of power as domination, feminist philosophers of science have attempted to restore agency to nature. Thus, it becomes important for Keller that nature not be bound by the necessity of law but be understood as generative and resourceful. She argues that Noble Laureate Barbara McClintock's work on genetic transportation was possible because McClintock had a belief in the resourcefulness of the natural order. Her discovery of the capacity of organisms to reprogram their own DNA points to the existence of forms of order more inventive than those we had previously envisioned.^[3]

Haraway also suggests that we grant the status of agent to the objects in the world. She speaks of the world as having a sense of humor and as resisting being

reduced to mere resource. To express a concept of nature that eludes being completely captured by the human imagination, she appropriates the metaphor of the coyote trickster of Native American legend. We keep pursuing the wily coyote all the time, knowing that it will inevitably get the better of us.^[4]

Chaos theory also finds nature to be generative and resourceful. If the Newtonian paradigm reduced nature's complexity and simplified it into predictable, law-like behavior, chaos would allow for an understanding of nature rich in disorder and surprise. Even though a chaotic system is deterministic, we cannot predict what it is going to do next; this indeterminacy can make for resourcefulness. For example, the chaotic path of a prey in flight can be used to evade a predator.

ATTENDING TO DIFFERENCE

The model of scientific law in the Newtonian paradigm is based on a universalism in which difference disappears. In contrast, feminist epistemology advocates a methodology that is sensitive to difference and that makes difference intelligible rather than making it disappear as a mere anomaly to a general rule. Keller notes that for McClintock, the discovery of anomalous kernels of corn pointed, not to disorder or unlawfulness, but to a larger system of order that is not reducible to a single law. Similarly, Haraway's concept of "situated knowledges" reflects a commitment to pluralism and difference. Knowledge that is situated demands the participation of as many perspectives as possible and recognizes that the answers that come out of a plurality of perspectives can never be reduced to unity.

In the same manner, chaos theory by designating disorder to be complex information makes difference meaningful. Researchers in human physiology are now looking to the chaotic dynamics of the body to better diagnose and predict illness. Epileptic seizures and heart attacks were once understood to be rhythmic disorders in contrast to healthy, ordered rhythms. Now physiologists believe that chaos provides a healthy flexibility to the heart and brain. And in fact, healthy systems are characterized by innate variability, while impaired systems are marked by a transition to more ordered, less complicated states. Investigators are beginning to think that, contrary to accepted belief, healthy bodies do not seek homeostasis but rather that chaos is the natural way to put together the systems of the body.^[5]

Researchers are learning, therefore, that a lot of useful information can be attained by attending to the variability of a system as well as to its stable order. Chaos theory by alerting scientists to attend to difference validates the feminist insistence that

difference matters and that discoveries come out of a plurality of perspectives that can never be reduced to a unity.

NON-REDUCTIONISM

The method of the Newtonian model is reductionistic. To understand how a clock works, you break it down into its parts, study each element individually, and reassemble it to understand the whole. Study of the individual unit is more important than the study of constituent interactions. Keller argued that scientific techniques contribute to and are product of the ideology of a culture. She analyzed how the reductionist model when applied to biology reinforces values stereotypically associated with masculinity: individuality, autonomy, competition, and dominance. Both population genetics and mathematical ecology, for example, give precedence to competition over cooperative interactions. Genes are understood to be inherently selfish. And when the existence of cooperative behavior is acknowledged, it is reidentified as "cooperative competition."^[6] In genetics, DNA is understood to be a "master" molecule that encodes and transmits all instructions for cellular development.

Chaos theory helps undermine the methodology of reductionism. It privileges qualification over quantification and recursive symmetries over individual units. In the Newtonian paradigm, a scientific theory without quantification was not a theory, but for dynamic systems, quantification is frequently not possible.

Instead, qualitative understanding is achieved by studying what the whole looks like as it moves through changes by watching the activity of a strange attractor. To study turbulence or a fractal, it is necessary to assume that the part is the whole rather than to make the reductionist assumption that the whole is the sum of its parts.

To conclude, chaos theory has opened up a space within scientific inquiry for a notion of knowledge that is fallible, for a restoration of agency to nature, and for a methodology that attends to difference and that moves beyond reductionism. And in so doing, chaos theory represents the revisioning of nature and scientific inquiry sought by feminist epistemologists.

REFERENCES

1. Harding, S. *Whose Science? Whose Knowledge?* Cornell University Press: Ithaca, NY, 1991; 138–163.
2. Keller, E. *Reflections on Gender and Science*; Yale University Press: New Haven, CT, 1985; 75–176.
3. Keller, E. *A Feeling for the Organism: The Life and Work of Barbara McClintock*; W. H. Freeman: New York, 1983; 121–138.
4. Haraway, D. *Simians, Cyborgs, and Women: The Reinvention of Nature*; Routledge: New York, 1991; 183–202.
5. Poole, R. Is it healthy to be chaotic?. *Science* **1989**, *243*, 604–607.
6. Keller, E. *Refiguring Life: Metaphors of Twentieth-Century Biology*; Yale University Press: New Haven, CT, 1995; 99–118.

Feminist Comparative Policy

Amy G. Mazur

*Department of Political Science, Washington State University,
Pullman, Washington, U.S.A.*

INTRODUCTION

It has been over thirty years since the second-wave women's movements took to the streets to change society in western democracies. In part, as a response to women's movement demands, governments across the ideological spectrum have developed a complex and varied set of machineries and policies. These policies and agencies are specifically charged with promoting women's rights and status and, in some cases, with addressing the deep-rooted causes of gender-based inequalities between men and women. Like the women's movements that inspired them, they have had varying levels of success in achieving ambitious goals. Moreover, government policies and structures that promote women's status and rights, as well as gender equality, have the potential to better represent women as a group, both substantively and descriptively, and in doing so, to make democracies more democratic.

BACKGROUND INFORMATION

Since the late 1980s, a new area of study has taken shape within the purview of political science to study, assess, and develop empirically based theory about the impact, both direct and indirect, of these gender-based political dynamics between women's movements and the state. This field is distinct from non-feminist work in political science and policy studies that has tended to ignore gender-specific dynamics and is also distinct from work in feminist studies that takes a critical view of social science inquiry and state action. Filling in the gaps of and building from scholarship in Comparative Politics and Policy, as well as Gender and Politics, Feminist Comparative Policy (FCP) focuses specifically on feminist policy dynamics through a distinct approach and research community. FCP scholars seek to answer the highly complex questions of if, how, and why contemporary, Western, post-industrial democracies are feminist. FCP focuses on Western, post-industrial democracies and not on any countries outside of the West, given that the women's movement and its state responses in these countries share basic, common patterns that are a product of the particular level of economic, political, and social development found in the 23 countries in this category.

An ever-growing community of over one hundred researchers in Europe, North America, Australia, and New Zealand, share a common comparative approach to the study of gender, policy, and the state. By 2006, FCP scholars produced a significant body of literature deserving attention and created permanent research and teaching infrastructures on three continents. The goal of this article is to map out this relatively new area of study. It first examines the approach and core concepts of FCP and then turns to a discussion of the people, infrastructure, and some of the more prominent projects in the discipline.

Given the centrality of the term feminist for FCP, it is important to provide a working definition of this oft-contested concept. Feminism is defined here using the approach of recent FCP work that recognizes the core ideas of western feminism from which academics, activists, and policy practitioners choose. These ideas include: (1) a certain understanding of women as a group within the context of the social, economic, and cultural diversity of women; (2) the advancement of women's rights, status, or condition as a group in both the public and private spheres; and (3) the reduction or elimination of gender-based hierarchy or patriarchy that underpins basic inequalities between men and women in the public and private spheres.

APPROACH AND CORE CONCEPTS

Feminist policy scholars in Western Europe first acknowledged the empirical gaps and gender biases in theory and methodology used in the study of the state and public policy in the early 1980s. By the early 1990s, researchers in North America and Australia joined their Western European counterparts in the new feminist academic enterprise that sought to systematically study the interface between gender and the state. In the mid 1990s, a loose consensus formed in this transnational community around conventions for conducting research, developing theory, and reporting findings, moving the field into a new stage of vitality and institutionalization. In 2006, with (1) over 400 published pieces, over one-fourth of which were published in the past five years, (2) an estimated 20 million euros received in research funding since the mid 1980s,

and (3) four journals that serve as major outlets for FCP work—*Social Politics*, *International Journal of Feminist Politics*, *Politics and Gender*, and *Women and Politics*—FCP scholars and their work constitute a formidable presence in political science and public policy studies.^[1] The FCP approach includes the following six features.

Feature 1: An Applied Feminist Empirical Approach

Labels commonly used to identify different strands of Western feminist thought are not useful in identifying the feminist approach of FCP, e.g., liberal, radical, or socialist. Individuals who contribute to this new field mostly carry-out a feminist agenda through designing, conducting, and disseminating their research and teaching about the topics they study. Some FCP scholars participate in women's movements and organizations and serve periodically as experts on government women's eliminate commissions. Most researchers, however, pursue their political action in academia. Unlike many feminist scholars who seek to transform or revolutionize the study of politics, FCP scholars have a more moderate agenda aimed at working within the rules of the scientific arena and contributing to nonfeminist and feminist bodies of knowledge. They tend to eschew a separatist approach to research; most would agree that men and women should conduct feminist research.

Reflecting what Sandra Harding refers to as an “empirical feminist” approach, FCP scholars agree that past social science research was androcentric, dominated by men, with research problems and methodology excluding the considerations of gender and women.^[2] At the same time, they do not seek to dismiss the entire social scientific process. FCP scholars recognize that social science inquiry is not entirely value-free. In their view, social scientists that claimed to be conducting objective science in the past essentially ignored empirical gender policy issues in the formulation of research problems and designs.

Most FCP practitioners use empirical observation to test and explore hypotheses in order to contribute to theories about gender, politics, and the state. FCP studies are often designed so that findings can be used to help policy practitioners and activists in women's policy agencies, political parties, movements, and organizations learn more about the causes of gender-based inequities and the complex range of solutions, including different ways of designing good practices. Since the early 1980s, FCP scholars have been consulted regularly in their expert capacity by the European Union (in the context of the Action Plans on gender equality), the Council of Europe, the Nordic Council,

and the United Nations, as well as numerous country-based commissions at national and sub-national levels.

The applied feminist empirical approach underpins FCP scholar attitudes about identifying with specific feminist currents of thought and about adhering to a single fixed definition of what constitutes feminist policy, feminist states, and feminist political action in general. FCP scholars tend to avoid articulating a specific political stance on feminism in their research and teaching. They do not deny adhering to a certain type of feminist politics; they just do not apply their specific political views of feminism to their research. The definitions of feminism that FCP researchers use in their analyses, therefore, come from comparative studies of feminist politics, rather than their own feminist politics. FCP scholars tend to be leery of pinning down a single narrow definition of feminism connected to a specific current of Western political thought. In recent years, many agree that there is a growing consensus around broad definitions of feminism, like the one proposed above, that take into consideration the diversity of feminist ideas in Western countries and extend across national boundaries.

Feature 2: Operationalizing Normative Feminist Theory on Democracy

Normative political theory has been an integral part of the development of feminist studies in general. A major question asked by the extensive literature on feminist theory is whether Western democracies are as democratic as observers think, particularly given the degree to which women and women's issues have been excluded from politics, particularly in the context of the formal articulation of universal, gender-blind values of equality, freedom, and representation.^[3] Feminist theorists who write on democracy argue for a better inclusion of women and ideas that favor women's rights in the political process through “descriptive” and “substantive” representation.^[4] These themes of representation and democracy are at the center of FCP studies in so far as they ask the empirical question of whether, how, and why democratic states can be feminist. The question is less about the specific form and design of democracy than its capacity to incorporate women's interests and women themselves into the political process and, in doing so, to promote gender equality and a more complete democratic system.

Feature 3: Bringing the Patriarchal State Back in as a Research Question

As most political scientists agree, particularly since the state was “brought back in” in the 1980s, the concept

of the state—government structures as opposed to country—is not a simple idea. For many feminist theorists, the state is highly problematic, given that it is a product of systems of power based on male domination, or patriarchy. From the assumption of the patriarchal nature of the state, where state actions, structures, and actors seek to perpetuate systems of gender domination that keep women in their inferior positions in the public and private spheres, many feminist analysts dismiss or are highly critical of the state as an arena for positive social change. Other feminist theorists provide a more malleable view of state patriarchy and argue that certain state arenas may be appropriate sites for feminist action. FCP analysts do not entirely dismiss the possibility of a patriarchal state; they see the issue of state patriarchy as a question for empirical research. Some parts of the state may be patriarchal, while other parts may have the potential to be quite woman-friendly. FCP places the state and its institutions at its analytical core. The four major areas of FCP research, outlined in the next section, focus on some aspect of the state or state action.

Feature 4: Using Gender as a Category of Analysis

Since the mid 1980s, feminist research across different disciplines has shifted its focus from sex, a more-or-less dichotomous variable based on biological differences between men and women, to gender, the social construction of sexual difference between men and women. As Joan Scott first asserted in 1986, the relational concept of gender should be the prime “category of analysis” in theoretical frameworks and research designs.^[5] This holistic approach to the use of gender is intended to push analyses beyond an “add women and stir” phase where sex or women are added as an analytical afterthought. Since the mid-1990s, FCP scholars have incorporated gender into their research designs in a variety of ways.

Feature 5: Comparative Theory-Building in Western Post-Industrial Democracies

FCP scholars seek to develop theory through culturally sensitive, comparative, and cross-national analysis. They often employ principles of research design and methods developed outside of a feminist perspective. Up until 2000, with a few exceptions, FCP work utilized small *n* analysis—case studies and the comparative method. In recent years, many studies include a large number of observations and, hence, use the statistical tools of large *n* analysis. Other studies take a bridging approach, incorporating both qualitative and quantitative analysis. FCP studies take a *most similar*

systems approach, where economic and political development in Western, post-industrial democracies are the control variables and variations in nation-based culture, state-society relations, women’s movement mobilization, and government design are examined as they influence gender, state, and policy issues. Most FCP work assumes that post-industrial countries in the West, unlike other countries of the world, share certain environments and institutions, with notable cross-national variations. A part of the common heritage is that women’s movements have developed strategies aimed, at least in part, at influencing the democratic policy process and the development of large welfare states. Designing and using concepts like feminism that are able, as Giovanni Sartori first elaborated, to “travel” across national boundaries without “stretching” the core meaning are also a key part of the comparative agenda of FCP.^[6]

Feature 6: One-Way Intersections with Nonfeminist Political Science

Many FCP scholars since the mid-1990s have actively sought to intersect their work with nonfeminist literature. Instead of completely rejecting traditional political studies or uncritically using feminist studies, FCP work purposefully develops the strengths and shores up the weaknesses of each to advance knowledge in both areas. In general, efforts to intersect work with nonfeminist political science tend to be one-way. FCP scholars seek to contribute to nonfeminist work, and nonfeminist policy analysts ignore FCP theory building that speaks directly to central analytical issues. Even within the context of the increasing methodological pluralism of Political Science, few studies outside of the purview of feminist analysis use the findings or concepts of FCP research or even focus specifically on gender policy dynamics. This trend may be reversing, with the growing academic influence and reputation of FCP scholars and their work.

FCP IN ACTION: RESEARCH COMMUNITY, INFRASTRUCTURE, AND PROJECTS

The vast and growing FCP literature examines four different substantive areas with regards to the state and its institutions. *Feminist Policy Formation* work scrutinizes the ways in which public policy promotes women’s status and strikes down gender hierarchies through studying the obstacles, actors, content, and processes of policy as purposefully feminist. *Feminist Movements and Policy* research is concerned with the interplay between women’s movements, the state, and policy. A major issue of interest here is to evaluate

the success of women's movements in influencing the content of public policy and the structures of the state. *State Feminism* scholarship considers whether state structures and actors can promote feminism through focusing on: (1) women's actors in the state as policy-makers; (2) the gendered nature of state agencies that influence women's (and men's) roles; and (3) the activities of women's and gender-equality policy machineries and the bureaucrats, or femocrats, that staff them, in a wide variety of government agencies and branches. State feminism work assesses the links between the substantive representation of women's interests in policy and the state and the presence of women in elected office or descriptive representation. *Gender and Welfare State* literature examines the welfare state as a prime obstacle and promoter of both gender discrimination and equality. Much of this scholarship looks at the link between women's and men's roles in the public and private domain in relation to the policies of the welfare state.

In 2006, there are roughly one hundred scholars who work regularly on FCP research. To be counted as a part of the FCP community, individuals need to have two or more publications on an FCP topic. The list of active FCP scholars used for this analysis is available on request.

Most members of the FCP scientific community are women. More than a result of any collective decision to exclude men, their absence is a result of the realities of graduate education; women tend to be more interested in the study of feminist politics, and male students tend to be channeled away from what is often perceived as a less prestigious area of study. Less than one-fourth of FCP scholars are from the United States and nearly three-fourths are from Europe. One-half of FCP scholars is based in English-speaking countries—Ireland, the United Kingdom, Canada, Australia, and the United States. The other one-half is based in continental Europe—the Netherlands, Germany, Austria, Italy, Spain, Denmark, Norway, Sweden, Finland, France, Belgium, and Greece. Controlling for national population and size of university infrastructure, the FCP community is quite evenly spread among the different countries. Many FCP researchers are trained in one country and then work and live in another.

FCP practitioners began to develop research networks in the early 1980s. These networks often meet at the conferences of the European Consortium of Political Research, the American Political Science Association, the International Political Science Association, or the International Studies Association. At least ten FCP books have come out of ECPR workshops. In the 1990s, FCP scholars increasingly developed multinational research projects. Once in place, the network members secure funds from a combination of supranational, national, university, private funding agency,

and other institutionalized venues that are not overtly feminist. These projects implement a complex, long-term, gendered research program, produce a series of publications, and often create more permanent infrastructure, such as newsletters, journals, Web sites, graduate funding, and permanent teaching programs.

A selection of FCP projects is presented here. Several research networks have been active since the mid-1990s in the area of state feminism. A three-person British research effort on gender and urban government, entitled *Gender and the New Urban Governance*, received funding from the U.K. Economic and Social Research Council and was based at the London School of Economics.^[7] The Andrew and Rodgers edited book, *Women and the Canadian State*, originated in a conference sponsored by the SSHRC, the Department of Justice, the Department of the Secretary of State, and the University of Ottawa.^[8] It assesses the impact of twenty years of state feminism in Canada. A partnership between nationally based feminist researchers and the European Union, through the Community Action Programs on Equal Opportunities for Women and Men, has been an important catalyst in the development of a consolidated FCP community.

An update of a ground-breaking study of gender, policy, and state in Nordic countries was published in English, as well as in all of the Scandinavian languages, in 1999.^[9,10] This study cuts across all four areas of FCP research. The Gender, Empowerment, and Politics project in Denmark is also conducting research that addresses the analytical issues from the FCP paradigm. Highlighting the important role of Norwegian scholars in the FCP community, the Institute for Women's Studies and Gender Research at the University of Oslo created an e-mail list to report on women's studies-related networking throughout the world. A new Centre for the Advancement of Women in Politics at Queens University of Belfast promotes research and teaching on FCP-related topics as well (<http://www.qub.ac.uk/cawp/>).

The Women and Restructuring Network in Canada, funded by a Social Sciences and Humanities Research Council grant (SSHRC) and a grant from the Canadian International Development Agency, examines how state restructuring has affected women and women's movements in the context of the global economy. At least three publications have been inspired by this network.^[11–13] The project entitled *Public Policy Discourses, Mass Attitudes and the Governance of Gender Equality in Europe*, at the University of Brèmen, funded by the German government, is a seven-nation study on the institutionalization of gender equality norms in European Union member states from 1975 to 2000.^[14] The long-term study of the impact of European Union gender mainstreaming on policy debates and public discourse, MAGEEQ, in

six European countries, has received over 5 million Euros in funding from the European Union (<http://www.mageeq.net/>). A trans-Atlantic effort comparing women's movements and state dynamics in North America and Europe recently published its findings in *Women's Movements Racing the Reconfigure State*.^[15]

Founded in 1995, the Research Network on Gender, Politics, and the State (RNGS) is composed of 40 members. RNGS designed and is currently completing a large-scale comparative research project that examines if, how, and why women's policy offices, through their relations with women's movements, make post-industrial democracies more democratic and the state more feminist. In order to examine both substantive and descriptive representation of women, the research explores changes in the policy making process, through the gendering of policy debate frames, brought about by the interface between women's movements and women's policy machinery from the late 1960s to 2002. Fifteen teams in fourteen countries—the United States, Canada, Italy, the United Kingdom, Ireland, Germany, the Netherlands, Spain, Austria, Belgium, Sweden, France, Finland, and Australia—have been collecting data for this project since 1997.

The interface between agency and movement is explored in individual policy debates on five key issues: job training, abortion, prostitution, political representation, and a top-priority issue of choice specific to each country in the study. Data for 140 policy debates will be gathered by the project's completion. Results have or will be published in five edited volumes, one for each issue area,^[16–20] an article that analyzes the findings quantitatively across all five issue areas, and a capstone book that brings together the qualitative and quantitative strands. The qualitative findings of the five issue areas are in currently being transposed into a data set that can be used by researchers and policy makers.^[21] The dataset will be available for public use in 2007 and will be archived at the major social science data archives in the countries of the study, as well as on the network's website, <http://libarts.wsu.edu/polisci/rngs>. RNGS has received over one-half of a million dollars in grants to fund research and network costs from the French Ministry of Social Affairs, the National Science Foundation (United States), the British Economic and Social Science Research Council, the Canadian National Research Council, the European Science Foundation, and members' home institutions.

CONCLUSIONS

To be sure, FCP research is in a period of vitality and growth. With an empirical feminist approach, FCP

scholars are increasingly contributing to the study of policy and politics in Western, post-industrial democracies. It is increasingly difficult to conduct a study of state-based politics without mentioning gender-specific elements. So, FCP researchers are making non-feminist theories more complete and sound. At the same time, it is of fundamental importance that scholars who do not share the FCP agenda take account of FCP work. It is still quite common to find policy studies and textbooks on adjacent topics that completely ignore, or only marginally discuss, gender considerations. More than any systematic resistance to feminist policy work, we are often all too busy to learn about new areas. Thus, this article provides a go-to source for busy policy scholars unaware of this new field. FCP scholars throughout the world are reaching out to policy makers and activists, beyond the halls of academia. Through their work on government reports, expert consultations, and as members of state-based commission, practitioners of FCP, arguably more than many fields of study, have the potential to contribute to making our democracies more democratic.

REFERENCES

1. Mazur, A.G. *Theorizing Feminist Policy*; Oxford University Press: Oxford, 2002.
2. Harding, S. *Conclusion: Epistemological Questions Feminism and Methodology*; Indiana University Press: Bloomington and Indianapolis, 1987.
3. Squires, J. *Gender in Political Theory*; Polity Press: Cambridge, 1999.
4. Pitkin, H.F. *The Concept of Representation*; University of California Press: Berkeley, 1967.
5. Scott, J. Gender: a useful category of historical analysis. *Am. Hist. Rev.* **1986**, *91*, 1053–1075.
6. Sartori, G. Concept misformation in comparative politics. *Am. Pol. Sci. Rev.* **1970**, *64*, 1033–1053.
7. Abrar, S.; Lovenduski, J.; Margetts, H. Sexing London: the gender mix of urban policy actors. *Int. Rev. Pol. Sci.* **1998**, *19*, 147–171.
8. Andrew, C., Rodgers, S., Eds.; *Women and the Canadian State*; McGill-Queen's University Press: Montréal, 1997.
9. Haavio-Mannila, E.; Dahlerup, D.; Eduards, M.L. *Unfinished Democracy: Women in Nordic Politics*; Pergamon: New York, 1985.
10. Berquvist, C., Borchorst, A., Christensen, A.-D., Ramstedt-Silén, V., Raaum, N.C., Styrkaskottir, A., Eds.; *Equal Democracies? Gender Politics in the Nordic Countries*; Scandinavian University Press: Oslo, 1999.
11. Bakker, I., Ed.; *Rethinking Restructuring: Gender and Change in Canada*; University of Toronto Press: Toronto, 1996.
12. Brodie, J. *Politics on the Margin: Restructuring and the Canadian Women's Movement*; Fernwood: Halifax, 1995.

13. Brodie, J., Ed.; *Women and Canadian Public Policy*; Harcourt Brace: Toronto, 1996.
14. Liebert, U., Ed.; *Europeanisation and Gender Equality. Reframing Public Policy in EU Member States*; Peter Lang: Bruxelles, 2003.
15. Banaszak, L.A., Beckwith, K., Rucht, D., Eds.; *Women's Movements Facing a Reconfigured State*; Cambridge University Press: Cambridge, 2003.
16. Mazur, A.G., Ed.; *State Feminism, Women's Movements, and Job Training: Making Democracies Work in the Global Economy*; Routledge: New York, 2001.
17. Stetson, D.M., Ed.; *Abortion Politics, Women's Movements and the Democratic State: A Comparative Study of State Feminism*; Oxford University Press: Oxford, 2001.
18. Outshoorn, J., Ed.; *The Politics of Prostitution: Women's Movements, Democratic States, and the Globalization of Sex Commerce*; Cambridge University Press: Cambridge, 2004.
19. Lovenduski, J., Sainsbury, D., Baudino, C., Guadagnini, M., Meier, P., Eds.; *State Feminism and Political Representation*; Cambridge University Press: Cambridge, 2005.
20. Haussman, M., Sauer, B., Eds.; *Gendering the State in the Age of Globalization. Women's Movements and State Feminism in Post Industrial Democracies*; Rowan and Littlefield, 2006.
21. Mazur, A.G.; McBride, D. The RNGS dataset: Women's Policy Agencies, women's movements and policy debates in western post industrial democracies. *French politics* **2006**, 4, 209–236.

Financial Condition

Jane Beckett-Camarata

Department of Political Science, Kent State University, Kent, Ohio, U.S.A.

INTRODUCTION

The regular reoccurrence of state and local governments' fiscal problems over the last 30 years has pointed to the need for an effective system to provide citizens, creditors, managers, legislative and oversight bodies, and others with early warning signs of pending fiscal stress. The widespread financial problems that occurred in the 1970s in cities throughout the United States resulted in changes in financial reporting nationwide. Pressure from the credit market and effective leadership of financial organizations have had major impacts, leading to improvements in financial reporting standards as well as compliance with those standards. Because of the implementation of those standards, potential fiscal problems are likely to be discovered sooner and prevented.

Thus, preventing financial problems requires early alert analysis to proactively identify trends and practices that may adversely impact financial condition. Assuring that government has the financial capacity to sustain desired public services is the primary reason for internal managers to monitor financial condition. Bond investors and creditors also have a similar interest in evaluating financial condition, particularly to assess a government's ability to make future interest and principal payments, even in the face of adverse economic trends or natural disasters. Knowledge that internal managers are employing a system to track financial trends provides investors and creditors with added confidence in the quality of the government's financial management. This is true particularly if such trend data are shared with credit analysts. Credit analysts also have more than a passing interest in the government's ability to provide services and pay down debt in the long run. Experience has shown that, in times of fiscal crisis, expenditures for critical services often take precedence over debt service payments.

Local governments produce many financial reports, both internal and external, every year. Financial data are presented in accordance with state and/or professional oversight bodies' requirements. Some states impose uniform financial accounting and reporting systems that all municipalities within the state must follow. Local governments may also be required to submit annual financial reports to a state oversight

body. State legislatures are responsible for helping formulate sound fiscal policies for the government. However, oversight bodies are responsible for establishing and possibly monitoring fiscal policies of governmental units they are responsible for. Both legislative and oversight groups are responsible for monitoring executive compliance with relevant laws and regulations.

OVERVIEW

The Comprehensive Annual Financial Report (CAFR) includes useful information for financial condition analysis. The essential information is contained in the basic financial statements (BFS), including the notes that are part of them. The BFS are financial statements that must be included in a government annual report to comply with GAAP under the Governmental Accounting Standards Board (GASB) statement no. 34. The required statements are: a government-wide statement of net assets; a government-wide statement of activities; fund financial statements for governmental, proprietary, and fiduciary funds; and Notes to the Financial Statements. The notes are particularly important for determining the structure of the organization, accounting methods used, assumptions made, and other background information that gives the reader a perspective as the financial statements are reviewed. Careful review and analysis of management's discussion and analysis (MD&A) and other required supplemental information (RSI) are fundamental to the new process. Furthermore, financial condition analysis must also take into account additional factors such as debt per capita, tax rates, and citizen income and wealth. However, financial condition analysis can be very complex. Prior to GASB statement no. 34, ratio analysis was limited. However, with government-wide financial statements prepared on a full accrual basis, it will be somewhat easier to use ratios to analyze governments' financial condition. Government financial statements usually do not provide information for more than 1 year, probably because of the amount of information already being displayed on one page. MD&A provides 2 years' worth of data for some numbers. CAFRs

frequently provide 2 years of statements. In performing a financial condition analysis, an understanding of financial trends is critical and, generally, 3 years' worth of data is considered adequate for ratio analysis.

What is the utility of the CAFR, given the complex analytical framework for analyzing financial condition? Many people who analyze public financial data find the introductory and statistical sections and the Notes to the Financial Statements to be more useful than the financial statements for assessing financial condition. This is because the introductory and statistical sections of the CAFR and Notes to the Financial Statements provide critical information on environmental and organizational factors and their relationship to financial factors. The BFS provide helpful information for calculating financial ratios essential to understanding the organization's financial position, which is an important component of financial condition.

WHAT IS FINANCIAL CONDITION?

Basically, financial condition is a local government's ability to finance its services on a continuing basis. It is the ability to: 1) maintain existing service levels; 2) endure local and regional economic disruptions; and 3) meet the demands of natural growth, decline, and change.^[1] The International City/County Management Association (ICMA) defines financial condition as "a government's ability to generate enough cash over thirty to sixty days to pay its bills."^[1] This definition of financial condition is called *cash solvency*. The Governmental Accounting Standards Board defines financial condition as: "The probability that a government will meet both its financial obligations to creditors, consumers, employees, taxpayers, suppliers, constituents, and others as they become due and its service obligations to constituents, both currently and in the future."^[2] In both the ICMA and GASB definitions, financial condition is similar to the accounting term *solvency*. Whether a local government can maintain existing service levels or increasing service levels is related to the concept of *interperiod equity*. Interperiod equity is defined as: "whether current-year revenues are sufficient to pay for the services provided that year and whether future taxpayers will be required to assume burdens to pay for current services to future taxpayers may threaten the government's ability to sustain the current level of services or to expand services to meet future population growth."^[3] Financial condition can also apply to a government's ability to provide services at the level and quality that are required for the health, safety, and welfare of the community and that its citizens desire. This is its *service-level solvency*. A government that does not have

service-level solvency might, in all other respects, be in sound financial condition, but be unable to support police and fire services at an adequate level. If in such condition, the local government tries to provide adequate services; it suffers cash, budgetary, or *long-term solvency* problems.

FACTORS AND INDICATORS USED IN ASSESSING FINANCIAL CONDITION

Financial condition analysis is the examination of the financial status of a government organization based on a financial statement analysis as well as an evaluation of many external factors that affect the financial condition of the government, such as the wealth of the citizenry, employment rate, and general economy. Financial condition analysis is a complex, multidimensional concept that requires the study of a multitude of factors. GASB identifies the major categories of these factors as: economy and demographics; revenue base; current and capital expenditures; debt, pensions, and other postemployment benefits; internal resources; management capabilities; infrastructure; and willingness to raise revenues and provide required public services.^[2] Khan,^[4] in his study of municipal bond ratings, listed several indicators used for assessing and making comparisons of financial condition. These include Per Capita Income, Percent Below the Poverty Line, Age of Housing, and Unemployment Rates. Economy and demography are two of the categories used by ratings agencies to assign bond ratings to municipalities. The ICMA uses indicators such as Vacancy Rates, Property Value, Population, Business Activity, Poverty, Households, Personal Income Per Capita, and Employment Base in its handbook for analyzing financial condition. It identifies these categories as "Community Needs and Resources Indicators." ICMA's financial trend monitoring system was developed for financial managers to use in tracking their governments' financial performance. The ICMA points out that these factors are important because they indicate the interrelationship between community demands for services and the ability of the government to provide them.

"Changes in community needs and resources are interrelated in a continuous cumulative cycle of cause and effect. For example, a decrease in population or jobs lowers the demand for housing and causes a corresponding decline in the market value of housing; this, in turn, reduces property tax revenues. The initial population decline also has a negative effect on retail sales and personal income, causing local government revenues to drop even further. But because of fixed costs in its expenditure structure that are impervious

to declines in population or business activity, the government cannot always balance the revenue loss with a proportionate reduction in expenditures. The government may, in fact, be forced to raise taxes to make up for lost revenues: this puts a greater burden on the remaining population. As economic conditions decline and taxes rise, the community becomes a less attractive place to live and population may decline further. The cycle continues.’’^[5]

WHAT IS THE DIFFERENCE BETWEEN FINANCIAL CONDITION ANALYSIS AND FINANCIAL POSITION ANALYSIS?

The primary difference between financial condition analysis and financial position analysis is that financial position “tends to be a shorter-run concept compared with financial condition.”^[2] According to GASB, “financial position for governmental funds focuses on assets and liabilities that require cash, or are normally converted to cash in the near future and can generally be determined from the financial statements alone.”^[2] Thus financial position is associated with the concept of liquidity.

Internal and external analysts can compare specific ratios to national and state medians published by Moody’s Investors Service, Selected Indicators of Municipal Performance (Moody’s annual), or comparative ratios published by Dr. Kenneth Brown for both cities and counties. Dr. Brown’s Ten-Point Test^[7] can be used to compare the financial position of cities or counties. The Ten-Point Test consists of a comparison of 10 key ratios for a city, county, or school district with similar entities across the nation. The *Test* measures the following elements of financial condition.

- Revenues—the ability of annual revenues to finance government services
- Expenditures—the manner in which revenues are utilized to provide government services
- Operating Position—the extent to which a positive balance between revenues and expenditures exists and the level of sufficiency of liquid assets and reserves
- Debt Structure—the levels of debt, both short term and long term, and the burden of annual principal and interest payments.

Of the 10 key ratios, three measure the adequacy of revenues, one measures expenditure allocation, three measure operating position liquidity, and three measure the debt structure of the entity.

DR. BROWN’S TEN-POINT TEST OF FINANCIAL CONDITION

1. Total Revenues/Population
2. Total General Fund Revenues from Own Sources/Total General Fund Revenues
3. General Fund Sources from Other Funds/Total General Fund Sources
4. Operating Expenditures/Total Expenditures
5. Total Revenues/Total Expenditures
6. Unreserved General Fund Balance/Total General Fund Revenues
7. Total General Fund Cash and Investments/Total General Fund Liabilities
8. Total General Fund Liabilities/Total General Fund Revenues
9. Direct Long-Term Debt/Population
10. Debt Service/Total Revenues.

The analysis in [Table 1](#) uses data from the four similar Ohio cities to compare financial position.

CITY OF CANTON, OH

The City of Canton is the county seat of Stark County in northeastern Ohio. It was incorporated as a village in 1828 and became a city in 1854. The 1998 U.S. Census Bureau estimate of population for the City of Canton was 79,259, a decline of 5.8% from the 1990 census figure of 84,161. Canton is a statutory city under the laws and regulations set forth in the Ohio Constitution. It operates under the mayor–council form of government. The main source of revenue for the operations of the city government is the collection of a 2% income tax. The city provides police protection, firefighting and prevention, street maintenance and repairs, building inspection, parks and recreation, water, and sewer and sanitation.

The results of the ratio analysis for Canton show relative strength in ratios 1, 5, and 10. Ratio 1, Total Revenue/Population, shows Canton to be in the upper quarter of cities. Higher per capita revenues indicate that the city has adequate annual resources to meet its obligations. Ratio 5, Total Revenues/Total Expenditures, is a measure of operating position and measures relative levels of operating surpluses or deficits. Canton’s high score indicated that in FY 1998, it ran a relatively high operating surplus, which helps to build margins for the future. Building margins in good times helps soften the blow of bad times.^[6] The third ratio, where Canton scored in the upper 25th percentile, was Debt Service/Total Revenues. A relatively low indicator of debt structure is important for a city’s ability to meet its debt service obligations in a timely manner.

Table 1 Comparison of financial condition

		Springfield	Canton	Loraine	Mansfield
Ratio 1	Numerator (US\$)	70,628,228	37,502,751	39,124,038	47,261,029
	Denominator (US\$)	79,259	68,857	49,802	65,568
	Ratio	891.107	544.647	785.592	720.794
	Score	2	(1)	2	1
Ratio 2	Numerator (US\$)	40,781,273	18,758,142	5,269,803	24,045,204
	Denominator (US\$)	49,138,452	23,228,448	10,152,822	28,572,261
	Ratio	0.830	0.808	0.519	0.842
	Score	0	0	(1)	0
Ratio 3	Numerator (US\$)	0	0	10,866	0
	Denominator (US\$)	49,138,452	23,228,448	10,162,688	28,571,261
	Ratio	0.000	0.000	0.001	0.000
	Score	1	1	1	1
Ratio 4	Numerator (US\$)	60,468,964	39,817,799	33,725,217	38,045,931
	Denominator (US\$)	66,850,333	42,319,044	38,065,918	47,282,401
	Ratio	0.905	0.886	0.886	0.805
	Score	0	0	0	1
Ratio 5	Numerator (US\$)	70,628,338	37,502,751	39,124,038	47,261,029
	Denominator (US\$)	66,850,333	42,319,044	38,065,918	47,282,401
	Ratio	1.057	0.866	1.028	1.000
	Score	2	0	2	1
Ratio 6	Numerator (US\$)	7,801,374	3,068,901	2,139,373	4,659,711
	Denominator (US\$)	49,138,452	23,228,448	10,153,688	28,571,261
	Ratio	0.159	0.132	0.210	0.163
	Score	0	(1)	0	0
Ratio 7	Numerator (US\$)	5,351,800	1,889,351	2,033,739	3,301,564
	Denominator (US\$)	5,754,138	4,086,572	2,941,238	2,935,719
	Ratio	0.930	0.462	0.691	1.125
	Score	0	(1)	(1)	0
Ratio 8	Numerator (US\$)	5,754,138	4,086,572	2,941,238	2,935,719
	Denominator (US\$)	49,138,452	23,228,448	10,163,688	28,571,261
	Ratio	0.117	0.176	0.289	0.103
	Score	1	0	(1)	0
Ratio 9	Numerator (US\$)	16,530,000	9,202,634	4,300,000	17,289,225
	Denominator (US\$)	79,259	68,857	49,802	65,568
	Ratio	208,557	133.648	86.342	263.684
	Score	1	1	1	0
Ratio 10	Numerator (US\$)	2,726,289	6,292,582	1,189,939	93,527,285
	Denominator (US\$)	70,628,228	37,502,751	39,124,038	47,261,029
	Ratio	0.039	0.168	0.030	1.979
	Score	2	0	2	(1)
Total score		Canton 9	Loraine (2)	Mansfield 5	Springfield 3

Although Canton did not score in the lower 25th percentile in any of the ratios, there are some areas of comparable weakness (25th–50th percentile score). Canton is somewhat reliant on intergovernmental revenue in its general fund (see ratio 2). This leaves the city more exposed to the whims of the Congress or the State legislature. Another area where the city may need more work is in ratio 4, Operating Expenditures/Total Expenditures. Its higher score indicates that the maintenance and building of infrastructure

may not be getting adequate attention. Canton also scored lower in unreserved fund balances and relative liquidity. Lower relative fund balances indicate that Canton might not be as well suited to meet a fiscal emergency as a similar city. Lower levels of cash and short-term investments could indicate an inability to meet short-term obligations.

Canton's overall score was nine. This puts Canton in the 75th percentile according to Dr. Brown's Test. The relative strength in the current position means that

Canton, at least in the short run, should be able to meet its obligations in both providing services to its citizenry and meeting its financial commitments.

CITY OF LORAIN, OH

Loraine is the largest city in Loraine County, located in northeastern Ohio. It was incorporated in 1874, and is located about 30 miles west of Cleveland. The 1998 U.S. Census Bureau estimate of population for the City of Loraine was 68,857, a decline of 3.4% from the 1990 figure of 71,245. Lorain, too, operates as a statutory city under the laws and regulations set forth in the Ohio Constitution. Legislative authority is vested in a city council and the executive authority in the mayor (mayor–council government). As in Canton, the main source of revenue from the operations of the city government is the collection of income tax, which represented 63% of general fund revenues. Loraine provides police protection, firefighting and prevention, public health services, street maintenance and repairs, building inspection, parks and recreation, water, and sewer.

Ten-Point Test scores for Financial Position were very low (e.g., no score in the upper quarter in any ratio, two scores in the 50th–75th percentile, and four scores in the bottom quartile of the 10 ratios). Scores for the ratio of revenues to population indicate a deficiency in annual resources. The ratio of operating expenditures to total expenditures is too high, indicating substandard attention to infrastructure. The unreserved general fund balance is too low compared with total general fund revenues, and Loraine seems to have relatively low short-term liquidity as evidenced in its low ratio of cash and short-term investments to total general fund liabilities.

Overall, the City of Lorain scored a –2, putting it in just the 5th percentile. That means that 95% of cities scored higher. A long-term trend analysis would seem to be prudent to rule out whether there are some underlying long-term issues that need to be addressed.

CITY OF MANSFIELD, OH

Mansfield is the county seat and the largest city in Richland County, located approximately 65 miles northeast of the city of Columbus. Mansfield was incorporated in 1808 and operates with the mayor–council form of government. The 1998 U.S. Census Bureau population estimate was 49,802, a decline of 1.6% from the 1990 figure of 50,627. Like most cities in Ohio, Mansfield relies on municipal income tax as its main source of revenue (50.5%). Mansfield provides traditional municipal services (police, fire, recreation,

streets, etc.) as well as operates an airport, a parking garage, and water and sewer utilities.

Mansfield had three areas of relative strength (75th percentile), and they were the same as Canton. The city's per capita revenue indicated adequate resources, its ratio of total revenues to total expenditures was strong, and its debt service levels were manageable. Mansfield did score in the lower quartile in three areas. The city seems too dependent on intergovernmental revenues in its general fund, indicating some vulnerability to outside governmental entities. Its liquidity was relatively low, whereas the ratio of general fund liabilities to general fund revenues was too high, indicating that short-term obligations might not be easily serviced by the normal flow of annual revenues. Mansfield scored a five overall on Dr. Brown's Ten-Point Test, which puts the city firmly in the middle at the 50th percentile. That means half of the cities analyzed will score higher and half will score lower.

CITY OF SPRINGFIELD, OH

Springfield Ohio is the county seat and largest of Clark County, located in southwest Ohio. It is the only city in this analysis that is located in the southern part of the state. It was incorporated in 1850, although its roots date back to the 18th century. The city enacted a charter in 1913 that provides home-rule powers to the city under a city commission–manager form of government. The 1998 U.S. Census Bureau estimate of population for the City of Springfield was 65,568, a decline of 6.96% from the 1990 figure of 70,487. Municipal income taxes represent 71.5% of general fund revenues. The city provides traditional municipal services, as well as operates an airport, golf courses, and a sewer system.

Springfield did not score in the upper quartile in any ratio in Dr. Brown's Ten-Point Test. It did score in the 50th–75th percentile range in 4 of 10 ratios. Per capita revenues seem relatively healthy, whereas the general fund seems relatively independent from outside revenue sources. The city appears to be investing an adequate amount in infrastructure, and its current year operating position (total revenues/total expenditures) was also slightly above the median.

Springfield scored in the 25–50th percentile in 5 of 10 ratios. Fully 90% of its scores was within the 25th–75th percentile. Springfield scored in the lowest quartile in the 10th ratio, Debt Service to Total Revenues, indicating a weakness in its ability to meet debt service requirements. This performance was reflected in its final score, a three. The score puts Springfield in the 25th percentile, which means that 75% of the cities analyzed scored higher.

In summary, according to Dr. Brown's Ten-Point Test, the four Ohio cities rank as follows:

Rank	City	Score	Percentile
1	Canton	9	75th
2	Mansfield	5	50th
3	Springfield	3	25th
4	Loraine	(2)	5th

The data to calculate the 10 financial ratios are readily obtainable from most CAFRs. Except for population, which is usually disclosed in the statistical section, data for the first four ratios can be obtained from the Statement of Revenues, Expenditures, and Changes in Fund Balances—Governmental Funds and Balance Sheet—Government Funds. Data for the remaining ratios usually can be found in the Notes to the Financial Statements. In determining ratios 1–8, some analysts prefer to utilize general fund data only, whereas others use combined data for all governmental fund types. This decision will depend, in part, on how large the general fund is relative to all governmental fund types. In determining operating revenues, capital project fund revenues should be excluded because the capital project fund is not an operating fund.

In addition to analyzing ratios, the stability, flexibility, and diversity of revenue sources; budgetary control over revenues and expenditures; adequacy of insurance protection; level of overlapping debt; and growth of unfunded employee-related benefits should also be evaluated. Socioeconomic and demographic trends should be analyzed as well, including trends in employment, real estate values, retail sales, building permits, population, personal income, and welfare. This information is found in the CAFR and the U.S. Census.

CONCLUSION

Financial condition analysis relies heavily on analysis of BFS, including the notes that accompany them. The notes are especially useful for determining the

structure of the entity, accounting methods used, assumptions made, and other background information that gives the reader a perspective as the financial statements are reviewed. In addition, financial condition analysis must take into consideration additional external factors such as debt per capita, tax rates, and citizen income and wealth. The use of ratios, such as Dr. Brown's Ten-Point Test or ICMA ratios, is also appropriate for financial condition analysis. Financial condition analysis is complex because the business of government is complex. For example, a local government's population might be growing but property values might be stagnant. There may be high property values but also high debt per capita. In any financial condition analysis, comparison with a relevant group should be made, and strengths and weaknesses should be considered as a whole in making a reasoned assessment about financial condition.

REFERENCES

1. Nollenberger, K. *Evaluating Financial Condition: A Handbook for Local Government (Revised)*; International City/County Management Association: Washington, DC, 2003; 1 p.
2. Berne, R. *The Relationships Between Financial Reporting and the Measurement of Financial Condition*; Governmental Accounting Standards Board: Norwalk, CT, 1992; vii, 16–17, 25.
3. Governmental accounting standards board concepts statement no. 1. In *Objectives of Financial Reporting*; GASB: Norwalk, CT, 1987, Part 61.
4. Kahn, A. Classification models and municipal bond ratings. *Public Budg. Financ. Manage.* **1994**, 6 (2), 183.
5. Groves, S.M.; Valente, M.G. *Evaluating Financial Condition: A Handbook for Local Government*; International City/County Management Association: Washington, DC, 1994; 112 pp.
6. Pollock, R.; Suyderhoud, S.J. The role of rainy day funds in achieving fiscal stability. *Natl. Tax J.* **1986**, 39 (4), 197–485.
7. Brown, K. The 10-point test of financial condition: toward an easy-to-use assessment tool for smaller cities. *Gov. Finance Rev.* **1993**, 9 (6), 21–26.

Financial Emergencies

Jane Beckett-Camarata

Department of Political Science, Kent State University, Kent, Ohio, U.S.A.

INTRODUCTION

Financial emergencies in U.S. local governments are an ongoing, significant, and complex public policy issue for legislators, state, and local officials and citizens. For our purposes, a financial emergency is defined as a formal legal declaration by a state government of local government financial default. Such formal legal declarations are in response to a local government's negative account imbalances (e.g., general fund imbalances); that is, the inability to provide basic services; the inability to meet payroll, service debt, pay vendors, or pay retirement fund obligations; or the inability to meet its significant financial obligations.

Financial (fiscal) emergencies have serious negative consequences for the local government, the community, the region, and the state. The end result of a financial emergency is reduced services, lower quality of life, and a threat to the physical infrastructure and safety. Each state has its own unique way of dealing with inadequate local government fiscal conditions. However, most attempt to address these inadequacies through a four-stage fiscal health/financial emergency continuum. The four stages in the continuum are: 1) fiscal health; 2) fiscal watch or monitoring; 3) fiscal stress; and 4) fiscal emergency. Local governments with fiscal emergencies must develop and implement strategies for eliminating the actual fiscal emergencies as well as the causes. Inadequate fiscal conditions at the local level are a problem for the state as a whole as well as for the local community. Most states have laws that govern fiscal emergencies. For example, the cities of Miami, Florida, and Washington, DC, have been in a state of public receivership. The financial management of both cities was handed over to public boards until fiscal integrity was restored. In the case of Washington, DC, a Memorandum of Agreement between the District of Columbia Financial Responsibility and Management Assistance Authority conveys back to the Mayor of Washington, DC, control over operations, but retains statutory oversight responsibilities.

In addition to state laws governing fiscal emergencies, local governments can file for bankruptcy under Title 9 of the U.S. Bankruptcy Code. If a local government files for bankruptcy under Title 9 and the state is silent on the filing, it is tantamount to authorizing the municipality to file under federal bankruptcy law.^[1]

Bankruptcy applies only to instances for which there has been a formal filing of a bankruptcy petition under [Chapter 9](#) of the Federal Bankruptcy Act (www.accessgpo.gov). To file a bankruptcy petition, the government must declare itself insolvent. Either a general-purpose local government or a special district can file this petition. Orange County, California, filed for bankruptcy because of unsound investment practices.

State-authorized eligibility to file for bankruptcy under Title 9 is a Tenth Amendment issue that is central to bankruptcy proceedings.^[1] Georgia, for example, prohibits municipalities from filing and Pennsylvania requires prior state permission. Other states expressly and specifically grant municipalities the authority to file for bankruptcy.

THE FISCAL HEALTH/FISCAL EMERGENCY CONTINUUM

A local government's fiscal health can be viewed as part of a four-stage fiscal health/fiscal emergency continuum ([Fig. 1](#)). The term fiscal emergency represents stages two, three, and four of this continuum, with each succeeding stage representing a declining level of fiscal health and adequacy. For example, a local government in stage one is in sound fiscal health. In stage two, the locale is showing early signs of fiscal problems such as general fund account imbalances. In stage three, the fiscal concerns are much more significant in that the locale not only has a general fund imbalance, but is also unable to provide adequate basic services to its community. In stage four, in addition to the items in stage three, they cannot meet payroll, service debt, pay vendors, pay retirement fund obligations, or meet other significant financial obligations. Stage four is a true fiscal emergency.

ANALYZING FISCAL HEALTH

Over the past several years, the regularity of financial emergencies in several states has confirmed the need for an effective system to provide taxpayers, creditors, managers, and legislatures, and others with timely notice of the situation. In most instances, financial constraints, coupled with financial reports (both internal

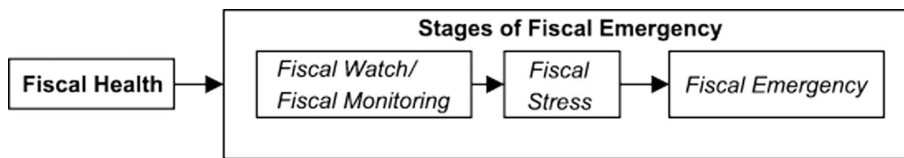


Fig. 1 Fiscal health/fiscal emergency continuum.

and external) of the local governments, did not provide information as to the nature of the fiscal problems until it was too late. In spite of improvements in financial reporting, fiscal emergencies continue to occur. Preventing fiscal emergencies requires an ongoing system of financial monitoring that can provide a warning of fiscal problems that may affect the community.

Credit Analysis

Credit agencies (e.g., Moody's, Standard and Poor's) examine both *structural* and *budgetary* factors, such as 1) economic; 2) financial; 3) debt; and 4) administrative/political. Standard and Poor's emphasizes debt factors (25). Moody's emphasizes the issuer's economic base.^[2] However, both Moody's and Standard and Poor's examine factors in each of the four groups. The ratings are viewed as one indicator of long-term financial condition. For example, Moody's Aaa rating indicates a city is likely in better financial position than a city with a Baa rating (the lowest *investment grade* rating). Further, ratings assigned to general obligation (GO) bonds often apply to all GO bonds of the same issuer.

A local government's fiscal health is its ability to meet its financial obligations as they become due and to deliver fundamental public services to its residents. One way that fiscal health can be assessed is by analyzing a local government's fiscal health over time. In addition to looking at the changes in financial indicators of fiscal health and relationships among indicators, environmental factors such as community needs and resources, economic conditions, political culture, intergovernmental constraints, and natural disasters and emergencies largely determine revenue capacity and demand for services. The way in which a local government responds to environmental demand and changes in the environment is a major factor in determining how environmental factors are translated into fiscal health.

State Government Analysis

There are various ways to analyze fiscal problems. A municipality in Ohio, for example, may fail certain policy criteria for fiscal emergency at year-end, but correct its fiscal situation in time to avoid state control. For example, existence of any year-end emergency criteria

is not a fiscal emergency if the emergency criteria do not exist at the time of the fiscal emergency audit.^[3] Ohio policy requires calculations at the prior year-end at the time of the fiscal emergency audit. Although the Ohio fiscal emergency policy delineates fiscal emergency, it is silent on the causes. It is also silent on fiscal emergency prevention strategies. The Ohio fiscal emergency policy is primarily intended as a remedial action to restore a local government's fiscal soundness. The Ohio State Auditor declares that a particular local government has a fiscal emergency. The State Auditor may initiate a review based on problems through its audit function of the state, or the governor, mayor, or presiding officer of the local government can request a fiscal emergency analysis.

Ratio Analysis

An important part of analyzing fiscal health is the use of ratios. Ratios help management and external users proactively understand the local government's strengths and weaknesses. A better understanding of strengths and weaknesses can lead to corrective action before a fiscal emergency occurs. In financial statement analysis, ratios are used to compare numbers taken from the financial statements. For instance, knowledge about how much a local government had in assets compared with current liabilities may be desired. If the Town of Bainbridge, for instance, had \$61,900 in current assets compared with its current liabilities of \$14,000, the comparison is done by dividing one number by the other. If we divide \$61,900 by \$14,000, the result is 4.4. This means that there are 4.4 dollars of current assets for every one dollar of current liabilities. This would be referred to as a ratio of 4.4 to 1 or simply a ratio of 4.4. This ratio is called a *current ratio*.

Ratios are used because the comparison between two numbers generates information that is more useful than either or both of the numbers separately. If we know that the town has \$61,900, that does not tell us whether it has adequate short-term resources to meet its obligations. Knowing that we have 4.4 dollars of liquid assets for each dollar of short-term liabilities tells us much more about the relative safety of the organization. Each local government has to develop its own set of useful ratios about its fiscal health. For example, the ratio of total revenues to total expenditures indicates the relationship of inflow from

revenues to outflow for expenditures; a high ratio is viewed as positive.

In addition to evaluating ratios, the stability, flexibility, and diversity of revenue sources; budgetary control over revenues and expenditures; adequacy of insurance protection; level of overlapping debt; and growth of unfunded employee-related benefits. Socio-economic and demographic trends should also be analyzed, including trends in employment, real estate values, retail sales, building permits, population, personal income, and welfare. Much of this information is contained in the statistical section of the Comprehensive Annual Financial Report (CAFR), if the local government prepares a CAFR.

Operational Definitions

There are two basic *operational* definitions of fiscal emergencies: 1) *structure* and 2) *budget*. The *structural* definition outlines a fiscal emergency in terms of the *external* constraints that are placed on the local government's fiscal health by the local economy (its economic base) and by the population demographic structure. The *budget* definition characterizes fiscal emergencies in terms of internal *budget* actuality. That is, the inability of local governments to balance their budgets, regardless of the causes of the problem. For example, fiscal mismanagement can cause a *budgetary* fiscal emergency, as well as sudden changes in local revenues (loss of major taxpayer or intergovernmental aid) or change in community needs (e.g., a catastrophe or sudden growth).

Legal Definitions

Within the budgetary category, there are legal definitions for three categories of budgetary fiscal problems. For example, in Ohio, the state auditor declares: 1) *fiscal watch/fiscal monitoring*; 2) *fiscal stress*; and 3) *fiscal emergency, under the following conditions*:

1. *Fiscal Watch/Fiscal Monitoring*: This status is declared to provide early warning to local governments whose finances are approaching fiscal stress. Local governments are monitored by the state when a deficit exists in the General Fund or other funds and when such deficit exceeds one-twelfth of total General Revenue Funds. It is declared when any one of the following factors exist:
 - All accounts that are due and payable from the General Fund for more than 30 days, less the year-end balance of the General Fund

exceeds one-twelfth of the General Fund budget for the year.

- All accounts that are due and payable from all funds for more than 30 days, less the year-end balance in these funds, exceeds one-twelfth of the available revenue for the preceding fiscal year from these funds.
- Total deficit funds, less the total of any balances in the General Fund and in any special fund that may be transferred to meet such deficits, exceeds one-twelfth of the total General Fund budget for that year and the receipts to those deficit funds during that year (other than transfers from the General Fund).
- Money and marketable investments, less outstanding checks, less total positive fund balances of general fund and special funds, exceeds one-twelfth of the total amount received during the preceding fiscal year.

2. *Fiscal Stress*: This status is declared when and one of the fiscal watch/fiscal monitoring factors are present, and there is an inability to fund service delivery needs and requirements because of a decrease in resources and an increase in demands for resource utilization.
3. *Fiscal Emergency*: This status is declared when, in addition to fiscal stress conditions, there is a failure to meet significant financial obligations. This classification includes cases in which governments have failed to meet principal and interest payments on their debt; failed to meet payrolls, pay vendors, or pay retirement fund obligations; or failed to meet other significant financial obligations. In most instances, these emergencies are confirmed by the state declaring the local government in a financial emergency.^[3]

CAUSES OF FISCAL EMERGENCIES

There are five broad *causes* of fiscal emergencies:

1. Socioeconomic decline (i.e., "regional or national forces, principally economic and demographic, that have weakened city tax bases and strained local services").
2. Local political factors largely relating to disproportionate demands for public services from specific groups or from unions, changing electoral coalitions that call for increased government spending to reward the partners of winning coalitions, and local political cultures that are supportive of different levels of public spending.

3. Government growth through inevitable dynamics of bureaucratic expansion, giving rise to increased expenditures.
4. Poor fiscal management practices.
5. State policies limiting revenue and expenditures.^[4]

The first cause relates to the *structural* definition of fiscal problems, as discussed above. The last four relate to the *budgetary* definition of fiscal problems also discussed above.

Some cities are candidates for fiscal emergencies because some local government managers take a short-term view, surviving each fiscal year by “getting by” or short-run tactics that may detract from the local government’s long-term fiscal integrity. Although they cannot correctly be described as having financial mismanagement, two types of shortcomings still exist. First, it is difficult to determine a local government’s vulnerability to a fiscal emergency from its published budget, balance sheet, and income and expense statements without also being familiar with its local policies and accounting method. Second, comparison of operating revenues and expenditures is also challenging if there is no capital budget, or if there are numerous operating funds.

Inferences drawn from the both the literature and state fiscal emergency laws might suggest the following interrelationships (Fig. 2).

VISIBLE SIGNS OF FISCAL EMERGENCIES

Some visible signs of a fiscal emergency are:

1. A decline or inadequate growth in revenues relative to expenditures.

2. Declining property values.
3. Declining economic activity (e.g., increasing unemployment, declining retail sales, declining building activity).
4. Erosion of capital plant, particularly infrastructure.
5. Increasing levels of unfunded pension and other postemployment obligations.
6. Inadequate capital expenditures.

Signs such as these, especially if several exist simultaneously, may indicate a potential fiscal emergency unless the local government takes immediate action to increase revenues or decrease spending. The typical sign of a fiscal emergency is fund deficits. There are three fundamental explanations for municipal fund deficits: 1) increased costs associated with state mandated (funded and unfunded) programs; 2) local plant closings; and 3) decreased income tax base.^[5] Mandates, which are typically new unfunded programs or procedures, approved by a state legislature or Congress and implemented at the local level, contribute to fund deficits. Many of these unfunded mandates relate to health care for the indigent population, the administration of the court system, the administration of the prison system, the education system, and the mental health system.

Fund deficits are affected by other constraints placed on municipal governments. For example, municipalities confronting a declining tax base may be unable to raise additional revenue in the short run because of state laws limiting tax increases, bond indebtedness, or the use of special revenue mechanisms. In addition, some states such as Ohio are overly reliant on local income taxes, which are adversely affected by local plant closings.

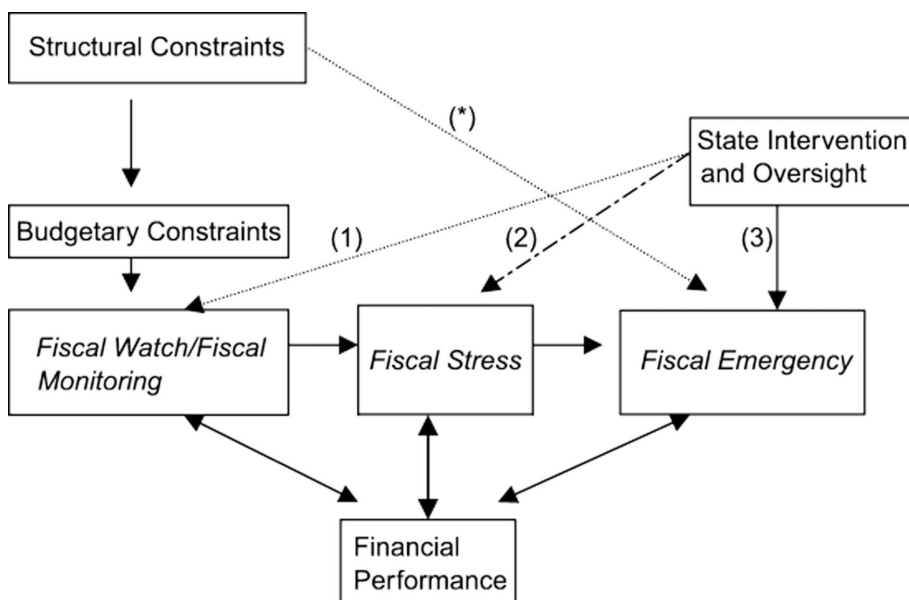


Fig. 2 Fiscal emergency model. (*) Unique catastrophic environmental occurrence; (1) minimal informal monitoring; (2) extensive informal monitoring; (3) formal intervention and oversight.

Taxing powers have a central role in fiscal emergencies. Taxing powers are important because their applicability defines different credit instruments, including general obligation bonds. Taxing powers may be constrained by tax-and-expenditure limitations. Some local tax-and-expenditure limitations have been on the books for decades. The ostensible purpose behind statewide tax-and-expenditure limitations on local jurisdictions is often enforcing governmental economy; the structure, rather than the size, appears more likely to be affected, and the long-term outcome is the centralization of the state-local fiscal system. In some instances, a tax increase may be counterproductive to the community because it is also accompanied by a reduction in services. In the case of Bridgeport, Connecticut, unless the city increased services without increasing property taxes, it would suffer property value losses, job flight, loss of businesses, and a fall-off of middle-class residents. Municipalities are in competition with their surrounding suburbs; therefore, a high proportion of relocatable retail and service-sector jobs and businesses will leave.

Measures associated with loss of financial flexibility include: 1) the proportion of own source revenues dedicated to matching funds; 2) the percentage of expenditures funded from grant funds; 3) the debt burden indicated by the ratio of general obligation long-term indebtedness to municipal population or the ratio of debt service payments to total own-source revenues; 4) the budget overruns indicated by the year-end expenditures over the original budget; 5) the ratio of the legally authorized tax rate over the current tax rate; and 6) the proportion of municipal expenditures allocated for covering mandates.^[6]

STATE FINANCIAL RECOVERY LAWS

States respond to signs of a fiscal emergency through the use of fiscal recovery laws. In Ohio, the fiscal emergency law is primarily intended to restore a local government's fiscal soundness through the use of a fiscal commission and state auditor oversight. The Ohio State Auditor is responsible for certifying that a particular local government has a fiscal emergency. The state auditor may initiate a review based on problems through its audit functions of the state or the governor, mayor, or presiding officer of the local government can request a fiscal emergency analysis.

State Oversight Commission

Once the state auditor has determined that a fiscal emergency exists, the governor appoints a Financial Planning and Supervision Commission to work with

the local government. The commission has eight members:

1. State Treasurer.
2. Director of the State Office of Budget and Management.
3. For municipalities, the mayor and presiding officer of the local government legislative authority.
4. For counties, the president of the Board of County Commissioners and the county auditor.
5. For townships, a member of the board of township trustees and the county auditor.
6. Three appointed members chosen out of five names provided to the governor by the mayor and presiding officer of municipal legislative authority, or by the county board of commissioners or board of township trustees. These individuals must be residents of the declared government (by home or office address) with at least 5 years private sector business/financial experience.^[3]

Responsibilities of a State Oversight Commission

In Ohio, for example, the Oversight Commission is responsible for approving a financial recovery action plan to essentially eliminate fiscal emergency conditions, balance the budget, and avoid future deficits and market long-term obligations. The mayor, board of commissioners, or board of trustees must submit the plan to the commission within 120 days of its first meeting. The commission can either accept or reject the plan. The above officials then have 30 days to resubmit a plan, and the process is repeated until a plan is accepted.

The Oversight Commission has far-reaching authority to: 1) review all revenue and expenditure estimates to ensure a balanced budget; 2) require the government (by ordinance or resolution) to maintain monthly levels of expenditures and encumbrances consistent with the financial plan; 3) approve and monitor expenditure and encumbrance levels; 4) approve the amount and purpose of any debt issues; 5) make and enter into all contracts and agreements necessary to the performance of its duties; and 6) make recommendations for cost reductions or revenue increases to carry out the financial plan.

All financial transactions require commission review and approval. The commission reviews the local government's budgets, tax levy ordinances, and appropriation measures. It approves the purpose and amount of any debt obligation. It also serves as an adviser on the structure and terms of debt obligations, on

methods to increase revenues, and on ways to reduce costs.

The use of a commission or board forces fiscal discipline on the municipality and typically represents state control over financial performance. A commission or board, by having state representation, provides leverage in remedying the problem and holds the local officials accountable for restoring fiscal health.^[7] To accomplish its work, a commission needs such extensive oversight. Municipalities in financial emergency must comply with the commissions' requirements. Municipalities under financial emergency must also comply with certain legal requirements. In Ohio, as noted above, the municipality must present its financial plan to the commission within 120 days of its first meeting. The plan must include the municipality's proposals for eliminating all fiscal emergency conditions and deficits, paying all overdue bills, restoring fund moneys, balancing the budget, and regaining the ability to market long-term general obligation bonds. Also included is an action plan for carrying out the plan and preventing a new fiscal emergency from arising in the future. Failure to submit a timely financial plan results in an automatic spending restriction in the general fund. The spending ceiling is 85% of general fund expenditures for that month in the preceding year. Fiscal discipline imposed by commissions usually results in managing within existing resources and not increasing the tax rate.

The Oversight Commission assists the municipality in developing sound financial practices to prevent future financial problems. The change in practices requires an improvement in the local government's internal capacity to identify and manage fiscal problems before they escalate. The local government must have a decision-support system that provides adequate and timely data for management decision making. It must also have accountability at each level of government.

ADDRESSING FISCAL EMERGENCIES

Local governments respond to fiscal emergencies by using several different strategies. First, they immediately tighten fiscal controls. Many use short-term strategies such as freezing/cutting the workforce and non-workforce expenditure reductions. Long-term strategies, the key to sound financial management, include putting more clarity and order into the local government's fiscal process and initiating an economic development program. This means 1) revising the budget formulation process; 2) controlling expenditures; 3) ensuring that an economic development program is in place and is maximized; and 4) initiating productivity and efficiency programs.

The local government response to the fiscal emergency should have a positive impact on its fiscal health. In New York, like Ohio, a municipality in financial emergency must prepare a multiyear financial plan for approval by the commission. These multiyear plans allow local decision makers to decide service allocations, but within revenue estimates and spending plans monitored by the commission.

Termination of the State Oversight Commission

To end state supervision, Ohio municipalities, for example, must comply with the following:

- An effective financial accounting and reporting system is being implemented with expected completion within 2 years.
- All fiscal emergency conditions have been or are in the process of being eliminated, and no new emergency conditions have occurred.
- The financial recovery plan objectives are being met.
- The entity has a 5-year financial forecast that the State Auditor determines is "non-adverse."^[3]

Other Fiscal Emergency Oversight

In Florida, the tests for meeting a municipal fiscal emergency are similar to those in Ohio, but the governor has the discretion to design a response to the local problem. The Florida Statute allows the Florida governor's office to informally consult with officials regarding the appropriate steps necessary to rectify the problem.^[8] New Jersey's laws are similar to Ohio and Florida, but in New Jersey discretion resides in a board rather than the governor.^[7]

New York State operates on a case-by-case basis, and a state law dictates the response to each particular fiscal emergency. In Pennsylvania, the head of the Department of Community Affairs appoints a department employee or a consultant as the coordinator of the municipality in fiscal emergency. The Pennsylvania Municipalities Financial Recovery Act (Act 47) governs fiscal emergencies. The act provides a range of grants local governments in fiscal emergency, including:

1. *Shared Municipal Services*: Grant funds are provided to promote cooperation between neighboring municipalities enabling various functions to be provided more efficiently and effectively.
2. *Local Government Capital Projects Loan Program*. Provides low-interest loans (2%) to municipalities (under 12,000 in population) to assist with the purchase of equipment and facilities.

3. *Land Use Planning and Technical Assistance Program*: Provides grants to municipalities, counties, and other units of local government for multimunicipal planning efforts that further the sound land-use initiatives.
4. *Regional Police Assistance Grant Program*: Provides grants for the start up of consolidated police departments for a period of up to 3 years.
5. *Floodplain Land Use Assistance Program*: Provides reimbursement for 50% of the eligible expenses connected with a municipality preparing, enacting, and implementing or administering floodplain management regulations.^[9]

CONCLUSION

Fiscal emergencies are a significant and complex public policy issue because such emergencies have serious negative consequences for local government, community, region, and the state. A fiscal emergency can lead to low quality of life and to reduced services. Fiscal emergencies are a state problem as well as a local problem. Each state has its own unique way of dealing with the emergency, although there are four generic fiscal strategies for eliminating fiscal emergencies. These strategies generally operate through state laws governing fiscal emergencies. As discussed above, states generally respond to fiscal emergencies by oversight through the use of Fiscal Emergency Laws and oversight boards.

The State Fiscal Emergency Laws used in Ohio, for example, in aiding municipalities' experiencing fiscal

problems appear to have a positive effect on long-term local government financial performance.^[5] What was once a financial "last resort" has become to some degree a less exceptional, more acceptable political option for assisting local governments in restoring fiscal health. Proactively preventing financial problems is the joint responsibility of states, local governments, and citizens. State legislative bodies are responsible for sound fiscal policies for the government they serve. Oversight bodies such as fiscal emergency oversight boards have responsibility for monitoring and, in some cases, establishing the fiscal policies of governmental units for which they have oversight responsibility.

REFERENCES

1. Lewis, C.W. Municipal bankruptcy and the states. *Urban Aff. Q.* **1994**, 30 (1), 3–26.
2. Lynch, T.D. *Public Budgeting in America*; Prentice Hall: Englewood Cliffs, 1995.
3. State of Ohio Website. www.st.oh.gov.
4. Pammer, W.J. *Managing Fiscal Stress in Major American Cities*; Greenwood Press: New York, 1990.
5. Beckett-Camarata, J. *Understanding Ohio Fiscal Emergencies*; 2001, Unpublished Manuscript.
6. Wilson, E.R.; Kattelus, S.C.; Hay, L.E. *Accounting for Governmental and Nonprofit Entities*; McGraw-Hill-Irwin: New York, 2001.
7. Hildreth, W.B. Financial emergencies. In *International Encyclopedia of Public Policy and Administration* Eds; Westview Press: 1998; (2) (D–K).
8. State of Florida Website. www.st.fl.a.gov.
9. Commonwealth of Pennsylvania Website. www.inventpa.com.

Financial Resource Management in Local Government

Kenneth A. Klase

*Department of Political Science, University of North Carolina at Greensboro,
Greensboro, North Carolina, U.S.A.*

INTRODUCTION

Financial resource management in local government involves a multitude of techniques and approaches focused on the financial management of resources to accomplish local government aims. Financial resource management occurs within diverse local government environments that profoundly influence it. Decision making in the local government fiscal system is shaped by the underlying local social and economic conditions, the formal authority structure of the local government, and by other stakeholders in the decision-making process. Fig. 1 shows the relevant factors involved and the interrelationships among factors in the process of decision making in local government financial resource management, including social and economic factors (community needs and resources); formal authority structure (degree of local government autonomy or home rule granted in state-local relations, especially with respect to financial management decision making); interest groups and other government agencies (political strategies developed and coalitions formed around particular issues). All these factors define the specific political and organizational environment, ultimately affecting financial decisions and outcomes.^[1]

STRATEGIC MANAGEMENT AND MANAGEMENT CONTROL

Although an environment such as that described above places constraints on the potential goals and strategies of local governments, local government financial resource management is most successful when performed within a strategic management context. Resource management provides the critical link between formulation of goals and objectives in a strategic planning process and the actual performance of local government activities. Local government is concerned about the effective allocation of public resources to carry out plans, strategies, and programs. Financial management is involved in the acquisition of resources and in the tracking of performance, resulting from the allocation of financial resources used to satisfy the demands of constituents within financial constraints. Thus, financial resource management

focuses on planning for achievement of desired results, implementation of those plans, and taking appropriate actions to control results so they conform to the plans. What has actually occurred must be determined, and then these actual outcomes must be reported through the financial reporting process.^[2-4]

These aspects of financial resource management are accomplished through the management control process. Management control occurs through a management control structure consisting of responsibility centers (e.g., program structures or service units) working toward the accomplishment of specific objectives. The management control process involves determining appropriate programming or activities to address strategic goals and strategies; budget preparation of a budget plan for implementing program objectives; operations and measurement in which resources are consumed and actual results attained; and reporting and evaluation where accounting and other information is summarized, analyzed, and reported. The information from reporting is used to help achieve local government goals by controlling current operations, evaluating operating performance, and performing program evaluation.^[5]

FINANCIAL MANAGEMENT FUNCTIONS

Basic functions common to financial management emphasize control, cost management, and planning. These functions include accounting and control mechanisms, allocation and management of existing resources through cost management, and long-term financial planning techniques for assessing the long-term needs of the organization and for acquiring the needed resources. Fig. 2 illustrates the principal financial management functions in local government and their emphases. Accounting and financial reporting emphasize control. Functions emphasizing cost management include cost allocation and cash management. Planning is emphasized by the functions of operations budgeting, capital budgeting, and debt administration. It is the individual functional components of financial management described in Fig. 2 that enable financial resource management to be accomplished and the management control process to function. These

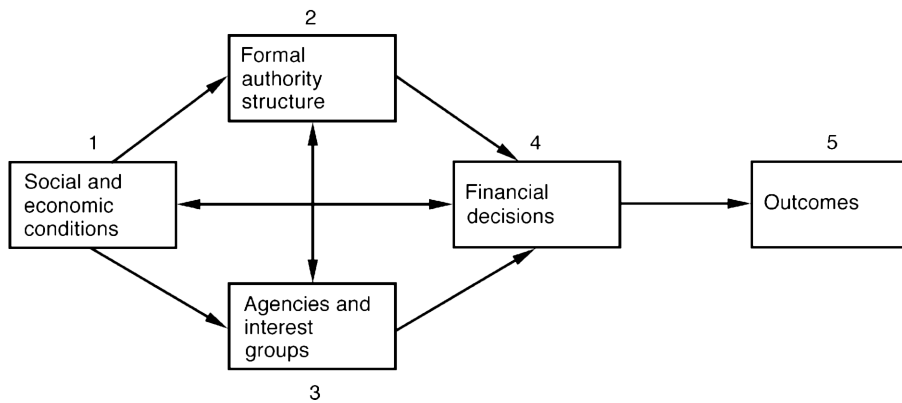


Fig. 1 A model of local government financial resource management. (Source: Ruchelman, Leonard I. "The Finance Function in Local Government." In Aronson, J. Richard and E. Schwartz, eds. *Management Policies in Local Government Finance*, 3rd edition. Washington D.C.: ICMA, 1987, p. 23.)

essential tools for managing public resources include appropriate mechanisms for accounting control and reporting of financial information, rational procedures for allocating resources and controlling costs, and techniques for assessing long-term needs and for acquiring needed resources.^[6]

Accounting and Financial Reporting

Financial accounting control systems are important components of financial resource management because of their role in record-keeping and financial control

functions. They are important not only for financial record-keeping and external reporting, but also because accounting serves as the basis for management planning, decision making, and control. Financial accounting is concerned with the results of fiscal transactions and the resulting financial position of the local government. Generally accepted accounting principles (GAAP) are the rules, guidelines, or accounting conventions used for reporting on financial transactions and creating an organization's financial statements so that they fairly represent the organization's financial position and results of operations. For state and local governments, GAAP are established by the

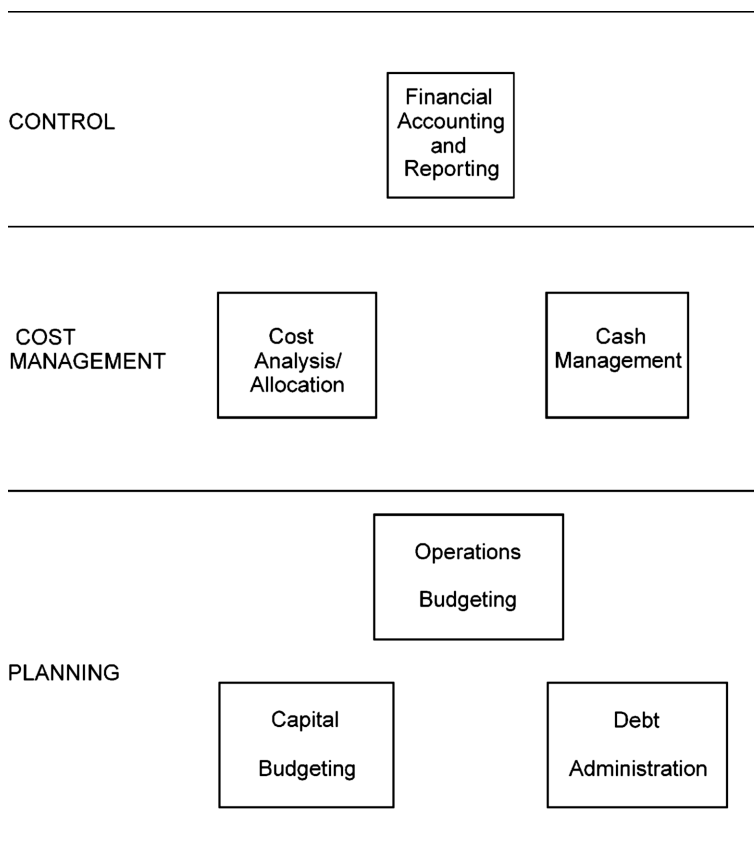


Fig. 2 Emphasis of principal financial management function in local government. (Source: Adapted from Steiss, Alan Walter. *Financial Management in Public Organization*, Brooks/Cole Publishing Company: Pacific Grove, CA, 1989, exhibit 1-1, p. 5.)

Governmental Accounting Standards Board (GASB), an authoritative body associated with the Financial Accounting Standards Board, which determines GAAP primarily for private sector organizations; both are private sector organizations that promulgate accounting standards for accounting professionals.^[7] Traditionally, GAAP has required fund accounting as the primary mechanism for control of governmental activities. Under fund accounting, resources are segregated into individual funds according to designated purposes, and reporting is performed within the fund structure. The major categories and fund types used in state and local fund accounting include the following:

- *Governmental funds* used to account for general government operations. Generic fund types classified as governmental include the *General Fund* for all resources not required to be accounted for in another fund; *Special Revenue Fund* to account for resources legally or administratively restricted for specific purposes; *Capital Projects Funds* used to account for resources restricted for major capital outlays; *Debt Service Funds* to account for resources used to repay principal and interest on general-purpose long-term debt.
- *Proprietary funds* used to account for government activities similar to commercial activities in the private sector. *Enterprise funds* are used to account for operations financed and operated similar to a business enterprise (e.g., public utilities), and *internal service funds* are used to account for operations that provide goods or services to other departments on a cost reimbursement basis (e.g., motor pool, data processing, or print shop).
- *Fiduciary funds* used to account for assets held by government in a trust or agent capacity, including *agency funds*, *pension trust funds*, *expendable and non-expendable trust funds*.

To accomplish financial accounting and reporting of fiscal transactions in local government for these fund types, what is measured and when it should be measured must be determined. The spending measurement focus emphasizes spending or the financial flow of funds by reporting only current assets and liabilities, and the difference between them or what is available to be spent in the current accounting period. In contrast, the capital maintenance measurement focus reports on all assets and liabilities (short and long term), and the difference between them or the capital or equity of the fund. Accounting principles require that a different basis of accounting be used to determine what transactions should be recognized for reporting purposes, depending on measurement focus differences between fund types. The modified accrual basis of accounting emphasizes a financial flow or spending measurement

focus, records revenues when they are measurable and available, and records expenditures when they are measurable and incurred (to be paid from current resources). Because a financial flow or spending focus applies, fixed assets are an operating expenditure and only accounted for as a general fixed asset, accounting for depreciation (representing the orderly write off of some portion of the value of fixed assets used up in producing a product or service) is optional, and long-term debt is handled as a general liability rather than a liability of a specific fund. The accrual basis of accounting emphasizes a capital maintenance or equity focus, records revenues when measurable and earned (service has been provided), and records expenses when measurable and incurred (liability created). Because a capital maintenance focus applies, fixed assets are an asset of the specific fund where purchased, depreciation is an expense for the fund-owning fixed assets, and long-term debt is a liability of the specific fund where issued. Thus, measurement focus and consequent basis of accounting depend on fund type: governmental funds and fiduciary funds with a spending or flow of funds measurement focus are accounted for under modified accrual accounting, whereas proprietary funds and fiduciary funds with a capital maintenance measurement focus are accounted for under accrual accounting.^[8]

Accounting changes for state and local government accounting and reporting under GASB Statement 34 have added to this model of fund accounting additional requirements for comprehensive entity reporting by local governments. Implementation of changes by local governments are being phased in according to the size of annual revenues for the local government, but will be fully implemented by 2003. The changes in reporting requirements are being implemented to enhance the understandability and usefulness of financial reports to various users who often find fund accounting complex and unintelligible. The new requirements focus on moving from reporting separately for different types of funds to reporting for the entity as a whole through two governmentwide statements (Statement of Net Assets, essentially an accrual accounting Balance Sheet and Statement of Activities, essentially an accrual accounting Operating Statement) as part of the basic financial statements for the government. In addition, governments are also required to continue fund financial statements. Perhaps most useful to users of the basic financial statements, local governments are required under the GASB 34 changes to provide a management's discussion and analysis section and other required supplementary information, including a comparison of actual results with authorized budgets. These interpret the financial statement information in a more understandable manner for users.^[2,9-11]

Cost Allocation and Cash Management

The cost analysis/allocation and cash management functions in financial resource management emphasize cost management, as shown in Fig. 2. Cost management depends on information available from accounting control systems. Local government accounting systems do not generally generate good cost data, given the financial flow measurement focus of governmental fund types. Consequently, local governments engage in cost analysis/allocation. Analysis of costs is used to forecast short- and long-term needs, and provides input to cash management decision making and long-term financial planning. Cost accounting approaches attempt to determine the full costs of local government service delivery and to assign accountability for costs to the appropriate local government responsibility center. Through cost analysis, costs can be classified as direct (costs directly associated with a specific purpose) or indirect (costs associated with more than one activity and not able to be traced directly to any one of the activities). Cost allocation involves allocating indirect costs or overhead using some basis of allocation (e.g., number of square feet that a cost center occupies) to specific cost centers to arrive at a more accurate measurement of the full costs (direct plus allocated indirect) for those cost centers. In addition, determination of fixed and variable costs can enable differential cost analysis for alternative decision making.^[6,12]

Cash management also emphasizes cost management and is concerned with short-term cash flow management, cash mobilization, and investment. Short-term cash flow management involves ensuring that adequate cash resources are available in a timely manner when needed for current operations to avoid cash shortages. Many organizations prepare a cash budget, usually indicating month by month how much cash is forecasted to be received or paid. The most difficult problem is accurately estimating receipts. Forecasted deficits in cashflow require borrowing, while surpluses allow for short-term investments to be made. Cash management also involves cash mobilization as a technique to assemble funds and make them more readily available for investment. Cash mobilization is accomplished through accelerating receivables (reduction of time delays in collecting receivables or deposit float) and controlling disbursements (especially delaying cash outflows through disbursement procedures and consolidated concentration accounts). Cash is mobilized to maximize investment return on temporary cash balances. The principal factors considered in selecting appropriate investment options for public funds are safety/risk, liquidity/marketability, maturity, and yield. Appropriate investments for local governments are often confined to certificates of deposit, U.S. Treasury bills, other federal agency securities, and state

investment pools. Interest earned on investments is an important benefit of good cash management, but safety and liquidity of cash assets have to be balanced against possible yield.^[6,13,14]

Operations Budgeting

The operations budgeting component of financial resource management emphasizes planning as depicted in Fig. 2, but is traditionally viewed as a control mechanism to ensure legal compliance and accountability through appropriation to specific objects of expenditure linked to budgetary accounting. The most common operating budget format in local government, the line item budget, appropriates funds to specific line items in this fashion and is oriented toward controlling expenditures in accordance with fiduciary accountability. Significant management and planning capacity can be incorporated into the operations budgeting process where performance and program budgeting formats and approaches are used. The performance budget is organized around tasks, activities, or direct output performed. It emphasizes expenditure by workload or activity and unit costs by activity. Performance budgets have a management orientation because they emphasize efficiency in performance, maximizing unit output per unit input of resources. The program budgeting format emphasizes effectiveness by relating expenditures to public goals through its organization by programs. It focuses on achievements, outcomes, or final products as a result of its planning orientation and links results to strategic planning goals. Financial resource management in local government can be enhanced by incorporating elements of these approaches where they help to focus on efficiency in the utilization of resources and effectiveness in accomplishing desired goals and objectives. Thus, the operations budget can become a planning tool that helps local government to achieve strategic planning goals.^[15-19]

Capital Budgeting and Debt Administration

Capital budgeting in local government emphasizes long-term capital facilities planning and programming. It differs from operations budgeting in focusing on long-term, non-recurring projects that are of high value and are financed from other than operating budget revenues, usually debt financing. The capital budgeting process depends on assessment and forecasting of community needs, evaluation and ranking of capital projects to address those needs over a period of time, and determination of the financial capacity of the community to pay for capital improvements. It results in a capital improvement program (CIP) for a 5- or 6-year period by scheduling capital projects in priority order over the period.

The first year of the 5- or 6-year CIP becomes the next year's capital budget.^[6,15,20]

Debt administration associated with capital financing requires establishing appropriate debt policy (in particular, no issuance of debt with a maturity longer than the useful life of the project). It also involves determining the structure and design of debt to be issued: type of security—general obligation bond debt backed by full faith and credit vs. limited liability revenue bond debt backed by revenue to be generated from the project; term to maturity of the bond—either term or serial with multiple maturities; and call provisions before maturity. The borrowing costs of bond issuance is affected by a local government's bond rating, an assessment of creditworthiness or risk to potential purchasers for a particular bond issue as determined by commercial rating firms. Because the marketing and long-term administration of bond debt is problematic for local governments, bonds are usually sold to an underwriting firm that purchases the entire issue and resells it at a profit to investors.^[6,15]

CONCLUSION

Financial resource management in local government is constrained by its environment, but nonetheless is a means for accomplishing strategic management. Through management control, the component functions of financial management enable goals and objectives from strategic planning to be linked to actual results. The component functions of financial management accomplish these ends through their emphasis on financial management's core values: control, performance (efficiency), and planning (effectiveness).

As a dynamic process, financial management in local government is continually changing. This is evident in the dramatic changes in accounting and financial reporting standards that will have significant impacts on local governments. It is evident in the increased emphasis on cost analysis to use resources more efficiently. It is also evident in the increased emphasis in operations budgeting on performance and program approaches to incorporate strategic planning to achieve planned results, and thus to enhance local government effectiveness and accountability. All these changes are further evidence of the continuing influence of the core values that component functions of financial management emphasize.

REFERENCES

1. Ruchelman, L.I. The finance function in local government. In *Management Policies in Local Government Finance*, 3rd Ed.; Aronson, J.R., Schwartz, E., Eds.; ICMA: Washington, DC, 1987; 23.
2. Finkler, S.A. *Financial Management for Public, Health, and Not-for-Profit Organizations*; Prentice Hall: Upper Saddle River, NJ, 2001; pp. 21, 370–371.
3. Bryant, S. Strategic management: developing and realizing a strategic vision. *Public Manage.* October 1997, 79, 28–32.
4. Berry, F.S. Innovation in public management: adoption of strategic planning. *Public Adm. Rev.* July/August 1994, 54, 322–330.
5. Anthony, R.N.; Young, D.W. *Management Control in Non-profit Organizations*, 6th Ed.; Irwin/McGraw-Hill: Boston, MA, 1999; 3–23.
6. Steiss, A.W. *Financial Management in Public Organizations*; Brooks/Cole Publishing Company: Pacific Grove, CA, 1989; pp. 1–4, 8–25, 53–120, 169–224.
7. Antonio, J.F. Role and future of the governmental accounting standards board. *Public Budg. Finance Summer* 1985, 5 (2), 30.
8. Glick, P.E. *How to Understand Local Government Financial Statements: A User's Guide*; GFOA: Chicago, IL, 1986.
9. Governmental Accounting Standards Board (GASB), Statement no. 34 of the governmental accounting standards board. *Basic Financial Statements—and Management's Discussion and Analysis—for State and Local Governments*; Financial Accounting Foundation: Norwalk, CT, 1999.
10. Crain, G.W.; Bean, D.R. What users think of the new reporting model for governments: the results of the GASB's focus group sessions. *Gov. Finance Rev.* February 1998, 14, 9–12.
11. Gauthier, S. GASB reporting model exposure draft. *Gov. Finance Rev.* June 1997, 13, 60–61.
12. Brown, R.E.; Sprohge, H.-D. Government managerial accounting: what and where is it? *Public Budg. Finance August* 1987, 7 (3), 35.
13. Mattson, K.; Hackbart, M.; Ramsey, J.R. State and corporate cash management: a comparison. *Public Budg. Finance Winter* 1990, 10 (4), 18–27.
14. Petersen, J.E. Managing and investing public money. *Governing March* 1993, 6, 43–54.
15. Mikesell, J. *Fiscal Administration: Analysis and Applications for the Public Sector*, 5th Ed.; Harcourt Brace & Company: Fort Worth, TX, 1999; pp. 181, 223–261, 533–567.
16. Allen, J.R. The uses of performance measurement in government. *Gov. Finance Rev.* August 1996, 12, 11–15.
17. Clifford, C. Linking strategic planning and budgeting in Scottsdale, Arizona. *Gov. Finance Rev.* August 1998, 14 (4), 9–14.
18. Gross, J. NACSLB recommended budget practices: what they are and how to use them. *Gov. Finance Rev.* 1998, 14 (3), 9.
19. Thompson, F. Mission-driven, results-oriented budgeting: fiscal administration and the new public management. *Public Budg. Finance Fall* 1994, 14 (3), 90.
20. Forrester, J. Municipal capital budgeting: an examination. *Public Budg. Finance Summer* 1993, 13 (2), 85–101.

Financial Statement Analysis

Martin Ives

*R.F. Wagner Graduate School of Public Service, New York University,
New York, New York, U.S.A.*

INTRODUCTION

General-purpose governments, governmental enterprises, and not-for-profit entities meet their public accountability requirements partly by preparing annual financial statements. Financial statements communicate the entity's financial position at the end of the year, for which the statements and the entity's operating results for the year then ended are prepared. The statements array the numerous day-to-day economic transactions and events in such a way that readers can make informed decisions about the entity and assess its accountability.

Financial statement analysis requires use of the statements, related notes, and auditor's opinion to develop information about the entity's liquidity and short-term financial position, operating results, and longer-term financial flexibility. For general-purpose governments, financial statement analysis also assists in assessing budgetary accountability. Financial statement analysis is also a part of the broader assessment of an entity's financial condition; that is, the likelihood that the entity will meet its future financial obligations to bondholders, employees, and stockholders as well as its service obligations to the citizenry.

Financial statement analysis requires conversion of financial data to more useful forms, primarily ratios. Effective analysis also requires using data for more than 1 yr as well as comparison with other entities (as a frame of reference for reaching conclusions about the entity studied). Many of the concepts and ratios discussed in this paper apply equally to general-purpose governments, governmental enterprises, and not-for-profit entities. Those that are unique to a specific type of entity are identified as such.

USERS AND USES OF FINANCIAL STATEMENT ANALYSIS

Persons both internal and external to a particular entity may analyze financial statements. (When prescribing financial reporting standards, accounting standards setters generally consider the needs of external users, because internal users have ready access to whatever information they need.) Users of financial statement analysis include a governmental entity's elected

and appointed officials; officials of oversight governments; managers and trustees of not-for-profit entities; donors and potential donors; citizens and citizen advocate groups; providers of goods and services, including employee unions; and financial resource providers and their representatives (e.g., banks, municipal credit-rating agencies, and insurers of municipal debt).

Financial statement analysis leads the user not only to assess financial position, results of operations, and accountability, but also to act on the information. For example, the analysis may lead managers or trustees to conclude that the data indicate a deteriorating financial position, causing them to act to shore up the entity's finances. These are some specific concerns an analyst might have when analyzing financial statements:

- To determine the likelihood of the entity's ability to pay its short-term obligations, such as tax anticipation notes and amounts due to vendors and employees.
- To determine the likelihood of the entity's ability to pay its long-term obligations, such as long-term bonds and pensions.
- To see if the entity's margin of revenues over expenses provides a sufficient cushion to weather economic downturns, finance certain expenditures, or meet contingencies.
- To ascertain the composition of an entity's revenues and expenses (e.g., the portion of revenues obtained from sales taxes or the portion of expenses used for program, rather than administrative, purposes).
- To determine what portion of an entity's resources may be used for general or unrestricted purposes, rather than being restricted for specific purposes by law or donor requirement.

INFORMATION CONTENT OF FINANCIAL STATEMENTS

Accounting standards setters have changed financial reporting requirements during the past few years. For example, new financial reporting standards require state and local governments to prepare a management discussion and analysis (MD&A) to precede the financial statements. The MD&A contains condensed

financial information that compares the current year with the previous year. The accompanying analysis of the government's financial position and results of operations helps financial statement users assess whether the entity's financial position improved or deteriorated as a result of the year's operations. The analysis also includes a discussion of significant variations between budgeted and actual amounts, as well as currently known reasons for variations that are expected to affect future services or liquidity. Important economic factors that significantly affected operating results for the year must also be discussed. Therefore, the MD&A provides a good starting point for analyzing a governmental entity's financial statements.

Regardless of the type of entity analyzed, the financial statements, the accompanying notes, and the auditor's opinion need to be read thoroughly. The formats and titles of financial statements prepared by governmental and not-for-profit entities (including not-for-profit hospitals) are somewhat different, but the kinds of data they convey are similar. This section contains a brief description of the contents of the various types of financial statements and the related notes.

Governmental and not-for-profit entities generally use fund accounting. Each fund is a separate fiscal and accounting entity, established for the purpose of carrying on specific activities in accordance with laws, regulations, and restrictions. Despite the use of funds for internal purposes, both governmental and not-for-profit financial reporting standards emphasize the entity as a whole, rather than individual funds or fund types. Government financial reporting standards require two levels of reporting, a government-wide level (which distinguishes between governmental and business-type activities) and a fund level (where the focus is on major funds). Not-for-profit entity reporting standards require reporting on the entity as a whole but permit entities also to provide data by fund type in external financial reports, if they want to do so. Within their entity-wide reporting, not-for-profit entities must report year-to-year changes in net assets by three net asset classes—unrestricted, temporarily restricted by donors, and permanently restricted by donors.

The most common financial statements prepared by governmental and not-for-profit entities fall into these general categories:

- Statement of financial position (or statement of net assets or balance sheet).
- Statement of operations (or statement of activities or statement of revenues, expenditures/expenses, and changes in fund balances).
- Cash flow statement.

In addition, depending on the circumstances, some not-for-profit entities may prepare a statement of

functional expenses and a statement of changes in net assets.

Statement of Financial Position

A statement of financial position provides information about an entity's assets, liabilities, and net assets at a point in time. When considered together with the notes to the financial statements, the statement of financial position helps the analyst assess liquidity and financial flexibility. Assets and liabilities appearing in the statement of financial position are arrayed generally in terms of liquidity, with assets shown in order of nearness to being converted to cash and liabilities shown in order of nearness to being paid. Assets and liabilities may be grouped, based on whether they are current or non-current. In that context, the term "current" generally refers to the expectation that the assets will be converted to cash or consumed in operations in the next year and that the liabilities will be paid off in the next year. (In governmental fund balance sheets, however, the term "current" connotes a shorter period of time.)

The net asset section of the statement of financial position—that is, the difference between assets and liabilities—is segregated as between unrestricted and restricted. (Use of resources in a governmental entity may be restricted by law or by creditors; use of resources in a not-for-profit entity may be restricted by donors.) Resources classified as unrestricted may be used for any purpose consistent with the entity's functions.

Statement of Operations

The statement of operations, when used with the related notes and other data, provides information to help users evaluate the entity's financial performance during a specified time period. It shows the composition of the various types of revenues and expenses and how revenues were used in providing programs or services. There are some differences in format and emphasis among financial statements prepared for governmental entities, not-for-profit entities in general, and not-for-profit hospitals.

Government-wide operating statements, prepared on the full accrual basis of accounting, show the relative financial burden each governmental function places on its taxpayers. The statements do this by reporting the expense of each function net of revenues (e.g., charges for services and operating grants from other governments) directly related to the function. Taxes, other general revenues, special items, and extraordinary items are then deducted from the sum of the net expenses of all the functions to show the change in

net assets for the year. (Special items, sometimes called “one-shots,” are transactions within the control of government that either are unusual in nature or do not occur frequently.) The change in net assets provides a measure of the extent to which the year’s revenues financed the services provided by the government.

Governments must also prepare operating statements for each major fund. There is likely to be more than one set of fund-level statements, because those statements are prepared on the modified accrual basis of accounting for governmental funds and the full accrual basis for proprietary funds. After considering the revenues, expenditures, other financing sources, and special items in the governmental fund statements, the result is the net change in fund balances. By definition, governmental fund accounting focuses on short-term revenues and expenditures; also, other financing sources may contain non-revenue inflows. Therefore, the analyst must use these statements with care.

The statement of activities prepared by not-for-profit entities shows the changes in net assets by net asset class (unrestricted, temporarily restricted, and permanently restricted) and in total. This statement shows major revenue sources (e.g., contributions, fees, and investment income) for each net asset class. Because reporting standards require that all expenses be reported in the unrestricted class of net assets, the amount of net assets released from restriction for use in operations is also shown. To help analysts to assess service efforts, not-for-profit entities report expenses by functional classification, such as major classes of program services and supporting activities. Not-for-profit entities classified as voluntary health and welfare organizations must report expenses both by major function and by natural classification (e.g., salary, rent, and utility expenses).

Not-for-profit hospitals follow the same reporting standards as other not-for-profit entities. Those concerned with hospital financial reporting, however, believe that it is important for hospitals to have a clear bottom line measuring financial performance. They, therefore, recommend using two activity statements, a statement of operations and a statement of changes in net assets. In the statement of operations, the caption “excess of revenues over expenses” provides the analyst with a performance indicator.

Cash Flow Statement

An entity could report an excess of revenues over expenses in the statement of operations because it uses the full accrual basis of accounting, yet experience financial difficulties because of a shortage of cash. The cash flow statement shows the factors that result

in cash inflows and outflows and provides a useful link to the other financial statements.

Notes to Financial Statements

Because the statements themselves do not provide a complete picture of financial position and results of operations, they are accompanied with notes. The notes are considered necessary for fair presentation and are an integral part of the financial statements. They provide details on such matters as significant accounting policies, contributions, restrictions, pension plans, contingencies, and financing arrangements. For example, note disclosures on the rate of amortization of long-term debt help in assessing an entity’s debt burden, and note disclosures on pensions help in assessing pension-funding status.

FINANCIAL STATEMENT ANALYSIS METHODOLOGY

Financial statement analysis requires use of time-series analysis and comparative analysis. It also requires conversion of financial and other data into forms that will facilitate time-series and comparative analyses, such as ratios, percentage change information, and common size analysis. There are limits, however, to what one can do with the individual bits of information derived from each aspect of the analysis. A few ratios may suffice for some purposes, for example, if the analyst wants to know only the present composition of the entity’s expenses. If the analyst wants to draw conclusions about the long-term financial viability of a governmental entity, however, many matters (including data outside the financial statements) must be examined. Financial statement analysis is an art, not a precise science.

Time Series Analysis and Comparative Analysis

Time-series analysis looks at changes within an entity over time to see if conditions have improved or deteriorated. Time-series analysis requires a minimum of 2 yr. In assessing a state or local government’s financial condition, time-series analysis covers a period of at least 5 yr. Time-series analysis allows one to draw conclusions about changes in an entity’s financial position and results of operations, but it suffers from the lack of an external reference point that allows one to assess the implications of the level of a particular statistic.

Comparative analysis provides that external reference point. Comparative analysis in this context means comparing an entity with a peer group of similar entities, credit-rating agency norms or medians, or

conclusions reached in research studies. Care must be taken in developing a reference group for comparison purposes, because data may be distorted by such factors as differences in mission, size, location, and mix of population served. Thus, a peer group might include hospitals of generally similar bed size that perform similar functions and, also, governments of roughly similar population that perform similar functions in the same state. For example, an analyst might compare financial data of a village with, say, 10 villages of roughly similar population in the same state.

Although having a reference point allows the analyst to draw inferences about an entity, using comparative data also has some limitations that need to be recognized. For example, if a peer group of similar entities suffers from the same problems experienced by the entity under study, the analyst may fail to recognize a particular weakness.

Conversion of Data to More Useful Forms

To make time-series and comparative data useful, financial statement data elements must be converted to more useful forms, such as ratios. Ratios are obtained by relating one financial statement data element to another or to an economic or demographic data element. By converting financial and other data to a ratio, you can gain insights not previously evident. Further, ratios make it possible to track data over time and to compare it with other entities. For example, when the numbers are considered separately, you cannot draw inferences from the fact that the financial statements show accounts receivable to be \$90,000 and revenues to be \$405,000. But, when those numbers are considered together and converted to ratios, you find that it takes an average of 81 days to collect the receivables. Comparing that number with last year's number of 75 days and an industry norm of 60 days allows the analyst to reach conclusions and raise appropriate questions.

A simple form of data conversion is calculating the percentage change from an earlier period to the current period. Another form of data conversion, the location quotient, is obtained by using the number for the entity under study as the numerator and the reference group number as the denominator. When used in time-series analysis, the location quotient readily reveals the direction of the data element for the entity under study relative to the reference group. Converting certain financial data to per capita information (e.g., to compute property tax revenues or education expenditures per resident) also facilitates comparison with a peer group.

Converting financial statements to "common size" facilitates financial statement analysis, because it results in conversion of numbers to percentages of

the whole. For example, if the total expenses of a not-for-profit entity are calculated as 100%, the amount reported for administrative and fund-raising expenses might be, say, 28% of the total. By itself, that percentage might not be meaningful. However, if that percentage was 24% last year and if similar not-for-profit entities report an average of 20%, the percentage change from last year and the location quotient relative to the peer group might cause the analyst to raise questions about the reasons for the disparities.

Some of the more common ratios developed in financial statement analysis are discussed in the ensuing paragraphs.

LIQUIDITY AND SHORT-TERM FINANCIAL POSITION RATIOS

Liquidity ratios (sometimes called short-term financial position or cash solvency ratios) provide information on an entity's ability to meet its short-term obligations, such as salaries, bills for supplies received, and payments on borrowings coming due in the near future. Paying short-term obligations in a timely manner requires the current availability of cash or the ability to convert other assets to cash in time to pay the bills.

The current and the quick ratios are the most commonly used liquidity ratios. The current ratio compares all current assets with current liabilities. The quick ratio compares cash and cash equivalents (and, sometimes, other current assets) with current liabilities.

Cash equivalents are included in the quick ratio because they are short-term, highly liquid investments, readily convertible to known amounts of cash, and are so close to maturity (having had an original maturity of 3 mo or less) that there is little risk of loss in value. Analysts may add other short-term, highly liquid investments and receivables when computing the quick ratio. General-purpose government analysts often do not do that, however, because current liabilities are generally due to be paid in 30 days, whereas investments and receivables may not be readily convertible to cash in time to pay bills. Some analysts use a quick ratio of 1.0 as a norm, but a ratio of less than 1.0 does not necessarily mean that the entity is illiquid.

$$\text{Current ratio} = \frac{\text{Current assets}}{\text{Current liabilities}}$$

$$\text{Quick ratio} = \frac{\text{Cash} + \text{Cash equivalents}^*}{\text{Current liabilities}}$$

*Sometimes, other current assets are added.

Another perspective on cash liquidity can be obtained by calculating the number of days' cash on hand. This ratio tells the analyst how long the entity can continue to pay its operating expenses without further cash inflows. It may also reveal that the entity has excessive cash, which might be better used to acquire new facilities and equipment.

Days' cash on hand:

$$\frac{\text{Cash} + \text{Cash equivalents} + \text{Short-term investments}}{\text{Cash needs per day}^*}$$

*Total operating expenses less bad debts and depreciation, divided by 365.

The speed with which receivables are converted to cash is a significant indicator of both liquidity and efficiency. The quicker the entity converts receivables to cash, the less cash and other assets it needs for paying current obligations and the quicker the excess cash can be converted to interest-earning assets. Analysts of general-purpose governments calculate a ratio of real property taxes receivable to tax revenues and a tax delinquency rate (or its inverse). Analysts of hospitals and some governmental enterprises often express receivables liquidity in the form of number of days' revenue in receivables.

Property tax receivables rate:

$$\frac{\text{Real property taxes receivable}}{\text{Real property tax revenues}}$$

Tax collection delinquency:^a

$$\frac{\text{Uncollected current year real property taxes}}{\text{Current year real property tax levy}}$$

Days' revenue in receivables (hospitals):

$$\frac{\text{Net patient accounts receivable}}{\text{Net patient service revenues}/365}$$

DEBT BURDEN AND LONG-TERM FINANCIAL FLEXIBILITY RATIOS

Debt burden and long-term financial flexibility ratios measure an entity's ability to meet its long-term financial and service obligations. These ratios help answer questions such as: Is the entity likely to have sufficient resources to pay its long-term debts, including bonded debt, pension obligations, and postemployment health-care obligations? Does it have the capacity

to issue new debt to meet its current and future capital needs?

General-Purpose Governments

The most common debt burden ratios for general-purpose governments relate the outstanding long-term debt to a base that measures the entity's capacity to pay the debt, such as population, full value of taxable real property, and personal income. When making these calculations, the amount used for long-term debt should include all of the government's tax-supported debt, including general obligation debt, lease-purchase debt, and debt incurred through specially created agencies. Analysts generally exclude self-supporting debt from the total debt. They also make a separate calculation of debt burden that includes a share of the "overlapping" debt from other governments that tax its citizens from the same tax base.

Debt burden:

$$\frac{\text{Long-term debt}}{\text{Population, or full value of taxable real property, or personal income}}$$

Closely related to the debt burden is the debt service burden—the portion of the total revenues required to pay annual debt service (interest and principal) on outstanding debt. Obviously, the greater the debt service burden, the less the amount of resources that is available for program services. Norms regarding debt service burden differ by types of government and the functions they perform. As a general rule, a debt service burden below 10% is considered "affordable" (in terms of competition among programs for resources) and a burden between 15% and 20% is considered "high." When assessing debt service burdens, you need to consider the rate of pay-down of debt principal, because an entity that amortizes its debt principal rapidly is likely to experience a higher debt service burden ratio than one that amortizes its debt principal slowly. Credit-rating agencies suggest that, on average, 25% of outstanding debt should be retired within 5 yr of the balance sheet date and 50% should be retired within 10 yr.

$$\text{Debt service burden} = \frac{\text{Total debt service}}{\text{Total revenues}}$$

Governmental Enterprises and Not-for-Profit Entities

Financial statement analysts of governmental enterprises and not-for-profit entities (including hospitals)

^aIn the inverse, the numerator is the amount of real property taxes collected in the year of the levy.

generally calculate leverage ratios, which relate debt to equity. The concept here is that the more leveraged the entity, the more difficult it will be to borrow to finance future capital needs. There are several ways to calculate the debt to equity ratio. The numerator is generally the total long-term debt, including the amount reported as current (due to be paid in the coming year.) The denominator is generally either the total assets or the net assets (total assets minus liabilities).

$$\text{Debt to equity ratio} = \frac{\text{Total long-term debt}}{\text{Total assets (or net assets)}}$$

Analysts focus on debt service coverage, that is, the number of times the total debt service is covered by the revenues or the number of times the interest on the debt is covered by the revenues. The greater the “coverage,” the more likely it is that the entity will be able to meet its debt service requirements.

Debt service coverage:

$$\frac{(\text{Revenues} - \text{Expenses}) + \text{Depreciation} + \text{Interest}}{\text{Principal payment} + \text{Interest expense}}$$

Times interest earned:

$$\frac{(\text{Revenues} - \text{Expenses}) + \text{Interest expense}}{\text{Interest expense}}$$

Pensions

If pensions are a significant element of an entity's expenses, the extent of the funding of the pension system often provides evidence of an entity's financial health. The most common measure of pension funding adequacy is the funded status—the ratio of pension assets to the pension obligation. This is a required disclosure in government financial reporting. If the funded status is not available, pension payout coverage (ratio of pension assets to pension benefit expenditures) provides a useful measure.

Funded ratio:

$$\frac{\text{Pension fund assets available for benefits}}{\text{Pension benefit obligation}}$$

Pension payout coverage:

$$\frac{\text{Pension fund assets available for benefits}}{\text{Pension benefit paid last year}}$$

OPERATING RESULTS RATIOS

Ratios that depict operating results are sometimes called budgetary solvency ratios or profitability ratios. The most common operating results ratio is similar to the return-on-sales ratio found in business enterprise. Although neither governments nor not-for-profit entities operate for the purpose of making profit, breaking even or generating at least some surplus for the year is considered a positive indicator.

Depending on the format of the operating statement, the names of the numerator and the denominator may be somewhat different for the various types of entities. For example, the numerator may be the excess of revenues over expenses, the increase in net assets, or the increase in unrestricted net assets. The denominator may be the total revenues and other financing sources or the total revenues, gains, and other support. In any event, the governmental analyst, in particular, must take care to eliminate (or separately consider) “one-shot” resource inflows from borrowings, unusual financing arrangements, or other special items, so as not to distort the operating results for a particular year.

Operating surplus or operating margin:

$$\frac{\text{Excess of revenues over expenses (or expenditures)}}{\text{Total revenues, gains, and other support}^*}$$

*Or total revenues and other financing sources.

Those who analyze state and local government financial statements are particularly concerned with cumulative results of operations and the resulting fund balance. This is because fund balance provides a cushion against future budgetary problems caused by economic contraction or emergency expenditure needs. A general rule is that unreserved fund balance of 5–10% of revenues is sufficient to cover most contingencies. The budget cushion of entities whose tax revenues tend to be economy sensitive (such as income and sales taxes) should be at the higher end of the range.

$$\text{Budget cushion} = \frac{\text{Total unreserved fund balance}}{\text{Total revenues}}$$

In addition to the overall operating results, those who analyze governmental financial statements focus on trends in the composition of revenues and expenditures. This is because increases in tax rates for a particular revenue source may weaken the ability of the entity to raise rates in the future. Increases in certain expenditures may indicate particular social and economic problems that need to be addressed.

Those who analyze not-for-profit entity statements may be interested in the portion of the entity's resources that is spent on program activities vs. non-program

activities, such as administrative and fund-raising expenses.

$$\text{Program service ratio} = \frac{\text{Program expenses}}{\text{Total expenses}}$$

ASSESSING GOVERNMENTAL FINANCIAL CONDITION

Financial statement analysis provides insight into an entity's current financial position and results of operations. For a longer-term perspective on an entity's financial health, however, the analyst needs to look beyond the financial statements. Going beyond the financial statements takes the analyst into what some call financial condition analysis.

The ability of a governmental entity to finance future expenditures arising out of previous commitments depends heavily on the strength of its economy. It, therefore, requires the analyst to consider not only the analysis of financial statements but also the trends in economic and demographic data. For example, the financial condition analyst is concerned with the growth in property, sales, and income tax bases, caused by such factors as new residential and commercial construction and the employment base. The financial

condition analyst is also concerned with population size and composition, personal income, poverty rates, full value of taxable real property, unemployment rates, employment totals and composition by industry sector, and employer diversity.

Much of the economic and demographic data needed for financial condition analysis may be found at websites maintained by the U.S. Bureau of the Census, the U.S. Bureau of Economic Analysis, and the U.S. Bureau of Labor Statistics. Some of the data may be found in the statistical sections of governmental comprehensive annual financial reports.

BIBLIOGRAPHY

1. Berne, R.; Schramm, R. *The Financial Analysis of Governments*; Prentice-Hall: Englewood Cliffs, NJ, Chapters 6, 8, 1986.
2. Finkler, S. *Financial Management for Public, Health, and Not-for-Profit Organizations*; Prentice-Hall: Upper Saddle River, NJ, Chapter 12, 2001.
3. Groves, S.; Valente, M. *Evaluating Financial Condition—A Handbook for Local Government*; ICMA: Washington, DC, Part II, 1994.
4. Ives, M. *Assessing Municipal Financial Condition*; Ives & Hancox: Croton-on-Hudson, NY, Chapters 5, 8, 9, 11, 2006.

Fiscal Distress of Local Governments in Israel

Abraham Carmeli

Graduate School of Business Administration, Bar-Ilan University,
Ramat-Gan, Israel

INTRODUCTION

Fiscal and financial distress of local governments is a worldwide phenomenon. This phenomenon has been a key subject of inquiry, one that has attracted great interest to all local government stakeholders (academic researchers, practitioners, state and local government organizations, suppliers, media, and the public, among others) due to its significant implications on the political, social, and economic state of affairs. The present entry provides an overview of this phenomenon, with a particular emphasis on the fiscal distress of local governments in Israel.

FISCAL DISTRESS OF LOCAL GOVERNMENTS

Fiscal distress of local governments is a major public problem that affects the economic, social, and political well-being of both individuals and organizations. Facing localities' fiscal distress is of great concern, because it may reach the point of overall system collapse. This phenomenon occurs in the more liberal economically oriented countries (e.g., the United States) as well as the more welfare-oriented countries (e.g., Norway).

Over the years, many local governments have experienced fundamental fiscal difficulties: New York City in 1970; Philadelphia in 1990; Orange County in 1994; Bridgeport, which became the largest city at the time to file for bankruptcy; Miami in the 1990s; and Glasgow in the second half of the 1990s.^[1] This has also been the case with local governments in Britain, the Netherlands in the 1980s, Germany in the 1990s, and Japan in the second half of the 1990s. Many of the local governments in Israel have also experienced and, in some cases, are still experiencing fiscal distress. Research suggests that this is an ongoing acute problem.^[2] Various indicators, as explained below, are evident for ongoing fiscal and financial distress. The key problem is that the local government experiences considerable difficulties in raising sufficient revenues to cover its legally required expenditures, which often creates a structural public problem.

FISCAL DISTRESS AND CRISIS—DEFINED

Scholars and practitioners use various terms such as *fiscal distress*, *fiscal crisis*, and *financial crisis* to describe a situation in which local governments have difficulties meeting fiscal and financial needs.

"Fiscal distress" has been defined in various related ways and generally refers to both short- and long-term difficulties. Bradbury was a pioneer in suggesting a categorization by which fiscal distress can be quantified.^[3] She distinguished between *budgetary fiscal distress* and *citizen fiscal distress*. "Budgetary fiscal distress" occurs when local authorities encounter difficulties in balancing their budget, while "citizen fiscal distress" occurs when citizens are imposed upon to pay exorbitant taxes and fees for the municipal services they are provided with or when local authorities provide a low level of municipal services compared to other local authorities in similar situations.^[3]

A similar approach for defining financial conditions was suggested by Groves and Valente.^[4] These authors distinguished between *budgetary solvency* and *service-level solvency*. "Budgetary solvency" refers to the government's capacity to generate enough revenues over its normal budgetary period to meet its expenditures and not incur deficits, while "service-level solvency" refers to the government's ability to provide a community's desired services at the level and quality that are required for the health, safety, and welfare of that community.^[4] In addition, Inman refers to "fiscal crisis" as that which "occurs when a city's potential to raise revenues is insufficient to cover the city's legally required expenditures" (on page 378 of Ref.^[5]).

ASSESSING THE FISCAL DISTRESS OF ISRAEL'S LOCAL GOVERNMENTS

The Financial Structure of the Local Government

Local government activities in Israel are financed by two major sources: the regular (ongoing or current) budget and the non-regular (one-time or development) budget. The regular budget is financed by self-incomes (revenues collected directly from residents

and businesses, such as property taxes and development levies), transferred revenues (portions of taxes that are basically owned by the local government but collected by the central government and transferred to local authorities), general grants (transferred by the Ministry of the Interior in order to ensure sufficient resources, so that local authorities can deliver an acceptable municipal services basket), and specific grants (transferred by the central government to finance state-related services, particularly education and welfare).^[6] The regular budget should finance current expenses such as salaries, operational activities, and loan payments. The non-regular budget is financed by non-regular incomes, which typically are not drawn from the same sources such as development fees, improvement fees, donations, and loans (e.g., Ref.^[6]), but are ad hoc in nature. The non-regular budget should finance the more costly infrastructures such as water, transport roads, lighting, and drainage, namely developmental programs. These two sources are an indication of a local authority's fiscal and financial health. The higher the self-income ratio is to the regular budget, the stronger a local authority is fiscally. In addition, a high self-income ratio is an indication of a local authority's level of autonomy.^[7]

The fiscal state of the local government

A report of the Department of Economics affiliated with the Israel Union of Local Authorities, published on August 18, 2003, indicates that 109 LGAs experienced problems paying their employees' provident fund and continuing education fund; 128 localities still have not yet paid their taxes to the authorities; in 24 localities, monthly salaries have not been paid at all; and, in an additional 19 localities, monthly salaries were only partially paid. To learn more about the budgetary conditions of the local government, we referred to all localities in Israel, using aggregated data.

The current deficit of the entire local government was 655,217,000 and 1,278,094,000 NIS or 2.2% and 4.2% in 2000 and 2001, respectively. The accumulated deficit of the entire local government was 4,518,236,000 and 4,970,345,000 NIS or 15.5% and 16.3% in 2000 and 2001. These findings are subject to one's point of view regarding the desirable scope of deficit. Yet, these results also indicate that the local government as a whole experiences considerable difficulties in raising sufficient revenues to cover its legally required expenditures. Consistent with Inman's view (on page 378 of Ref.^[5]), Israel's local governments have been experiencing a fiscal crisis.

A review of the financial reports of the local government in Israel for 2003, as provided by the Ministry of the Interior report, suggests that the situation has not improved. The assessment of the fiscal conditions

of Israel's local government was conducted by implementing various measures. We first assessed fiscal solvency. To measure short-term fiscal solvency, we used two measures: current ratio and self-income ratio. We first measured the *current ratio*, which is defined as the ratio between the total amount of current assets (excluding investments, investments for covering budgeted funds, accumulated deficit in the regular budget, and temporary net deficits in the non-regular budget) and the total amount of current obligations (excluding development funds, budgeted funds, accumulated surplus in the regular budget, and temporary net surplus in the non-regular budget). The total amount of current assets consists of all the current properties for the short run (up to 1 yr), such as customers, bank remainders, investments, and other debtors. The total amount of current obligations consists of all the current debts for the short run (up to 1 yr), such as suppliers and budgeted funds. The data showed that the current ratio ranged from a minimum of 0.02 to a maximum of 4.77, with a relatively poor average of 0.38 (SD = 0.50).

To further assess short-term fiscal solvency, we calculated the *self-income ratio*, which is defined as the ratio between all the self-income and the overall income of the regular budget. Self-income consists of all the income (taxes, grants, and fees) that the local authority directly collects from the inhabitants, businesses, and other assets within its jurisdiction. Regular budget income consists of property taxes, fees, surcharges, general grants, finances from ministries, and individual sources of income. The closer the ratio is to 1, the better the financial state of the local authority. The data indicated that the self-income ratio ranged from a minimum of 0.06 to a maximum of 0.93, with an average of 0.48 (SD = 0.20). Although we used this to measure short-term fiscal solvency, it should be noted that this also reflects the fiscal independence of local authorities, which, according to the data, is poor.

Following Bradbury's suggestion to assess budgetary fiscal distress by analyzing the extent to which local authorities encounter difficulties in balancing their budgets, we calculated the *income-to-expenditures ratio*, which is defined as the ratio between the total amount of income and the total amount of expenditures in the regular (current) budget.^[3] Regular budget income consists of property taxes, fees, surcharges, general grants, finances from ministries, and single sources of income. The total amount of expenditures in the regular budget consists of general administrative services, local services, state services, enterprises, and other expenses. The data indicated a ratio that ranged from -1.04 to 0.05, with an average of -0.11 (SD = 0.13), and found that 78.9% of the local authorities had a deficit in their regular budget and only 4.9% percent had a (minor) surplus in the regular budget.

Finally, we assessed the long-term fiscal condition by calculating the ratio between the socioeconomic level of the population in local authorities and the self-income ratio. The socioeconomic level is a measure provided by the Israel Central Bureau of Statistics, which ranks the local authorities on a 10-point scale ranging from 1 = very poor to 10 = very strong. Results of a bivariate correlation indicate a strong correlation ($r = 0.77$, $p < 0.001$). This indicates that the tax burden is already too high and that citizens do not have the economic means to fund the budgetary deficits.

SOURCES OF FISCAL DISTRESS IN ISRAEL'S LOCAL GOVERNMENTS

There are various sources for the fiscal distress of local governments in Israel (Carmeli, 2007). Below, we have summarized some of the fundamental causes of these difficulties.

Local Authorities' Size

The *size* of a local authority is considered an important factor, as it may have a significant impact on its performance. A local authority can be measured by the number of permanent residents within its jurisdiction and/or the value of the municipal budget. It may be argued that smaller local authorities are unable to reach optimal levels of economic efficiency because of their limited resources and complex challenges. Razin's study indicated that larger localities enjoy better fiscal health than smaller localities.^[8] Findings of data from the financial reports of the local governments in Israel for 2003 indicate that larger local authorities enjoy better fiscal health than smaller local authorities. However, the correlation, although significant, is small.

Location

The *location* of a locality influences its ability to attract new businesses and a strong population. In Israel, about 80% of the population is concentrated within the triangle along the coast and inland toward Jerusalem, which accounts for slightly more than 20% of the country.^[9] Localities located along the periphery put the blame on the central government for its continuous neglect in developing infrastructures that would connect the periphery with the center.

Residents' Socioeconomic Status

The financial sources of municipal services (e.g., sanitation, recreation, and road maintenance) provided

by a local authority are derived from municipal taxes the latter collects from the permanent residents and businesses within its jurisdiction. Clearly, a locality with an inferior socioeconomic resident status encounters difficulty in raising the amount of revenue required for providing satisfactory municipal services. We found a strong correlation between residents' *socioeconomic status* and the self-income ratio ($r = 0.77$, $p < 0.001$). This confirms the above argument that a locality with an inferior socioeconomic resident status is limited in its ability to raise satisfactory revenues, be fiscally independent, and provide the expected level of municipal services.

Governmental Resource Allocation

The way the central government allocates its resources affects the capacity of a locality to provide satisfactory state-level services. Local government leaders make the argument that many local authorities are in a financial crisis due to their task burden and lack of resources. This position is held by several scholars who have investigated the resource allocation issue.^[10] In a policy paper, for instance, presented by Ephraim Torgovnik at the First Annual Local Government Conference held by the Harold Hartog School of Government and Policy at Tel Aviv University (2005), it was indicated that the extensive budgetary cut in central government aid for expenditures that had already been committed to services such as education had resulted in acute fiscal distress, and, even though this cut was later cancelled, the Ministries of Finance and the Interior allocated only 1.5 billion NIS for local authorities that would enter a recovery program. This, along with the fact that there are no clear and explicit definitions of whether the state-level services are state, municipal (local), or both postpolicy and implementation difficulties in allocating resources efficiently.^[10]

Transparency

Transparency is essential for improving the performance of local governments. Transparency should cover everything the organization does or does not do and the reasons behind the actions. This includes input, processes, output, and outcomes. In a policy paper, "Transparency in Local Government," presented at the First Annual Local Government Conference held by Harold Hartog School of Government and Policy at Tel Aviv University (2005), Amichai Ben Dror, Yossi Gross, Doron Navot, and Susan Tam noted that there was sufficient evidence to support the claim that there were patterns of corruption in a significant number of local authorities and that

institutional reform aimed at increasing *inter alia* transparency was needed. Improper management in organizations, whether intentional or unintentional, is a serious problem not only in public sector organizations but also in not-for-profit and for-profit organizations. Transparency facilitates a comparison among the local authorities and drives localities to perform in a better and more responsible way.

Performance Measurement

Although scholars and practitioners recognize the need to assess and measure how well local governments perform, there is no systematic evaluation mechanism through which stakeholders can assess the performance of localities in Israel. Stakeholders cannot make rational decisions without clear and detailed performance reports; however, even more importantly, one may question how local authorities make their decisions without a fundamental evaluation system.

A notable attempt to change this situation was led by the Committee for Implementing Changes in Financial Reporting of Local Authorities (Ministry of the Interior, 1999) to recommend 10 performance criteria that a local authority would be required to include in its financial report (as a management report). These criteria are concerned with 1) the total expenditure for education, sanitation, and infrastructures; 2) generated revenues such as property taxes; 3) changes in internal immigration; and 4) financial performance such as accumulated deficit and deficit per resident. Yet, this is not a comprehensive performance evaluation system and, at best, may be considered as an initial step in the right direction. Suffice it to say, much has yet to be done in this direction.

Management of Local Authorities

One of the major arguments, expressed mainly by the central government, claims that general mismanagement of the local authorities is the main source of their financial difficulties. In his response to local government heads' threats to strike and stop providing municipal services, Nehemia Strasler, a journalist who has been extensively covering the economic situation in Israel, said:

The truth is that even without the last cutback of the balance grants, the local government authorities would have gone on strike, because of their chronic deficits, derived from their continual failure in the management of their organizations. There is not one year in which they have avoided a strike, there is not one year in which they have not committed themselves to execute recovery plans, and yet they continue to implement

the same type of mismanagement that brings them down. A good local authority head is not one who keeps the budget balanced. On the contrary; as long as the (local authority head) pays for and showcases good objectives, as long as she/he provides employment for close associates, she/he would be considered as the better local authority head, regardless of deficits.

—Haaretz (February 2, 2004)

Managing a locality is a very complex task. Although we have witnessed certain improvements, it remains a major problem that affects the capacity of localities to enhance their fiscal health.

Relationships Between the Central and Local Governments

Although the local government operates closely with the central government, this relationship has still not become sufficiently institutionalized. This results in difficulties when attempting to divide responsibilities (who is responsible and accountable for what?) and causes a lack of coordination between the governments. Solving the fiscal crisis of the local government—a mutual interest—may be achieved by redefining the responsibilities for the services and resources required to build a healthy system. The fiscal crisis of the local government affects a wide range of related issues such as education, crime control, and job creation, among others. Neither the central government nor the local government wishes to be perceived as entities that do not deliver public value. Paradoxically, the fiscal state of affairs not only is affected by the problematic relationships between the local and central governments but also acts as a catalyst driving them both to resolve the framework of their responsibilities and to improve their level of coordination.

Differences Between Jewish and non-Jewish Localities

There is a substantial body of evidence suggesting that local authorities in which the majority of the residents are Jewish (hereafter “Jewish localities”) are in superior fiscal health compared to local authorities in which the majority of the residents are non-Jewish (Arab and Druze) (hereafter “non-Jewish localities”). For example, Carmeli examined a sample of 98 localities, which is a representative sample of the local government in Israel, and found that the correlation between sectors (1 = Jewish and 0 = non-Jewish) was relatively large and significant ($r = 0.44$, $p < 0.001$).^[11] One clearly should take into account differences between these types of localities when analyzing the fiscal distress/health of the local government

in Israel. Clearly, these differences are a complex issue and require particular and intensive attention in order to understand its sources, consequences, and changes over the years.^[12]

CONCLUSIONS

The fiscal distress of the local government in Israel is a complex phenomenon that may be seen, on the one hand, as part of a wider, international phenomenon, as well as a case that has some idiosyncratic characteristics on the other. The current entry has highlighted some of the key issues with which the local government has been coping over the years. Some of these issues are structural (i.e., require substantial restructuring), while others call for special and focused strategies. Hence, both macro- and microlevel issues need the attention of policymakers, in order to improve the fiscal conditions of the local government in Israel.

ACKNOWLEDGMENTS

The author wishes to thank the Area Editor, Eran Vigoda-Gadot, and an anonymous reviewer for their helpful comments and suggestions.

REFERENCES

1. For a review, see Honadle, B.W. The states' role in U.S. local government fiscal crises: a theoretical framework and results of a national survey. *Int. J. Public Adm.* **2003**, 26 (13), 1431–1472.
2. Carmeli, A. A conceptual and practical framework of measuring performance of local authorities in financial

terms: analysing the case of Israel. *Local Gov. Stud.* **2002**, 28 (1), 21–36.

3. Bradbury, K.L. Fiscal distress in large U.S. cities. *N. Engl. Econ. Rev.* **1982**, November/December, 33–44.
4. Groves, S.M.; Valente, M.G. *Evaluating Financial Condition: A Handbook for Local Government*, 3rd Ed.; International City/County Management Association: Washington, DC, 1994.
5. Inman, R.P. How to have a fiscal crisis: lessons from Philadelphia. *Am. Econ. Rev.* **1995**, 85 (2), 378–383.
6. Hecht, A. *Restructuring Municipal Finance in Israel*; The Floersheimer Institute for Policy Studies: Jerusalem, 1997 [in Hebrew].
7. Kalchheim, C.; Rozevitch, S. Deficits in local government budgets in Israel: a reflection of political cycles and an expression of local autonomy. *Public Budget. Finance* **1990**, 10 (1), 67–77.
8. Razin, E. *Fiscal Disparities Between Small and Large Municipalities in Israel*; The Floersheimer Institute for Policy Studies: Jerusalem, 1999 [in Hebrew].
9. Sharkansky, I. Israel: a metropolitan nation-state. *Cities* **1997**, 14 (6), 363–369.
10. Ben-Elia, N. *Government Finance and the Fiscal Crisis in Israeli Local Authorities*; The Floersheimer Institute for Policy Studies: Jerusalem, 1999 [in Hebrew].
11. Carmeli, A. The relationship between organizational and structural characteristics and local authorities' fiscal health. *Public Admin. Q.* **2003**, 26 (4), 426–464.
12. Al-Haj, M.; Rosenfeld, H. *Arab Local Government in Israel*; Westview Press: Boulder, CO, 1990.

BIBLIOGRAPHY

1. Carmeli, A. 1990. The fiscal distress of local governments in Israel: Sources and coping strategies. *Administration & Society* **2007** (forthcoming).

Fiscal Federalism

William Duncombe

Department of Public Administration, Syracuse University, Syracuse, New York, U.S.A.

INTRODUCTION

Fiscal federalism describes the division of fiscal resources and responsibilities among levels of government. Fiscal federalism is not just confined to countries with a federalist form of government, but can exist in any country where both central and local governments are assigned power over resource allocation and revenue-raising decisions.^[1] A brief review of theories and research on fiscal federalism is first presented. Then the concepts of fiscal federalism are discussed, by examining some recent changes in intergovernmental fiscal relations in the United States.

THEORIES OF FISCAL FEDERALISM

Theories of fiscal federalism develop a framework for deciding which level of government should provide a particular government service, with which sources of revenue, and what fiscal relationships should exist between them. Public finance economists, who have written extensively on this topic, emphasize that economic criteria, such as allocative and productive efficiency, should be used in making these decisions.^[2] Allocative efficiency refers to matching the type and level of government service provision with the demands of citizens, and productive efficiency is achieved when the maximum output is produced for a given budget.

Assignment of Functions

In his seminal work on fiscal federalism, Wallace Oates offered a convincing case for division of service responsibilities across levels of government.^[1] His framework for assignment of functions is based on the assumption that people and resources can flow freely across borders of local governments, but not across national boundaries. The economic functions of government can be divided into three major categories—stabilization, distribution, and allocation. Each function entails responsibilities, which may be best administered at a particular level of government.

Stabilization involves using monetary and fiscal policies to maintain employment, economic growth, and price stability. Macroeconomic functions of

government are best performed at the federal level, in part, because of the need for central control of the banking system and money supply. Also, the mobility of resources across local government borders greatly weakens the influence that any particular local government can have in the short run on economic growth within its borders.

Distribution describes the role government can play in assuring a socially desired distribution of resources across individuals within a country. It is generally argued that redistributive programs are best implemented by central governments, with a supplemental role for regional governments, such as states and provinces. Mobility across local government borders limits the ability of local governments to offer significantly better social services than neighboring governments. Evidence suggests that recipients of these services are mobile and will move to locations where services are better.^[3] Turning these social services over exclusively to local governments may promote a “race to the bottom” as local governments lower social services to encourage low-income households to move to a different community.^[4]

Allocation and production of socially desired goods and services remains the principal role of government. According to Oates’ *correspondence principle* the size of the government providing a service should be the smallest government, which includes all the beneficiaries of the service.^[1] National government should provide only those services, such as defense, immigration, and national transportation systems, that benefit a national constituency. Regional governments should specialize in services that have significant benefits extending beyond local boundaries, such as pollution control, corrections, and higher education. The remaining services should be decentralized to the local government level.

Local governments, by representing a smaller constituency, can best match government services to local citizen preferences, improving allocative efficiency. In addition, monitoring government performance is easier for citizens in smaller governments, which should also lead to improved productive efficiency. These efficiency benefits of smaller governments must be balanced against the potential cost savings from economies of scale for certain types of services, and the additional costs of running a number of smaller governments.^[1]

The efficiency advantages of decentralization could be enhanced by competition among local governments as households “vote with their feet” by choosing local governments with the best tax and service package.^[5] However, interjurisdictional competition over tax rates and services levels may lead to an underprovision of some public services.^[3] In an attempt to attract business, local governments may cut taxes and services not valued by business to levels below that desired by citizens in the community. The efficiency impact of competition may depend in part on whether the government is attempting to attract business or mobile households looking for the optimal mix of taxes and services.

Assignment of Taxes

An equally important part of an intergovernmental fiscal system is deciding which taxes are best suited for use at national and subnational levels of government.^[3] Key tax criteria include minimizing efficiency losses from the tax and enhancing vertical equity. In general, the larger the private sector response to avoid the tax, the larger the efficiency loss. Enhancing vertical equity involves designing tax systems that impose tax burdens on individuals in line with their ability to pay. In particular, we want to generally avoid regressive taxes, where taxes as a percent of income are higher for low-income households than high-income households.

Assuming mobility of resources within a country, subnational governments will have difficulty using progressive taxes. Any attempt to impose higher tax rates on upper-income individuals or owners of businesses could lead them to relocate to a local government with a more favorable tax climate. This mobility will increase the efficiency losses of the tax and reduce potential revenues. Instead, an ideal local tax system from an efficiency standpoint would involve the use of either user charges for services rendered, or taxes on relatively immobile tax bases.^[2] The most obvious example of the latter would be a property tax on unimproved land, which is completely immobile. Although property taxes and user charges may create relatively small efficiency losses, they are generally considered to be regressive and slow-growing sources of revenue.

The potential regressivity and inelasticity of local taxes suggests the need for the national government (and to a lesser degree, states and provinces) to use progressive taxes, which are responsive to economic growth. Income taxes, in particular, match these criteria, because they are tied directly to personal income changes, and they can be implemented with a progressive tax rate structure. General sales taxes are also responsive to economic changes, but can be regressive

unless necessities, such as food and clothing, are removed from the tax base.

Intergovernmental Aid

Intergovernmental grants represent a key tool in the design of an intergovernmental fiscal system. The most efficient division of functions across governments may involve assigning many of the most important and costly functions to subnational governments. Yet, optimal tax policy indicates that some of the largest and fastest-growing revenue sources should be at the national or regional level. The “vertical imbalance” between central revenues and local expenditures can be addressed through the distribution of revenues from the central government to local governments in the form of intergovernmental grants. Besides resolving the “vertical imbalance problem,” intergovernmental grants can be used by the central government to encourage local governments to increase their provision of services that have benefits beyond the boundaries of the local government.^[6] For example, local governments are often responsible for environmental infrastructure, such as waste treatment plants and solid-waste disposal facilities. Because the effects of water pollution will literally flow outside the boundaries of the local government, central governments can encourage local governments to improve waste treatment by providing grants. In addition, grants can be used to adjust for differences between local governments in their ability of finance public services. The largest grant programs in most state governments in the United States are to local school districts, and are used in part to equalize the fiscal capacity of these districts to finance education services.

The design of intergovernmental aid systems involves decisions over several dimensions. First, should the grant be targeted to the provision of specific services or resources, commonly called categorical grants, or should unconditional grants be provided to supplement general revenues (also called revenue sharing)? Another choice is whether the grant should be a lump-sum transfer of resources to a local government or in the form of a price change. By agreeing to match local contributions in the form of a matching grant, central governments can effectively lower the price for local governments to provide particular services. A third design decision is whether the grant is distributed by a predetermined formula (formula aid) or requires the local government to formally apply to receive funds (project grants).

Ideally, grant design should match the objectives of the grant. If the grant is to be used primarily to stimulate an increase in local spending on a particular service, then the grant should generally be a categorical

open-ended matching grant. Whether the grant should be distributed by formula or through a project application depends on the complexity of the service being provided, and whether there are concerns about local capacity to use the funds effectively. If the grant is going to be used primarily to correct for horizontal inequities across governments, then a non-categorical, lump-sum formula grant is preferred, with the formula carefully designed to adjust for differences in fiscal capacity and costs across local governments.^[7] If the grant is being used primarily to compensate local governments for using smaller, less elastic revenue sources, then non-categorical grants distributed roughly in proportion to the central government revenue collected in the local government might be justified. General revenue sharing (GRS), passed in 1972 under the Nixon Administration, was the only major attempt by the U.S. federal government to develop a general grant to address vertical imbalance problems. General revenue sharing was eliminated in 1986 by another Republican president, Ronald Reagan.

FISCAL FEDERALISM IN THE UNITED STATES

Recent Trends

In general, the intergovernmental fiscal system described above closely mirrors the division of fiscal responsibilities in the United States. The empirical evidence suggests that, for many government functions, fiscal federalism and the correspondence principle seem to be operative. Local governments, for example, tend to provide goods and services, such as public safety, housing, and recreation whose benefits are somewhat limited by jurisdictional boundaries. However, several of the major services provided by local government, particularly education, wastewater treatment, and solid waste extend beyond local borders. State governments are involved with goods benefiting larger groups of individuals, such as transportation, corrections, and higher education. Contrary to the theory, states are major providers of distributive programs in social services and health. The federal government is clearly responsible for national public goods, such as defense and foreign affairs, and has funded major social and income support programs in the past (Table 1).

In terms of assignment of revenues, the general pattern across levels of governments also fits recommendations. The federal government generally taxes the most mobile and elastic sources of revenue—corporate and personal income. Given that corporations are often mobile across state boundaries, the federal government should be the principal user of this tax. The federal government also receives significant revenue in the

Table 1 Government expenditures by function in the United States by level of government, 1992 (percent of total expenditures)

Function	Total	Federal	State	Local
General expenditure	74.7	70.2	87.3	86.9
National defense and international relations	14.1	23		
Postal service	1.8	2.9		
Space research	0.5	0.9		
Education and libraries	14.1	3.2	30.1	36.8
Social services and income maintenance	13.7	13.1	29.6	12.2
Transportation	3.5	1.7	7.7	5.6
Public safety	3.8	0.7	4.2	8.8
Environment and housing	6.1	6.1	2.5	9.6
Government administration	2.7	1.3	3.7	4.7
Interest on general debt	10.2	13.1	3.5	4.6
Other	4.2	4.2	6	4.5
Utility and liquor store	3.4		1.4	11.4
Insurance trust	21.8	29.8	11.3	1.6

Source: U.S. Bureau of the Census, *Compendium of Government Finance*, 1992 Census of Governments, Table 1.

form of payroll taxes for social insurance programs (social security). State governments rely most heavily on personal income and sales taxes. For both types of taxes, mobility across borders could be a problem, but only for residents and consumers living near state borders. As recommended by theory, local governments are heavy users of property taxes and user charges and fees (Table 2).

Table 2 Major revenue sources in the United States by level of government, 1996 (percent of total revenue)

Revenue source	Total	Federal	State	Local
Intergovernmental revenue	14.8	0.2	22.9	33.7
General revenue from own sources	59.9	64.5	56.8	54.6
Taxes	45.9	53.7	43.3	33.7
Property	6.3	0.0	1.0	24.8
Individual income	22.0	37.5	13.8	1.7
Corporation income	5.7	10.0	3.0	0.3
Sales and gross receipts	9.7	4.8	21.3	5.3
Other	2.2	1.4	4.1	1.5
Charges and miscellaneous revenue	14.0	10.9	13.5	20.9
Utility and liquor store revenue	2.3	0.0	0.7	8.5
Insurance trust revenue	23.0	35.3	19.6	3.3

Source: U.S. Bureau of the Census, *Statistical Abstract of the United States*, 2000, Table 494.

The general trend in fiscal federalism in the United States is toward increasing state and local government reliance on federal grants. Since the 1970s, total federal aid has represented roughly 3% of gross domestic product, and has grown from 5.3% of total federal outlays in 1955 to 16% in 2000 (Table 3). From 1950 to 2000, federal grants as a percentage of state-local revenues increased from 13.7% to nearly 30%. State-local reliance on federal aid peaked in 1980, when federal aid represented close to 40% of state-local revenue. Although federal aid to states and local governments declined in the early 1980s, by the early 1990s, its importance to state and local revenues began to increase again. Despite several attempts by Republican presidents to increase the use of block grants since the 1970s, categorical grants continue to dominate federal aid. In 1995, estimates suggest that nearly 90% of all federal aid to states and local governments took the form of categorical grants (Table 4).

Table 5 provides information on the changing functional composition of federal grants since 1965. Transportation-related grants have decreased in importance, whereas health related grants—Medicaid, in particular—have increased dramatically as a percent of total federal aid. Postwar federal investments in the national interstate highway system represented significant portions of federal aid prior to 1965, but they have diminished as the system nears completion. Transportation grants dropped from 38% of federal aid in 1965 to 15% in 2000. Despite federal funding increases for environmental infrastructure in the 1970s, federal investments in state and local infrastructure have plainly lost ground to other priorities.

Table 3 Trends in federal grants to state and local governments in the United States

Year	Federal grants as a percent of		
	State-local revenues	Federal outlays	Gross domestic product
1950	13.7	5.3	0.8
1955	13.1	4.7	0.8
1960	19.1	7.6	1.4
1965	20.4	9.2	1.6
1970	26.3	12.3	2.4
1975	33.7	15.0	3.2
1980	38.3	15.5	3.3
1985	27.9	11.2	2.6
1990	25.0	10.8	2.4
1995	31.0	14.8	3.1
2000	29.3	15.9	2.9

Source: U.S. Office of Management and Budget. *The Budget of the United States for Fiscal Year 2002*, Tables 12.1 and 15.1.

Table 4 Composition of federal aid types, percent of total dollar outlays, 1975–1995

Year	General purpose	Broadbased	Categorical
1975	14.1	9.2	76.7
1979	12.3	14.7	73.0
1981	7.2	10.6	82.2
1984	7.0	13.3	79.7
1987	1.9	12.1	86.0
1989	1.9	10.4	87.6
1991	1.4	10.8	87.8
1993	1.1	10.6	88.3
1995	1.0	10.0	89.0

Source: Advisory Commission on Intergovernmental Relations. 1995. *Characteristics of Federal Grant-in-Aid Programs to State and Local Governments*.

In inflation-adjusted dollars, federal spending on state and local environmental infrastructure dropped from \$6.5 billion in 1980 to \$2 billion by 1996.^[8] In contrast, federal grants devoted to health, income security, and employment and training increased from 48% of total federal aid in 1965 to 75% in 2000. Clearly, the distributive function of the federal government is crucially important in the current intergovernmental fiscal system.

Devolution and Block Grants

The most important current reforms in fiscal federalism involve the incorporation of more flexible, block grant-style federal aid mechanisms, with increased “devolution” of program responsibility to state governments. The bulk of federal aid, as well as the focus of most reforms, was that portion of the social safety net that was jointly financed by federal and state governments, namely Aid for Families with Dependent Children and Medicaid. In 1996, President Clinton signed the Personal Responsibility and Work Opportunity Reconciliation Act (PRWOR), which includes substantial reforms for federal and jointly funded income support programs. A critical component of the reform is a shift from an open-ended matching formula to a block grant for Temporary Aid for Needy Families (TANF), which replaces the 30-year-old AFDC program. Under the new law, the federal block grants require states to impose strict time limits for individual TANF benefits, as well as stringent work requirements.^[8]

The official justification for moving to state block grants has traditionally been to provide states with more flexibility in designing and administering their programs. Locating income support programs in the states may improve program design and efficiency

Table 5 Changes in functional composition of U.S. federal grants to state and local governments (percent of total federal aid)

Function and fund group	1965	1970	1980	1985	1990	1995	2000
National defense	0.3	0.2	0.2	0.1	0.1	0.1	0.2
Energy	0.1	0.1	0.1	0.2	0.5	0.5	0.3
Natural resources and environment	1.7	1.8	1.7	5.0	5.9	3.8	3.0
Agriculture	4.7	2.9	2.5	0.5	0.6	2.3	1.1
Transportation	37.6	31.6	19.1	11.3	14.2	16.1	14.9
Community and regional development	5.9	4.5	7.4	9.1	7.1	4.9	3.3
Education, training, employment, and social services	9.6	20.0	26.7	26.4	23.9	16.8	18.0
Health	5.7	9.0	16.0	16.3	17.2	23.1	30.1
Income security	32.2	27.8	24.1	17.7	20.2	25.7	26.7
Veterans benefits and services	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Administration of justice			0.2	0.7	0.6	0.1	0.4
General government	2.1	2.0	2.0	12.4	9.4	6.5	1.8

Source: U.S. Office of Management and Budget. *The Budget of the United States for Fiscal Year 2002*, Table 12.2.

through state experimentation. In addition, there may be allocative efficiency gains associated with enabling citizen preferences to be more closely reflected in state spending decisions. Whether the potential efficiency gains from state administration of welfare programs will materialize is an area of substantial debate.^[3] However, block grants have also commonly been used to cut federal aid to states.

These reforms constitute fundamental shifts in the design of federal aid to support the social safety net and will undoubtedly have a significant impact on the spending behavior of states. Devolving redistributive social programs from the federal government to state governments runs counter to theories of fiscal federalism that suggest such programs should be operated at the national level. Because PRWOR eliminates the price subsidy offered through the matching grant formula, state spending in this area could drop substantially in the coming years.^[4] Spending decreases on welfare programs could be exacerbated by competitive pressure between states to keep tax rates low by cutting social services.

The actions of the federal government in the United States to devolve functions to state and local governments pose several challenges to conventional theories of fiscal federalism. The federal government has drastically cut back funding for state and local environmental infrastructure even though the benefits from these facilities clearly extend beyond state and local borders. The devolution of key programs in the social safety net to state governments risks competition among subnational governments that could lead to

an underprovision of these services. It seems inevitable that state and local governments will come under intensified fiscal pressures as the economy slows down over the next decade. The ability of these subnational governments to maintain their core infrastructure and the social safety net with diminished federal support will indicate whether it is time to write a new chapter in fiscal federalism.

REFERENCES

1. Oates, W.E. *Fiscal Federalism*; Harcourt Brace: New York, 1972.
2. Bird, R.M. Threading the fiscal labyrinth: some issues in fiscal decentralization. *Nat. Tax J.* **1993**, *46* (2), 207–227.
3. Oates, W.E. An essay on fiscal federalism. *J. Econ. Lit.* **1999**, *37* (3), 1120–1149.
4. Brueckner, J.E. *Welfare Reform and the Race to the Bottom: Theory and Evidence*; Institute of Gov. and Public Affairs, U. Illinois, 1998; WP 64.
5. Tiebout, C. A pure theory of local expenditures. *J. Polit. Econ.* **1956**, *64*, 416–424.
6. Duncombe, W.D.; Johnston, J.M. Intergovernmental fiscal relations: revisions and reforms. In *The White House and the Blue House: Government Reform in the United States and Korea*; Cho, Y.H., Frederickson, H.G., Eds.; University Press of America: New York, 1997; 271–309.
7. Gramlich, E.M. A policymaker's guide to fiscal decentralization. *Nat. Tax J.* **1993**, *46* (2), 229–235.
8. Ladd, H.F.; Yinger, J. *America's Ailing Cities: Fiscal Health and the Design of Urban Policy*; Johns Hopkins Univ. Press: Baltimore, MD, 1989.

Fiscal Transparency

A. Premchand

Consultant, Irvine, California, U.S.A.

INTRODUCTION

The content of fiscal transparency has evolved over years. The cumulative developments in this area are expected to provide an overview on the wisdom, faithfulness, and economy in the use of resources, in the provision of services, and in securing macroeconomic stability. Despite the efforts of international financial institutions to strengthen fiscal transparency, about one-third of the world's population does not have full access to information on the fiscal operations of public bodies. More efforts are indicated to expand the scope of fiscal information and to establish bodies of oversight. What remains to be achieved is significant and urgent.

HISTORY OF FISCAL TRANSPARENCY

The contours of fiscal transparency and its content were shaped over the years by two distinct trends—a desire to make the public officials accountable for their actions, and the political arithmetic of the times reflecting the concerns of the financial class and their interest in investing money in instruments of indebtedness. The evolution of fiscal transparency to date can be analyzed in terms of five stylized stages. During the first stage, the concerns of financial accountability of the monarchs and the concerns of an active society seeking a role in the utilization of public money—a feature associated with the Athenian state in the pre-Christian era—dominated the fiscal scene. The concern of the king was the preservation of the wealth of his own domain, and this required him to devise ways and penalties that would have the effect of preventing his officials from stealing. Writing more than 2000 years ago, a prime minister of a small North Indian kingdom cautioned his king to “Look well to the Treasury, for it is the key to all.” Variants of these approaches were followed in China, and this subsequently became the main feature of the Cameralist school associated with European kingdoms during the Middle Ages. The emphasis in a monarchy was on transparency that was intended for only one person, viz. King, and his audit agency. The concerns were the same in the development of the Athenian state, where there was explicit recognition, amongst others, by Aristotle, of the risks

associated with handling of large sums of money by officials and the need for systematic accounts that would illumine the whole area.^[1,2]

In the second stage, there emerged an investing class, and the political and financial arithmetic associated with their investments in government during seventeenth-century England had an impact on the content of fiscal transparency. A continuous engagement in wars depleted the British Treasury and made it dependent on private sources of financing raised through borrowing. The investing class was, however, keen to have detailed accounts of fiscal transactions of the government so that the financial health of the government could be ascertained. As these accounts were not forthcoming, the investors resorted to the appointment of their own staff to undertake independent compilation of accounts. In due course, however, this had the effect of government appointing its staff and led to the emergence of accountants as a class in government. The overall development of a new nexus between governments and their investing public contributed to the colonization of the state by financial interests.^[3,4] During this stage, the need for transparency received wider acceptance, although it was seen primarily as that of the investing class.

During the third stage, reflecting the gradual development of the legislative institutions and their powers to review the “wisdom, faithfulness and economy” with which parliamentary grants were spent, efforts were made to appoint a Commission of Accounts and a Commission on Audit (a predecessor of audit as practiced now), culminating in the establishment of a permanent national audit agency to review the use of approved financial resources. This stage emphasized that transparency was not an end in itself but was a means for the legislature to ensure accountability of the government and its officials. What was however, within the domain of the legislature, was also within the reach of the public. These British practices, were in due course, replicated in the British commonwealth countries and variations of their approach were introduced in others. A result of these efforts was the explicit recognition of the need for fiscal transparency as a feature of a democratic government.

The fourth stage represents the more recent experience of governments, particularly since the 1950s. The growth of the welfare state, as well as the steady

expansion in the size of governments and the range of activities undertaken by them, contributed to an inexorable growth in expenditures and, over the years, to growing fiscal deficits and crisis. The emergence of fiscal crisis contributed to the erosion of the public trust in government and to greater demands for the involvement of the public, along with its legislative institutions, in shaping their own fiscal destiny. It was suggested that improvements in governance required the empowerment of the community and its participation, whenever possible, in the formulation of policies. In turn, this implied that there should be greater fiscal transparency so that the community recognizes where it is and determines where it should go. The idea of improved governance came to be dependent on the reality of fiscal transparency.

The fifth stage reflects the series of financial crises in the 1990s that had a substantive impact on the emerging market countries. These countries, dependent on capital inflows, were highly vulnerable to sudden changes in market perceptions (i.e., perceptions of the investing class) and their destabilizing effects. Preventive action, aimed at forestalling the crises, or at a minimum, reducing the intensity, required the strengthening of the fiscal system, and enhancement of fiscal transparency. Such transparency was aimed at correcting the information asymmetries and associated belief systems. Cumulatively, the result is that extensive fiscal transparency has become an essential feature of governments. The content of fiscal transparency has not, given the diverse developments, however, been articulated coherently, except to some extent by international financial institutions.^[5,6]

Objectives

Reflecting the various influences in the mutation of fiscal transparency, the objectives sought to be achieved comprise three groups:

- *Stewardship of resources:* Governments should provide data on the state of finances, for the past, present, and the future so that the community can make its own assessment about the viability of the policy stance, including the preventive actions taken or contemplated to reduce or avoid financial market failures. This requires that the information be comprehensive, including all activities, as well as contingent liabilities, on a consistent basis. The data must comply with specified standards.
- *Adequacy of the fiscal machinery:* Information is needed on the various aspects of tax administration, expenditure management, lending and borrowing operations, sales and purchase operations, and management of the financial portfolio. Efforts in

this regard are aimed, in part, to restore the credibility of the public management systems and to assure the community of the continuing effective functioning of the fiscal machinery. As an integral part of this effort, attention paid to ensuring the due process, prevention of opportunities for corruption, and the smooth working of the accountability channels associated with legislative or other forms of social action is revealed to the public.

- *Decision-making approaches:* There should be a window of opportunity for the community to be informed about the decision-making approaches behind the fiscal policies sought to be pursued. The window should enable an understanding, even as decisions are made (and not after they have been made) on the main components of fiscal policy—pursuit of macroeconomic stability, effective performance in the delivery of services, and pursuit of economy and efficiency.

These objectives aim at the complete fulfillment of fiscal transparency and, to that extent, go beyond the guidelines issued by international financial organizations. The guidelines issued by these organizations place more emphasis on the legal framework (intended more for international audiences), linkages with macroeconomic trends and policies, operational procedures, dissemination of fiscal data, and channels of accountability with legislatures. Correspondingly, less focus is laid on the adequacy of the fiscal machinery and delivery of services, aspects that are of critical importance to the community.

Content of fiscal information

The fiscal information provided to the public and the arrangements for transparency have their roots in the government accounting system, which is a homogeneous source of supply. The users of the information are, however, varied and the heterogeneous groups that include the community, legislatures, market analysts, investors, policy makers, external donors, and others. Information aimed at meeting these diverse requirements is provided through various channels. The features of the information supplied are shown, in detail, in [Table 1](#).

Uses and limitations

The demands for additional information on government fiscal operations have grown during recent years. Along with these demands, the capacity of government has also increased severalfold, to meet them. To a large extent, this has been facilitated by the application of computer technology to government transactions, transforming a long drawn-out, slow, and often

Table 1 Fiscal transparency: components, instruments, and features

Functional area	Instruments	Features
I. Structures and policy spheres		
(a) Structures		
• Functions and fiscal responsibilities of central, state, and local governments • Transfers from the central government to the state and local government; where revenue collections are decentralized, states, regions, or provinces may make transfers to the central government	• These are, in general, specified in the constitution and associated legal framework. • These transfers may be specified in the Constitution. Annual transfers determined with reference to legal criteria may be shown in the budget documents; where budgets are not available to the public (see below), the quantum of these transfers and their utilization may not be known.	• In federal types of countries, the financial relationships among these levels tend to be complex. • The experience of federal types of countries shows that these transfers, including devolution of resources, are determined in a quasijudicial process; the findings, as distinct from proceedings, are made public. • Experience of unitary governments is far more diverse; the details are contained in the budget documents (and are specifically shown as transfers), while in a few cases, they remain obscure.
• Autonomous bodies of the central governments	• These bodies have, in general, their own budgets, and the extent of dependence on government is shown therein.	• Access to these documents is problematic and dependent on the laws governing the autonomous bodies.
• State-owned enterprises	• Enterprises are obligated, in most cases, to publish their annual balance sheets and accounts. Their budgets, where the activities tend to be commercial, are not made public. The main instruments of transparency are the accounts and the periodic reports published by them. These reports may be more frequent where there is a high degree of dependence on the capital markets.	• Their accountability to the legislatures differs very widely: • The accounts and associated reports may not specify the costs of non-commercial objectives that the enterprises follow at the behest of governments. Moreover, subsidies given by governments may not have the specific end use, given the fungibility of resources, indicated by governments.
• Relationship with central bank/monetary authority and other public financial institutions	• These relationships are usually specified in the form of a law, inter-alia, indicating the tasks and responsibilities of the institutions in the management of public debt. The financial institutions publish their annual accounts and other reports that are within the public domain.	• Although the accounts are available, the quasifiscal accounts or activities undertaken by them may not be fully shown. Standards in this regard are evolving, and it is likely that more detailed information will be available on the quasifiscal activities.
• Relationships with the corporate private sector	• In most cases, the patterns of government equity ownership and the responsibility of governments in regard to regulation are specified in law (e.g., Companies Act). In some cases, regulation may be enforced through autonomous agencies.	• Laws and associated guidelines are generally matters that can be enforced through judicial means. In some countries, however, the ownership patterns are both highly complex and opaque.

(Continued)

Table 1 Fiscal transparency: components, instruments, and features (*Continued*)

Functional area	Instruments	Features
(b) Relationships with legislatures	- The patterns vary among countries. Broadly, there are three types of legislatures 1) where legislatures have the dominant role, 2) where the executive has the dominant role, and 3) where legislatures do not have any role in terms of annual management of finances, except that they are kept informed on all key aspects. In several countries, there is a framework of delegated legislation within which the executive wing of the government has powers to levy taxes, duties, fees, etc., and to spend the proceeds. - Legislatures are provided with medium-term fiscal strategy, development plans, medium-term budgets, annual budgets, and supplementary budgets. All these documents are in the public domain.	The usefulness of the documents prepared by governments to determine the current fiscal status and its viability varies considerably among countries. In some countries, budgets are not available to the public; instead, summaries of the main features are presented.
(c) Primary fiscal instruments	- The range of fiscal instruments has expanded over the years. In addition to the annual budget intended for purposes of legislation, several documents seeking to provide background information are provided. These include multiyear rolling budgets and other documents referred to above. The most important instrument continues to be the budget. To serve the purposes of transparency, budgets are required to be: <i>Comprehensive</i>, covering all transactions, including foreign aid, proceeds of taxation, gross spending, gross borrowing and lending, and gross buying and selling. - As <i>integral part</i> of the above and reflecting the full status of finances, data on extrabudgetary accounts, quasifiscal activities, tax expenditures, guarantees provided, and consequent contingent liabilities are also needed; - On a <i>gross basis</i> to avoid misleading conclusions, expenditure offsets, etc., are to be avoided.	- Although all countries have budgets, their features vary considerably. - In many cases, large chunks of expenditures may be organized in the form of extrabudgetary activities or as quasifiscal activities. This feature limits the usefulness of the main budget. More significantly, in the absence of data on guarantees and related contingent liabilities, the community may not have the data to assess critically the current status of government finances and their sustainability. - In general, many budgets are organized on a gross basis, but a frequent resort to netting is also common.

(Continued)

Table 1 Fiscal transparency: components, instruments, and features (*Continued*)

Functional area	Instruments	Features
	• <i>On an accounting basis</i>, that reflects the cash position as well as liabilities. • <i>A classification</i> of government transactions into a functional program, as well as economic types, to illustrate the broad purposes of expenditures.	• Most budgets are organized on a cash basis. During recent years, attempts have been initiated to introduce forms of accrual budgeting. • Patterns of classification vary considerably among countries. Different practices are in vogue for budgets and for international reporting. Categories such as defense remain, for the most part, opaque.
II. Fiscal Management		
(a) Objectives	• The range of instruments includes fiscal responsibility legislation, medium-term rolling framework, and development plans. These documents provide the essential context within which the annual budget serves as the most important fiscal instrument.	• During recent years, some countries, such as Argentina, Brazil, etc., have—following the experience of Australia and New Zealand—enacted fiscal responsibility legislation indicating the broad goals of fiscal policy.
(b) Consolidated budget	• The budget and its supplementary instruments seek to reflect the totality of government finances within a country. Because regional/state/provincial budgets have an important role, they need to be consolidated for the country as a whole.	• In unitary forms of government, budgets reflect the total picture. In federal types, however, the lower levels of government formulate their own budgets. In several countries, these are consolidated, usually after a lag, and made available to the public. Even here, the budgets of local governments (county and city governments), being too many, may not be included in the consolidation exercise.
(c) Macroeconomic framework	• The purposes of the annual budget reflect the times. To facilitate an understanding of this crucial element in fiscal policy, assumptions about the growth in the economy, exchange rate, inflation rate, aggregate demand factor, and estimated capital flows are indicated.	• The actual experience in this regard varies considerably. In most cases, the macroeconomic framework may not be explicit. In some cases, autonomous research organizations may publish them as a part of their studies. Objectives, where indicated, tend to be too general and may not constitute a strategy.
(d) Annual policy making	• The annual budget, being the main vehicle, is expected to indicate the following: • New policies to address the problems in the economy. As an integral part of this effort, new outlays are distinguished from continuing policies. • Changes in revenue policies including changes in tax expenditures. • Changes in lending policies.	• The approaches to annual budget vary among countries, depending on the role of the legislature. • Although new policies are highlighted, the expenses associated with them and their continuing financial implications may not be adequately analyzed. • Tax expenditures, guarantees, and changes in lending policies do not, in general, receive specific focus in the documents.

(Continued)

Table 1 Fiscal transparency: components, instruments, and features (*Continued*)

Functional area	Instruments	Features
	• Policies aimed at deficit containment including austerity management. • Changes in public debt patterns, including maturity. • Changes made in delivery of services and expenditure benefits. • Identification of high-risk areas (e.g., foreign aid). • Some of these areas may be made public before the presentation of the annual budget, through public announcements, white papers, and sessional documents.	• Deficit reduction has been a dominant theme during recent years. Austerity management, however, may not be specified in detail to prevent lobbying by interest groups. • Although data on outstanding domestic and foreign debt are provided, their holders and maturity patterns—in particular, in the area of foreign debt—may not be indicated in detail. • Expenditure benefits have not received adequate attention until recently. • Risk areas and associated efforts at risk management remain, where undertaken, at a nascent stage.
(e) Detailed objectives of departments and agencies	• The objectives to be achieved during a fiscal year are to be stated in detail in each agency budget.	• Objectives, where indicated, tend to be too general and may not constitute a strategy.
(f) Performance orientation	• The agency budgets should also indicate the achievements, in terms of delivery of services, expected during the fiscal year. This constitutes an important vehicle for full accountability. As an integral part of this data on cost–service–quality linkages are sought to be provided.	• Performance budgeting in governments has had a chequered career. During recent years, there has been a revival of some aspects of this system. Many countries have yet to implement this.
(g) Pursuit of efficiency and economy	• Agency budgets are also expected to indicate the specific efforts made to achieve economy and efficiency.	• To a large extent, this remains to be fulfilled.
(h) Changes in supporting administrative infrastructure	• Governments announce, through annual budgets, changes made in the tax collection machinery, expenditure management, and debt management. In tax administration, the effort is to augment the legal specification of the tax basis, while reducing the administrative discretion in its application. In expenditure management, the intent is to facilitate internal financial operations of the spending agencies. • In debt management, the instruments seek to provide more information on the currency of debt, separation of interest from principal, etc.	• In tax administration, the major area of darkness continues to be the use of discretionary powers by tax officials and related rent-seeking behavior. In expenditure management, procurement and related contracting remain, as noted below, less transparent. Debt transactions also remain, in some cases, shrouded in secrecy.

(Continued)

Table 1 Fiscal transparency: components, instruments, and features (*Continued*)

Functional area	Instruments	Features
III. Implementation of budgets	- Notwithstanding differences of degree, several stages of policy implementation are common to all countries. These include the following: - Phased release of budgetary authority - Organized cash management that facilitates a linkup with debt management - Award of contracts, procedures for tendering and contracting - Procedures to ensure budgetary outcome as intended - Specification of performance measures	Several of the activities enumerated are usually considered as “in-house” activities and, as such, may not be open to the public. The primary issue for the community is whether there is adequate machinery empowered to deal with these aspects and whether the machinery is working effectively. In practice, the most controversial area is the award of contracts. The range of contracts has widened considerably and now includes services such as day care centers. These contracts may not, in several cases, be in the public realm. Specification of performance is still at a nascent stage. The budgetary outcome may, in some cases, be technically managed through the accumulation of payment arrears.
IV. Accounting and reporting	The accounting system is intended to ensure that the tasks inherent in the stewardship of money are being handled. It should therefore specify 1) the basis-cash, accrual, etc., 2) the procedures for payment and budget monitoring, and 3) the compilation of periodic and year-end accounts. Accounts provide the data for the past, whereas the budget contains the present and projected trends.	- Emphases on procedures vary. In general, however, accounting systems are designed to handle the tasks enumerated. During recent years, however, significant leakages have become common. These relate to the exclusion of off-budget accounts, guarantees, contingent liabilities, quasifiscal activities, and tax expenditures. - Periodic reports on the status of government finances are now being released by several countries in conformity with the desire of the international financial institutions to publish standardized fiscal data. Intra-year trends may however, be significantly different from the year-end outcome as developments in the last quarter have a unique pace.
Administrative reports	Fiscal and other agencies are required to issue annual reports illustrating their activities. These agencies include the tax collection machinery.	These documents, which vary enormously in content and coverage, are in the public domain in most countries.
V. Evaluation	Government activities are evaluated by their own agencies to gain experience in the utilization of resources.	Evaluation is limited in most governments and, where conducted, may not always be available to the public.
VI. Audit	There should be adequate arrangements for an independent audit agency to verify the appropriation accounts and carry out accountancy, financial, and performance audits.	Although many countries have independent audit offices, their effective contribution is dimmed by their limited purview (policy matters are excluded) and lack of firm legislative arrangements for a review of the findings.

(Continued)

Table 1 Fiscal transparency: components, instruments, and features (*Continued*)

Functional area	Instruments	Features
VII. Independent standards	Standards are specified for the maintenance of accounts and audit. Similar arrangements exist in some countries (e.g., Sweden) for rating the overall financial management systems, including tax collection machinery.	Accounting standards for government are evolving and, as a result, more governments may have them in the future.
VIII. Legislative review	Submission of the budget, consideration of new policies or changes in the existing ones, and annual accounts provide several opportunities to the legislature to exercise the traditional control of purse.	In practice, the power of the legislatures vary considerably. Even where endowed with requisite powers and instruments, they may not exercise them in view of the party discipline and related procedural limitations.

tedious process into a quick and often cost-effective way of using information for policy making. Notwithstanding the assiduous efforts to improve the coverage and content of fiscal transparency, from the user's point of view, several problem areas remain to be addressed.

In the current situation, approximately one-third of the world's population has little access to the budget details and, thus, to the fiscal policies that have a profound and continuing impact on their lives. People living in centrally planned countries such as China, Vietnam, or quasi monarchies like the Gulf Kingdoms, and similarly placed governments receive few budget details. This is in stark contrast to the former Soviet Republics, which have made a quantum jump, in a short period, to the establishment of fiscal transparency as in other democracies.

The transactions of government and the structure of their portfolios have changed during recent years. In several cases, central governments have become funding and policy agencies, while the operations are conducted by the other levels of government, as well as by non-governmental organizations and the corporate sector. Information in regard to the latter tends to be brief, where provided, and aggregative in nature. To the extent that these shadow governments are not fully covered the usefulness of data on fiscal operations tends to be reduced.

Fiscal transparency does not yet take the community into confidence in the realm of public policy making. In taxation, concerns of secrecy, with regard to defense concerns of security, have contributed to a large degree of opaqueness (North American Treaty Organization countries have begun efforts to provide more information on defense costs). Similarly, little information is available about the effectiveness of the

fiscal machinery (see Ref.^[7], for a discussion of the limitations of audit and legislative control).

The data provided are mostly intended to meet the requirements of the investing class and do not go far in illuminating the fiscal status for the common man. The language of data is generally complex and, without the aid of a professional interpreter, may not be clear to the general public. The community's preference is for analytical data, rather than the voluminous raw data about operations.

Fiscal transparency, unaccompanied by channels of public accountability,^[2] is not likely to be effective. In many cases, however, these channels remain to be effective; thus, more remains to be achieved.

REFERENCES

1. Premchand, A. *Effective Government Accounting*; International Monetary Fund: Washington, 1995.
2. Premchand, A. Public financial accountability. In *Governance, Corruption and Public Financial Management*; Schiavo-Campo, S., Ed.; Asian Development Bank: Manila, 1999; 145–192.
3. Brewer, J. *The Sinews of Power, War, Money and the English State, 1688–1783*; Alfred K. Knopf: New York, 1989.
4. Ferguson, N. *The Cash Nexus: Money and Power in the Modern World, 1700–2000*; Basic Books: New York, 2001.
5. International Monetary Fund. *Code of Good Practices on Fiscal Transparency. Declaration of Principles*; Washington, 1998.
6. International Monetary Fund. *Code of Good Practices on Fiscal Transparency*; Washington, 2001.
7. Premchand, A. *Control of Public Money*; Oxford University Press: New Delhi, 2000.

Foreign Policy Analysis

Eric K. Leonard

Shenandoah University, Winchester, Virginia, U.S.A.

INTRODUCTION

In the 1950s, the field of international politics (IP) split in two. The study of IP became a systemic view of foreign relations, whereas foreign policy analysis (FPA) began the process of “opening up the state” and examining its contents. These two divergent paths examined very different causal factors for global events. However, the overlap between these two fields was always undeniable. This article begins by examining the split between IP and FPA, with a focus on the difference between the traditional theoretical perspectives of these two fields of study. I then turn to the primary goal of this article, which is to show how we can bring these two fields back together. The method of unification is social constructivism and its emphasis on the co-constitutive relationship between agent and structure. If employed as a method of FPA, it may assist scholars in a more complete understanding of the decision-making process.

THE EMERGENCE OF FPA

Prior to the 1950s, IP, as a field of study, was more or less cohesive. Those who studied IP tended to blend the analysis of state relations with an assessment of foreign policy decision making. Although the early study of IP contained its fair share of disagreements (in particular, the realist-idealist debate of the 1930s and 1940s), the field remained unitary. Foreign policy analysis, although clearly a part of our world since at least the rise of the modern state system, did not emerge as an autonomous academic field of study until the early Cold War era. Foreign policy analysis evolved as a part of IP, finding its voice in the classical realist and idealist writings. However, the scientific or behavioralist revolution within the social sciences allowed the study of FPA to branch out on its own.^{[1]a}

Foreign policy analysis “refers to a complex, multilayered process, consisting of the objectives that governments pursue in their relations with other governments and their choice of means to attain these

objectives.”^[2] Clearly, this form of analysis lies within the study of IP. However, during the 1950s, the IP scholars began to focus their attention on “actor-general” theories and turn away from any acceptance of “actor-specific” theories.^[3] The best way to describe the actor-general type of analysis is that it is systemic in nature. This model describes the state as a rational unitary actor which acts in accordance with the national interest. If scholars of international relations are searching for a causal factor for world events, they need to look at systemic factors and not within the black box of the state. In effect, this form of analysis entails the study of like units acting within an anarchical system. What this perspective does not take into account, at least not to any large extent, is the inner workings of the state and its domestic relations.

Realism is the quintessential example of actor-general theory, as one can see by examining its core assumptions. First, realists view the state as the most important actor within the international system. No other actor rivals the importance of the nation-state. Second, they consider the state a unitary, monolithic actor. In other words, the state is considered a “black box” that acts, at least in the international arena, in a unified fashion. The turmoil of domestic politics is not a consideration for realist scholars because the primary causal factor for states’ actions is external or systemic, not internal or domestic. Third, realists consider states rational actors. This “rational hypothesis” allows an IP scholar to view state motives as similar and unitary. This hypothesis also constructs a system of states that always act according to their national interest—which, for realists, is defined by security and relative power. Finally, the anarchical structure of the system is the underlying factor for all actions within the international system. Due to the fact that there is no overarching authority, states will act to maximize their relative power position within the system.^{[4,5]b}

The systemic turn of IP study is what FPA scholars find so unpalatable. Foreign policy analysis scholars believe that if we are to understand the relationship among states, we must first and foremost open up the black box of the state. This entails moving from

^aProvides an excellent description of this split along with a detailed discussion of the divergent models within IP and FPA.

^bMorgenthau is the quintessential classical realist, while Waltz is the definitive text on neorealism and its structural emphasis.

an actor-general model of understanding to an actor-specific model of understanding. Foreign policy analysis scholars began the process of unpacking the black box of the state in the 1950s.^[6] This break from traditional understandings of foreign policy decision making—understandings that are predicated on viewing the state as a rational unitary actor—forever changes the field of FPA.

TRADITIONAL FOREIGN POLICY MODELS

When foreign policy scholars opened up the black box of the state, they discovered a plethora of factors that influence foreign policy decision making. Foreign policy analysis was no longer tied to the static thinking of structural models. Foreign policy analysis scholars could now engage the inner workings of the state and examine their impact on the decision-making process. Although numerous different perspectives emerged from this new form of analysis, this article will examine the three most influential models: bureaucratic politics model, small group context model, and individual decision-making model.^{[7]c}

The statement, “where you stand depends on where you sit,” best describes the bureaucratic politics model. A bureaucratic politics assessment of foreign policy centers on the debate and negotiations that occur between competing bureaucratic agencies. Instead of focusing on the rational decision-making process of the state, bureaucratic politics analysts focus on the self-promotional style of competing organizations. Foreign policy, according to this model, is created from a negotiation process between competing organizations, which are each attempting to implement policies that benefit or highlight their organizational strengths. Thus rational unitary decision making is no longer sufficient for understanding the reality of intergovernmental competition.

This model gained credibility with the publication of Allison’s seminal text, “Essence of Decision: Explaining the Cuban Missile Crisis.” In Allison’s analysis of the crisis, three competing paradigms of decision making are compared. Allison concludes that the first model, the rational actor model, is insufficient due to the simplicity of its conclusions. Allison asserts that the decision-making process during the crisis was much more complex and involved a negotiation process among several competing governmental organizations. Therefore Allison tends to favor the examination of intraorganizational debate and interorganizational

debate (both of which engage an assessment of bureaucratic politics) as a means of FPA.

The small group context model focuses on the inner circle of foreign policy decision makers and the debate, or lack thereof, that arises within the group. Within the realm of American foreign policy, the analysis would center on the presidential cabinet. Most small group context analysts agree that this form of decision making is not the best. According to this model, the sociological dynamics of such a group tends to homogenize the perspectives of its participants. The result, often times, is a lack of policy options due to the desire for personal acceptance and group cohesion.

The seminal text on small group context, or at least its negative consequences, is “Victims of Group Think” by Janis. Janis’ sociological analysis of the dynamics of small group decision making asserted that group consensus was the prime motivator for many foreign policy decisions. This form of political decision making hinders the ability to ascertain good policy because of either the lack, or complete loss, of viable options. The small group becomes homogenized, thus prohibiting any of the positive components that may have resulted from a multiplicity of perspectives. In fact, there is a loss of multiple perspectives to the one dominant group consensus. Janis tested this hypothesis on several foreign policy case studies, including the Bay of Pigs incident, the Cuban Missile Crisis (which actually shows the positive attributes of small group decision making), the escalation of Vietnam, and others.

The individual decision-making model focuses on the primary authority figure within the nation-state—usually the President or Prime Minister. This psychological analysis of the leader depicts the decision-making process as one in which the ultimate ability to create foreign policy resides in one person. If, according to this model, FPA scholars are to understand the decision-making process, then they must examine the individual characteristics of the primary actors. This microlevel analysis would include an examination of personal characteristics (such as leadership style, conflict management, and personal perception/political history), along with the context within which the individual is making one’s decision.

There is not one seminal text for this model of FPA. However, one of the more influential texts is George’s, “The ‘Operational Code’: A Neglected Approach to the Study of Political Leaders and Decision-Makers.” This text describes the importance that a leader’s core political beliefs have on the foreign policy process, thus placing emphasis of foreign policy decision making on the individual. George’s assessment of the role of personal political beliefs, along with management styles within the decision-making process, exemplifies this most microlevel of analysis—the individual.

^cSee Ref.^[3] for a comprehensive discussion of foreign policy models and their historical evolution.

Ultimately, the message of individual-level FPA scholars is that decision making in the world of foreign policy can turn on the perceptions, personality type, or management style of the world's leaders.

What all of these FPA models have in common is a general acceptance that individual actors have the greatest influence on the foreign policy decision-making process. This conclusion completely counters the traditional understandings of FPA that emerged from the field of IP. These traditional understandings focused on the state as a rational unitary actor whose decision making is affected primarily by the systemic construct. But why does an analysis of foreign policy have to exist within this dichotomous relationship? Is the best understanding of foreign policy achieved by engaging in an either/or framework of analysis? Currently, many scholars feel that there is a way to incorporate both IP and FPA understandings of decision making into their analysis. The link between these two divergent understandings is social constructivism.^{[8]d}

SOCIAL CONSTRUCTIVISM AND FPA

Social constructivism rests on the premise that the world around us is irrevocably social, and that an examination of global social relations will provide us with a more complete understanding of IP and, in the case of this article, the foreign policy decision-making process. According to this analytical perspective, the world exists in its current form because we—agents in the world—have made it that way. Such a perspective is very amenable to the FPA scholars and their models of decision making because the focus rests on the agents involved in the process. However, social constructivism does not view the decision-making process as a one-way street. Yes, agents do make the world, but conversely, social relations also make agents.

Thus social constructivists also acknowledge the causal factors of the traditional IP literature. This complex form of analysis provides us with a co-constitutive model of foreign policy, with both agents and structures making each other. In general, social constructivists do not privilege one perspective over another, but recognize the complexity of decision making in world politics, thus acknowledging the importance of both actor-general and actor-specific perspectives. It

is the co-constitutive relationship of these two perspectives that social constructivists want to examine.^{[9,10]e}

When thinking about foreign policy, the first attribute that social constructivists purport is that no concept, term, institution, or agent of foreign affairs is static. Because of the socially constructed nature of the world, the reproduction or reconstruction of everything is possible. This reconstruction simply depends on the interaction between the relevant agents and social arrangements, and, more importantly, the rules that guide those agents and social arrangements. Therefore we cannot think of the defining concepts of world politics/foreign policy as static institutions or discourses.

Second, social constructivists agree that foreign policy decision making does not come about through one causal factor, as traditional FPA or IP scholars would have us believe. Instead, decision making results from the interplay between actors (which include more than just states and individual leaders) and context. Social constructivists privilege neither factor. However, what social constructivists do privilege is the social relationship between the two (agent and structure), which causes a certain decision-making process to occur.

One problem that arises from this rich and complex understanding of foreign affairs is qualifying the relationship that exists between agent and structure. If social constructivists hope to study foreign policy from both agent-general and agent-specific perspectives, how do we link the two forms of understanding? The link occurs in the form of action and interaction. Social constructivists do not accept that agents/actors and or context/environment are simple and static. Both agents/actors and structure/context are malleable entities, and the change that may or may not occur in both of these entities comes about via action and interaction. This is, in fact, the making of our world, or, in the case of FPA scholars, the making of foreign policy. An example from the current global context may help to clarify this approach.

The current situation in Iraq—and, in particular, the United States' decision to invade—is a hot topic of conversation among FPA scholars. How does one understand the motives behind the U.S. invasion? Was this action generated primarily by the agents involved, or were these agents simply acting according to the current structure? A social constructivist would answer yes to both questions. How does one understand the decision to invade Iraq if you disregard either the actors involved or the structure? The fact that George W. Bush is President of the United States is crucial to understanding the decision to invade. His boisterous, do-it-alone (unilateralist) attitude is a primary causal factor in the decision-making process, as is the individuals that he has surrounded himself with. However, an analysis of the President's personal

^dKubalkova's "Foreign Policy in a Constructed World" is one text that attempts to bring together these two divergent areas of policy study by engaging in constructivist ontology.

^e"Constructivism: A User's Manual" and "Making and Remaking the World for IR 101: A Resource for Teaching Social Constructivism in Introductory Classes" provide two excellent introductory sources to the study of world politics through a social constructivist perspective.

perceptions, management style, etc., along with an analysis of the aforementioned traits of the presidential cabinet, are not sufficient to understand the decision to invade. One major question looms: would the United States have invaded Iraq if 9/11 had never happened? I would argue no, thus showing the importance of context and its impact on the agents involved. 9/11 made the foreign policy decision to invade possible, but so did the election of George W. Bush. The crucial area to study for FPA scholars is not simply the context nor the individuals, but the interaction between these two. Social constructivism may be more complex than previous FPA approaches, but it is also more complete in its assessment of foreign policy.^{[11]f}

CONCLUSION

So where is the field of FPA headed in the coming years?^{[12]g} This article has attempted to show both the divergent positions within the study of FPA and a new form of analysis that might serve as a bridge within the field. With that said, the future of FPA may or may not find its theoretical basis in social constructivism. Constructivism may not solve the long-standing debate between traditional FPA and IP scholars, along with the subdebates contained within FPA. However, if FPA scholars are willing to engage in a social constructivist assessment of foreign affairs, they may find the bridge that brings these seemingly opposed fields of study back together. As this form of analysis grows in acceptance, I believe that its applicability to multiple fields of study will grow. Nevertheless, ontological shifts are slow, but constructivists can serve as a repairperson for the IR/FPA split, if the analysts allow it. The result would only benefit the field and our understanding of foreign policy decision making.

REFERENCES

1. Holsti, O.R. Models of international relations and foreign policy. In *Diplomatic History*; 1989, 13(1), 15–43.

2. Kubalkova, V. Foreign policy, international politics, and constructivism. In *Foreign Policy in a Constructed World*; Kubalkova, V., Ed.; M. E. Sharpe: Armonk, NY, 2001; 15–37.
3. Hudson, V.M. Foreign policy analysis yesterday, today, and tomorrow. *Mershon Int. Stud. Rev.* **1995**, 39 (1), 209–238.
4. Morgenthau, H. *Politics Among Nations*; Alfred Knopf: New York, NY, 1947.
5. Waltz, K.N. *Theory of International Politics*; Addison-Wesley: Reading, MA, 1979.
6. Snyder, R.C.; Bruck, H.W.; Sapin, B. *Decision Making as an Approach to the Study of International Politics. Reprinted in Foreign Policy Decision Making: An Approach to the Study of International Politics*; Free Press: New York, NY, 1962.
7. Hudson, V.M. *Foreign Policy Analysis*; Encyclopedia of Policy Studies, 1994; 281–304.
8. Kubalkova, V., Ed.; *Foreign Policy in a Constructed World*; M. E. Sharpe: Armonk, NY, 2001.
9. Onuf, N. Constructivism: a user's manual. In *International Relations in a Constructed World*; Kubalkova, V., Onuf, N., Kowert, P., Eds.; M. E. Sharpe: Armonk, NY, 1998; 58–78.
10. Ba, A.; Hoffmann, M.J. Making and remaking the world for IR 101: a resource for teaching social constructivism in introductory classes. *Int. Stud. Perspect.* **2003**, 4, 15–33.
11. Campbell, D. *Writing Security: United States Foreign Policy and the Politics of Identity*, revised edition. Minnesota University Press: Minneapolis, MN, 1998.
12. Karbo, J. Foreign policy analysis in the twenty-first century: back to comparison, forward to identity and ideas. *Int. Stud. Rev.* **2003**, 5 (2), 156–163.

BIBLIOGRAPHY

1. Allison, G.T. *Essence of Decision: Decision Making and the Cuban Missile Crisis*; Little, Brown, and Co.: Boston, MA, 1971.
2. George, A.L. The 'Operational Code': a neglected approach to the study of political leaders and decision-makers. *Int. Stud. Q.* **1969**, 13, 190–222.
3. Janis, I. *Victims of Group Think*; Houghton Mifflin Company: Boston, MA, 1972.

^f“Writing Security: United States Foreign Policy and the Politics of Identity” is an excellent example of constructivist work and its relation to foreign affairs.

^g“Foreign Policy Analysis in the Twenty-First Century: Back to Comparison, Forward to Identity and Ideas” provides a good assessment of probable future forms of study.

Forms of Government

Victor S. DeSantis

*Department of Political Science, Bridgewater State College,
Bridgewater, Massachusetts, U.S.A.*

INTRODUCTION

The legal relationships between states and local governments are extremely complex and vary widely from state to state. Local governments, herein used to describe municipalities and counties, are legally considered to be creatures and agents of state government and, as such, subordinate to that state. All power and autonomy that local governments possess are granted to them through state action, and that power can be enhanced and rescinded. This legal principle is referred to as Dillon's Rule and dates back to the late 1800s.

State law, as well as a region's political culture and traditions, affects the selection of a local government's organizational structure. Although three levels of state law establish the parameters of local government autonomy—the state constitution, the state legislature (in the form of statutes), and state charters—many forms of government issues are specifically addressed in state charters. Charters are the legal mechanism that states use to recognize the existence of local governments and generally provide detail on the local government organization, powers, and responsibilities. Charters may be proffered by legislation and chosen by the jurisdiction or, in the case of home-rule jurisdictions, drafted and approved locally by citizens. Much like a state constitution, charters enumerate the basic purposes and tasks of local governments, the selection and powers of local officials, the authority and process for passing local ordinances or bylaws and, finally, the basic structures or forms of municipal and county government.

MUNICIPAL FORMS OF GOVERNMENT

There are five general forms of municipal government found in the United States today. These are the mayor–council, council–manager, commission, town meeting, and representative town meeting forms. Although each form retains distinct structural characteristics, it appears that there has been a modest, but consistent, movement since the mid-1980s toward council–manager forms of government, and away from mayor–council and commission forms. Specifically, research indicates that there has been a general

convergence of the different forms over the past several decades that is especially apparent between the mayor–council and council–manager systems.^[1–3] In addition to the overall movement toward the council–manager model, other less noticeable but important changes in the structures of local government have also surfaced.^[4,5] First, an increasing percentage of communities with mayor–council or town meeting forms have hired chief administrative officers to assist with the day-to-day operations of the government. Second, these chief administrative officers have increased their budgetary and appointment powers relative to the elected officials. Finally, a decreasing percentage of municipalities report giving their mayor a veto power over council actions.

The 1996 national Municipal Form of Government survey, conducted by the International City/County Management Association (ICMA), indicated that the majority of American cities operate under the mayor–council or council–manager form. The 1996 ICMA survey data show that 51.9% of responding communities used the council–manager form, whereas another 31.9% used the mayor–council form. The remaining survey respondents used the commission, town meeting, or representative town meeting forms, or were unsure of their specific categorization.^[4]

Mayor–Council

The mayor–council form of government (Fig. 1) is most prevalent in the nation's larger cities and smaller jurisdictions. In this form, there is usually a legislative body, the city/town council, and a separately elected chief executive, the mayor. The actual titles afforded to the legislative body and executive official vary widely and are usually based on long-standing local or regional tradition. There are generally two types of mayor–council government, the “strong mayor” and “weak mayor” models. The substantive difference between these types depends on the administrative duties and power given to the mayor.

The strong mayor form of government centralizes the administrative powers in the hands of the popularly elected mayor, while vesting the legislative function in the elected council. In this form, the mayor

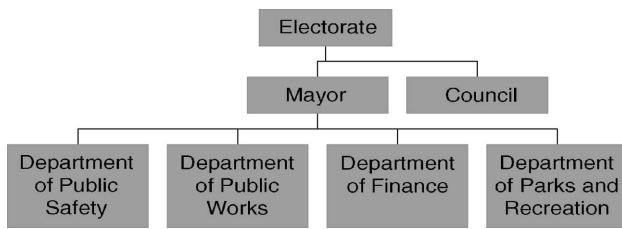


Fig. 1 Model of the typical mayor-council form of government.

has the authority to prepare and control the administration of the budget, appoint and remove department heads, preside over council meetings, and set the council agenda. The strong mayor may also have a veto power over council actions. Supporters of a strong mayor government argue that a strong mayor provides leadership to a community while being held responsible to citizens for their actions. Also, a strong mayor may be better able to exercise political power in bigger cities to bring together large and diverse interest groups.

Increasingly, strong mayors are also opting to appoint a chief administrative officer for the city who serves at their pleasure to focus on the details of day-to-day government administration, such as budget preparation and personnel management. This arrangement frees mayors to concentrate more on their roles out of city hall and in the community, and also increases their ability to deal on policy matters and agenda setting.

The weak mayor system is the older of the two mayor models, resulting from the traditional American distrust of strong executive power. The weak mayor is often selected from among the council members, instead of being elected directly by the people. The weak mayor generally has limited appointment and budgetary authority and does not have veto power over council. This system is also characterized by its decentralized and fragmented nature, with multiple department heads and other local officials directly elected by the community. The weak mayor form of government is the more commonly used in the United States, especially in smaller communities.

The dawn of the twentieth century was an important watershed period for municipal government in the United States, most specifically for the traditional mayor council model of government. Across the nation, political bosses and machine politics were the norm in many larger cities. Although the newly formed National Municipal League first supported the strong mayor form in 1898 in an effort to decrease the power of political machine bosses, this effort was short lived. Further efforts at reforming the structure and process of local government in reaction to the widespread

corruption and inefficiency surfaced during this period, which became popularly known as the Progressive Era. At the core of this “good government” movement were the upper middle-class residents that had grown tired of the powerful political machines and the immigrant groups keeping them in power.^[6] Their specific agenda included numerous election reforms and the promotion of more efficient and professional government through civil service systems, competitive bidding practices, and forms of municipal government that provided for an apolitical appointed (rather than elected) executive. These reform efforts led to experiments with the commission and council-manager forms of government.

Commission

Cities and towns with the commission form of government provide for a full centralization of power in the elected council to make policy and direct the executive operations of the local government. Each elected commissioner serves as a member of the legislative body and is also elected to lead one or more administrative departments. This form continues to decline and is currently found in only a few communities nationwide.

In 1900, in response to a disastrous hurricane and flood that crippled the city of Galveston, the Texas governor appointed a set of five businessmen (later elected by the community) to handle the policy and administrative functions of the community. Supporters of this model believed that this plan was a positive step toward improving local government because it unified all power in the hands of a board of commissioners. Many of the good government reformers, themselves from the business community, began promoting the commission form, and it was adopted widely over the next decade.^[7] However, one of the weaknesses of the commission form that began to surface was that it fragmented administrative functions and was too unwieldy. Reformers later latched on to the idea of the council-manager form, a hybrid of the strong mayor and commission forms of government.

Council-Manager

Another form of municipal government that grew out of the Progressive Era is the council-manager form (Fig. 2). Council-manager government combines the strong political leadership of elected officials with the strong managerial leadership of an appointed professional manager. This form of government rests the policy-making power in the hands of the elected legislative body. A professional manager serving at the pleasure of the elected body is charged with day-to-day

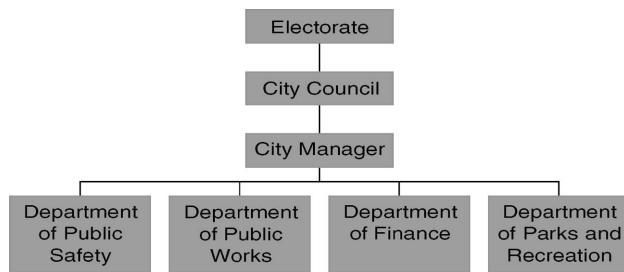


Fig. 2 Model of the typical council–manager form of government.

administration of the government. The manager commonly has the authority to direct administrative operations, appoint administrative officers, prepare and administer the budget, and advise the council on matters of policy.

Although many council–manager forms of government have a popularly elected mayor, many of the mayors in this form are selected from among the membership of the legislative body. These mayors have restricted powers that are generally limited to presiding over council meetings and making appointments to boards or commissions. Mayors are part of the council and vote as regular members, although they have little or no veto power. Proponents of the council–manager form of government argue that this structure allows supervisory and administrative responsibility to be centralized in one individual. This allows the individual's expertise and knowledge of administrative activities to be developed, while vesting all power in an elected governing body to promote representative democracy.

The council–manager form is actually a hybrid of two previous government structures: the strong mayor and commission. In an effort to improve on the decentralized and fragmented commission form, Richard S. Childs and other local government reformers proposed a plan that combined the commission form with a strong administrative element, concentrating all administrative powers in the form of a single professional manager. This became the council–manager plan and, in 1915, the National Municipal League revised its model city charter to reflect this new form of government.^[7,8]

Town Meetings

The town meetings of New England are often considered to be one of the last modern-day examples of direct democracy in action. This distinction comes from the mass public meetings, in which any registered voter of the jurisdiction may attend, speak, and vote on

various policy, administrative, and budgetary issues. The town meeting process as practiced in New England can be traced back to the Massachusetts Bay Colony of the late 1620s^[9] and is currently found only in the New England states of Maine, Massachusetts, New Hampshire, Vermont, Rhode Island, and Connecticut. However, in recent decades, the prevalence and operation of town meeting forms of government have experienced substantial tension and change.

The limited empirical evidence on town meetings indicates a growing trend toward abandoning open town meetings in favor of an elected town council or establishing a representative town meeting government.^[10] The representative town meeting involves the voters selecting a limited number of citizens to represent them and vote at town meetings. All citizens may attend and participate in the debate, but only the duly elected representatives may actually vote. Other New England communities have adopted a more incremental change by hiring full-time professional chief administrative officers, in addition to the open town meeting process.

Under the town meeting form of government, the town meeting is considered the local legislative body with the authority to pass ordinances and bylaws, and raise and appropriate funds. Each town with this form generally holds one annual town meeting and may also hold as many additional meetings (called special town meetings) as deemed necessary to carry out the town's business. A warrant (from the term “warning”) is issued to townspeople stating the date, time, place, and itemized agenda for annual and special town meetings. An elected moderator presides over the town meeting and has wide latitude in directing the debate and conducting business.

During the remainder of the year, the elected board (usually called selectmen) of three or five members is charged with running the daily operations of the town. Boards of selectmen carry out many appointing, financial, and licensing responsibilities granted to them by the town meeting. Many towns have added the position of chief administrative officer to assist the board in their administrative roles on an ongoing basis. These chief administrative officers serve at the pleasure of the board of selectmen and have their duties prescribed by board action or a town meeting bylaw.

COUNTY FORMS OF GOVERNMENT

As units of local government, counties have historically been considered administrative arms of state government and operate with more constraints than their municipal counterparts. In most areas of the United States, this means that counties operate with

less local autonomy or “home rule” authority and have their operations and structures dictated substantially by state law. There are three general forms of county government found in the United States today. These are the commission, council–administrator, and council–elected executive forms. Although the reform movement for county government came later in the twentieth century than the municipal reform movement, there has been modest improvement, especially over the last few decades. Many of these efforts have been aimed at improving and streamlining county operations and providing for greater professional leadership. Research points to increased usage of the council–administrator and council–elected executive forms. Indeed, the overall trend toward professional management in counties can be seen by the ever-increasing number of counties, even with the traditional form, that provide for a position of chief administrative officer.^[11]

Commission

The commission or plural executive form of government in counties is the oldest and most traditional organizational structure. A central governing board characterizes it with members usually elected by district. Although various names exist for this board, among the favorite are the board of commissioners or supervisors and the county court. Most often, the board selects one of its members as the presiding officer. Members of the governing board may act as department heads, but this is not a defining characteristic as in the municipal commission model. The governing boards share administrative and, to an extent, legislative functions with independently elected officials: the clerk, the treasurer, the sheriff, the assessor, the coroner, and others. Generally, no single administrator oversees the county’s operations. In some counties with the commission form, the structure includes an official (generally full time), such as the county judge, who is independently elected at-large to be the presiding officer of the governing board. As noted previously, many disadvantages associated with the commission form stem from the usual lack of a chief administrator to provide more professionalism, executive leadership, and accountability.

Council–Administrator

The council–administrator form of government for counties is similar to the council–manager form for cities, but three distinct variations are identifiable. In its strongest variation, the council–administrator form

provides for an elected county board or council and an appointed administrator. The county board adopts ordinances and resolutions, adopts the budget, and sets policy. The administrator, appointed by the board, has responsibility for budget development and implementation, the hiring and firing of department heads, and recommending policy to the board. In some counties, where a weaker version of the council administrator plan is in place, the administrator usually has less direct responsibility for overall county operations and less authority in hiring and firing, and may consult with the board on policy issues.

Council–Elected Executive

The third form of county government is the council–elected executive form. This system has two branches of government—legislative and executive—and more clearly resembles the strong mayor form of city government. Here, the county council or board assumes responsibility for county policies, adopts the budget, and audits the financial performance of the county. The elected at-large executive is considered the chief elected official of the county and often has veto power, which can be overridden by the council. The executive prepares the budget, carries out the administration of the county operations, appoints department heads (usually with the consent of the council), and suggests policy to the governing board. In addition, this official carries out appropriations, ordinances, and resolutions passed by the board and generally acts as the chief spokesperson for the county. When the executive is considered the chief political spokesperson, the executive often delegates the administrative responsibility for the daily county operations to a chief administrator. This form, along with the council–administrator form, has begun to receive more popular support over the past few decades and push counties forward in their quest for modernization.

REFERENCES

1. Renner, T. Elected executives: authority and responsibility. In *Baseline Data Report*; International City Management Association: Washington, DC, 1988; Vol. 20 (3).
2. Boynton, R.P.; DeSantis, V.S. Form and adaptation: a study of the formal and informal functions of Mayors, Managers, and Chief Administrative officers. In *Baseline Data Report*; International City Management Association: Washington, DC, 1990; Vol. 22 (1).
3. Frederickson, H.G.; Johnson, G.A. The adapted American city: a study of institutional dynamics. *Urban Aff. Rev.* **2001**, *36*, 872–884, (July).

4. Renner, T.; DeSantis, V.S. Municipal forms of government: issues and trends. In *The Municipal Year Book 1998*; International City/County Management Association: Washington, DC, 1998.
5. Hansell, W. Is it time to reform the reform. *Public Manage.* **1998**, *80*, 15–16, (December).
6. Renner, T.; DeSantis, V.S. Contemporary patterns and trends in municipal government structures. In *The Municipal Year Book 1993*; International City/County Management Association: Washington, DC, 1993.
7. Nolting, O. *Progress and Impact of the Council-Manager Plan*; Public Administration Service: Chicago, IL, 1969.
8. Stillman, R.J. *The Rise of the City Manager*; University of New Mexico Press: Albuquerque, NM, 1974.
9. Zimmerman, J.F. The New England town meeting: pure democracy in action? In *The Municipal Year Book 1984*; International City Management Association: Washington, DC, 1984.
10. DeSantis, V. *Open Town Meetings in Massachusetts: Issues and Evaluation*; Institute for Regional Development: Bridgewater, MA, December 1997; Vol. 97 (2).
11. DeSantis, V. County government: a century of change. In *The Municipal Year Book 1989*; International City Management Association: Washington, DC, 1989.

Freedom of Information Act—Federal

Suzanne J. Piotrowski

Graduate Department of Public Administration, Rutgers University,
Newark, New Jersey, U.S.A.

INTRODUCTION

The Freedom of Information Act (FOIA) is the premier open government statute in the U.S. federal government. FOIA works to ensure a more transparent federal administration. The openness afforded by FOIA has become a central aspect of federal administration. After a long debate, FOIA was passed by Congress in 1966 and has been amended periodically. FOIA allows individuals or organizations to request documents from a federal agency. While not all documents are releasable, many are. Federal agencies may withhold documents that fall into nine broad exemption categories. Included within this entry is an overview of the theoretical underpinnings of the act, a legislative history of FOIA, and a short description of the major provisions of the act.

WHAT IS FOIA?

The Freedom of Information Act is one federal law that deals with open government. Other open government federal laws include the Privacy Act (1974), the Government in the Sunshine Act (1976), and the Federal Advisory Committee Act (1972). Implementation of FOIA is a major administrative function which cost over \$270 million in fiscal year 2001. Open government laws are not unique to the federal government. All 50 states have laws governing the release of government documents. Freedom of information and the free exchange of information is a global concern. Over 50 countries ranging from Albania to Zimbabwe have laws facilitating access to government documents. Over half of these international freedom of information laws were passed within the last 10 years.^[1]

Governmental transparency and democratic accountability are the underlying tenets behind FOIA. Governmental transparency refers to the ability to find out what is going on inside of government. Democratic accountability is holding elected and unelected government officials responsible for their actions. While the initial Freedom of Information Act was passed in 1966,^[2] the debate surrounding transparency in

government long predates the act. James Madison wrote in a personal correspondence:

“A popular Government, without popular information, or the means of acquiring it, is but a Prologue to a Farce or a Tragedy; or, perhaps, both. Knowledge will forever govern ignorance: And a people who mean to be their own Governors must arm themselves with the power which knowledge gives.”^[3]

Madison’s sentiments were included in the *Senate Report*, which accompanied the initial FOIA.^[4]

The idea of opening up the government and holding officials accountable is an essential aspect of democracy. Francis Rourke^[5] made this point eloquently: “Nothing could be more axiomatic for a democracy than the principle of exposing the processes of government to relentless public criticism and scrutiny.” Policies which work toward a more transparent government combat administrative and executive secrecy. As explained later in the section “How FOIA Works?” not all types of documents are releasable through FOIA requests; however, at its core, FOIA is a disclosure statute.

FOIA also has two other provisions. The first requires agencies to disclose information automatically by publishing it in the *Federal Register*. Information that must be disclosed includes descriptions of agency organizations, functions, and procedures; substantive agency rules; and statements of general agency policy. The second requirement is the reading room provision of FOIA. Final agency opinions and orders rendered in the adjudication of cases, specific policy statements, certain administrative staff manuals, and some records previously processed for disclosure must be made available in agency reading rooms. Some of these disclosed documents must also be posted in an agency’s electronic reading room. Failure to disclose some types of information precludes agencies from enforcing or relying on them.

LEGISLATIVE HISTORY

Varying proposed pieces of freedom of information legislation were prepared and debated in Congress

for over a decade before the final bill was passed. An early advocate of freedom of information laws, Senator Thomas C. Hennings, Jr. (D-Missouri), referring to a proposed freedom of information legislation, noted that: “The aim . . . of this bill . . . is to make it clear beyond any doubt that the basic purpose of this section is to insure the dissemination of the maximum amount of information reasonably possible.”^[6] FOIA, which amended the Administrative Procedure Act of 1946, was passed on July 4, 1966 and went into effect in 1967.

Since the initial law was passed, a series of amendments have modified the statute. Some changes strengthened the statute while others weakened the reach of the law with regard to access of government. The first series of major amendments came in 1974 in the wake of the presidential Watergate scandal. These amendments significantly strengthen the reach of FOIA disclosures. A little more than a decade later in 1986, another series of amendments were passed by Congress during the Reagan administration. These changes weakened access to government afforded through FOIA by expanding the exemption for law-enforcement documents.

The last major set of amendments came in 1996 and are referred to, collectively, as the Electronic Freedom of Information Act or E-FOIA. These amendments brought FOIA into the Internet age. Electronic documents are now accessible through FOIA requests, and agencies are obligated to proactively post their most frequently requested documents in their electronic reading rooms. In 2002, the Homeland Security Act was passed which also included an amendment to FOIA. Senator Patrick Leahy (D-Vermont) called this amendment the “most severe weakening of the Freedom of Information Act in its 36-year history.”^[7] The Homeland Security Act amendment exempts all information deemed related to homeland security which corporations voluntarily disclose to the Department of Homeland Security. This is a very broad exemption that ensures the department will keep the disclosed information secret and allows companies to be free of potential civil liability and antitrust lawsuits if the information indicates wrongdoing.

HOW DOES FOIA WORK?

Essentially, FOIA is a disclosure statute. All agency records are accessible to the public unless specifically exempt from disclosure. Documents, not information, are requested and released under FOIA. This is a meaningful distinction. One cannot request the federal government to make a report or compile and summarize information. Under FOIA, one can only request preexisting agency records. There is no clear definition

of what is an agency record. Typically, documents must be in control of the agency from which they are requested. After the 1996 E-FOIA amendment, electronic files may be requested and will be released in electronic form. Prior to the 1996 amendments, databases released to requesters would usually have been printed and sent in hard-copy form. Currently, requesters may receive data sets electronically in the electronic format of their choosing. Generally, e-mails are considered agency records although the line between agency and personal records is at times blurred. For a comprehensive discussion of relevant case law, see *Litigation Under the Federal Open Government Laws 2002: Covering the Freedom of Information Act, the Privacy Act, and the Government in the Sunshine Act, and the Federal Advisory Committee Act*.^[8]

Virtually, anyone can make a FOIA request including U.S. citizens, foreign citizens, corporations, and governments. FOIA is a disclosure statute, meaning it opens up documents in the federal government for release. With that said, there are nine exemptions written into FOIA:

1. National security information.
2. Internal agency rules.
3. Information exempted by other statutes.
4. Business information.
5. Inter- and intra-agency memoranda.
6. Personal privacy.
7. Law-enforcement records.
8. Records of financial institutions.
9. Oil well data.

Most of these exemptions are discretionary; that is, if a document falls under one of the exemption categories, then the federal agency has the option whether to release the document or not. However, the first exemption is not discretionary. Documents that have been classified for national defense or foreign policy are not releasable. Documents that are properly classified are not appropriate for a discretionary disclosure. The second exemption refers to documents which relate to internal personnel rules and practices. Documents falling under exemption 2 may be discretionarily disclosed. The third exemption is one of the broadest. If documents are exempt from disclosure through another statute, then they are also exempt under FOIA and cannot be discretionarily disclosed. Examples of statutes that exempt the release of material include the Department of Homeland Security Act, the Immigration and Naturalization Act, the Federal Aviation Act of 1958, and the Ethics in Government Act.

Trade secrets and business information are exempt from release through FOIA by the fourth exemption and rarely are discretionarily disclosed. At times,

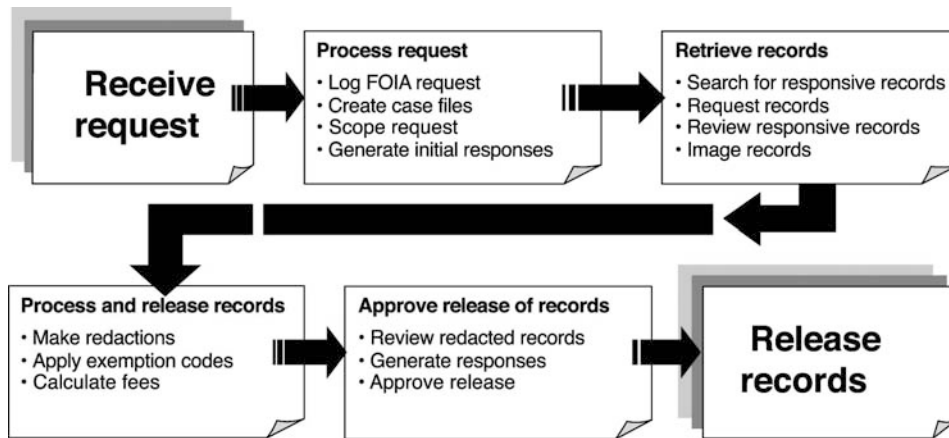


Fig. 1 Overview of generic FOIA process. (U.S. General Accounting Office (2002). *Information Management: Update on Implementation of the 1996 Electronic Freedom of Information Act Amendments* [GAO-02-493]. Washington, DC: U.S. General Accounting Office, p. 5.)

corporations or businesses seek to prevent the release of information gathered by an agency to a third party. These challenges are referred to as “reverse” FOIA litigation. Documents such as inter- or intra-agency memos regarding predecisional policy formation are covered under exemption 5. Documents falling under exemption 5 may be discretionarily disclosed. Individual’s personnel or medical files are not releasable because of the privacy exemption, number six, and generally not deemed appropriate for discretionary disclosures. The law-enforcement provisions, which were strengthened by the 1986 amendment, are covered under exemption 7. Records dealing with financial institutions and oil wells are covered by exemptions 8 and 9, respectively. Documents falling under exemption 7, 8, or 9 may be discretionarily disclosed.

To make a FOIA request, one must send a written request to the individual agency you would like to release the documents. This can be carried out either by the mail or, increasingly, by fax or e-mail. It is essential that you write to the correct agency and you are as specific as possible about your request. There is a wealth of good references on how to write a FOIA request. The U.S. House of Representative Committee on Government Reforms publishes *A Citizen’s Guide to Using the Freedom of Information Act and the Privacy Act of 1974 to Request Government Records*. This report is linked to the committee’s homepage: <http://www.house.gov/reform>. A summary of the steps a FOIA office takes to fill a request are presented in Fig. 1.

However, not all requests are filled. If a request is denied, either in full or in part, the requester has the right to appeal. Frequently, appeals are settled informally by the agency and never make their way to court. Judicial remedy is the last recourse, however.

By law, agencies are required to make a determination on a FOIA request within 20 working days. Agencies are also required to make a determination on administrative appeals within 20 working days of

receipt of the appeal. Many agencies do not meet these requirements and some have construed the 20 days as the time frame within which the request needs to be acknowledged. This was not the intent of the law. Agencies frequently do not meet the statutorily prescribed timelines and subsequently have large backlogs of unfulfilled requests. Some agencies have requests that are years old. The 1996 E-FOIA amendments attempted to address this issue by setting up a multi-track processing system. Instead of a first-come first-serve basis, they now have a system where simple requests are filled within the order they are received and complex requests are put into a different queue.

Agencies can charge requesters for the direct costs associated with searching for documents, the direct costs associated with reviewing documents to determine which portions are releasable, and the duplication costs. Depending on the category of requester, all, some, or none of these fees may be charged. Review, search, and copy charges are set by the Office of Management and Budget. Certain categories of requesters can apply and receive fee waivers. Individuals or organizations that fall under the category of public interest groups, representatives of the news media, or educational or non-commercial scientific organizations may receive these waivers. Fee waivers are also granted to any requester who can show that the disclosure of the information is in the public interest and will likely contribute to public understanding of governmental operations and activities.

CONCLUSION

Implementation of the Freedom of Information Act is a complex process. Inherent in this process are multiple tensions. The first tension is between efficiency and transparency. FOIA is expensive and FOIA offices are periodically underfunded. FOIA functions are not associated with the missions of most federal agencies.

When budgets are tight, it is essential to continue funding non-mission-based functions such as FOIA.^[9] Budgetary constraints are a major impediment to FOIA implementation. Statutorily FOIA requests must be filled and lack of funding does not absolve this responsibility. Most agencies have significant backlogs of unfilled requests and requests may take years to be filled. Requesters have little recourse for delayed filling of requests other than litigation.

The tension between privacy and transparency is perennial. Privacy advocates consistently want to decrease the amount of information that can be released with regard to personal privacy, while other groups of people, such as journalists, may push for greater release of personal information. Other areas of tension include national security, homeland security, and business information. These tensions are inherent in any freedom of information law and are most likely never going to be resolved. The need to periodically strike new balances among FOIA's major concerns is not necessarily a problem. Competing interests will help shape the nature of future Freedom of Information Act amendments.

REFERENCES

1. Banisar, D. *The www.freedominfo.org Global Survey: Freedom of Information and Access to Government Record Laws Around the World*; The National Security Archive, The George Washington University: Washington, DC, September 28, 2003. <http://www.freedominfo.org/survey/survey2003.pdf> downloaded September 30, 2003.
2. Freedom of Information Act 1966 (as amended 1974, 1986, 1996).
3. Madison, J. Letter to William T. Barry, August 4, 1822. In *James Madison: Writings*; Literary Classics of the United States: New York, NY, 1999; 790 pp.
4. *Freedom of Information Act and Amendment of 1974 Source Book: Legislative History, Texts, and Other Documents*; U.S. Government Printing Office: Washington, DC, 1975.
5. Rourke, F.E. Administrative secrecy: a congressional dilemma. *Am. Polit. Sci. Rev.* **1960**, *54* (3), 691.
6. Braverman, B.A.; Chetwynd, F.J. *Information Law: Freedom of Information, Privacy, Open Meetings, Other Access Law (Books 1 and 2)*; Practising Law Institute: New York, 1985; 12 pp.
7. Dalgish, L.A., LaFleur, J., Leslie, G.P., Eds.; *Homefront Confidential: How the War on Terrorism Affects Access to Information and the Public's Right to Know*, 4th Ed.; Reporters Committee for Freedom of the Press: Arlington, VA, September, 2003; 67. March.
8. Hammitt, H.A., Sobel, D.L., Zaid, M.S., Eds.; *Litigation Under the Federal Open Government Laws 2002: Covering the Freedom of Information Act, the Privacy Act, and the Government in the Sunshine Act, and the Federal Advisory Committee Act*; Epic Publications: Washington, DC, 2002.
9. Piotrowski, S.J.; Rosenbloom, D.H. Nonmission-based values in results-oriented public management: the case of freedom of information. *Public Admin. Rev.* **2002**, *62*, 6. November/December.

Futures and Options

Greg Chen

*School of Public Affairs, Baruch College, City University of New York,
New York, New York, U.S.A.*

INTRODUCTION

Like stocks and bonds, futures and options are financial instruments that are traded in exchange markets across the world. Unlike stocks and bonds, however, futures and options are derivatives whose value depends on the values of their underlying assets.

Futures and options serve important functions in the financial market. They provide the opportunity for hedging and speculation. Futures and options are a growing field in financial market. This is mainly due to the increasing volatilities in financial and commodities markets over the past decades.

This paper discusses the concepts and practices of futures and options in the financial market. It also introduces briefly the basic principles and the key models in option pricing, i.e., the binomial model and the closed form Black and Scholes model.

FUTURES AND FUTURES MARKETS

Futures are exchange-traded derivatives, representing agreements to buy or sell some underlying assets in the future for specified prices. Futures trade on a variety of underlying financial assets and commodities, including stocks, stock index, foreign exchange, gold, wheat, oil, live beef, etc.

Futures markets have a long history. They were initially developed to meet the need of farmers and merchants. The Chicago Mercantile Exchange (CME) was established in 1919. It is the most significant futures market in the world. Futures are also traded in many exchanges in many countries. Among the major ones are the London International Financial Futures Exchange, the Tokyo International Financial Future Exchange, Bolsa de Mercadorias Futuros in Sao Paulo, and the Swiss Options and Financial Futures Exchange, to list just a few.

The futures market in United States is regulated by the Commodity Futures Trading Commission (CFTC), an independent agency of the U.S. government. The market is also subject to regulation by the National Futures Association (NFA), a self-regulatory body authorized by the U.S. Congress and subject to CFTC supervision.

The futures exchange has a clearinghouse function. The clearinghouse is responsible for clearing trades

and for day-to-day settlement. This means the clearinghouse records all of the trades made through the exchange each day. At the end of the trading session, the clearinghouse matches or reconciles contracts bought and sold.

The participants in the futures market are mostly hedgers and speculators. A hedger tries to minimize risk by buying or selling now in an effort to avoid rising or declining prices in the future. Conversely, the speculator tries to profit from the risks by buying or selling now in anticipation of rising or declining prices.

The gains and losses in a futures contract are settled daily by a process called marking to the market. The broker transfers that amount from the trader's margin account to the clearinghouse when faced with loss. The clearinghouse transfers that amount to the broker, who then deposits it into the margin account, when there is a gain. Through this daily margining process, futures settle every day. This means that obligations under the futures contract are satisfied every day on an ongoing basis as mark-to-market profits or losses are realized. This essentially eliminates credit risk for futures.

It should be noted that futures trading is highly leveraged. This means that the profit gains and losses are large in comparison to the change in the underlying assets and relative to the initial margin that the traders deposit. For example, suppose the futures prices for S&P 500 is \$1100. Given that each S&P 500 is a multiple of \$250, the nominal value of a contract is $250 \times 1100 = \$275,000$ dollars. As normally requested, a 10% margin is deposited for the contract. Assuming that the next day, S&P 500 drops just by 5% to 1045, the long position loss is, however, $(1100 - 1045) \times 250 = \$13,750$, representing a 50% loss of the total margin deposit for the long term position holder. This example indicates that futures trading is risky as it is highly leveraged. Due diligent and substantial training is usually required to be successful in the futures market.

FUTURES PRICING

The price of a futures contract can be derived either from the time-value of money principle or from arbitrage arguments.^[1] From the time value of money viewpoint, the futures price represents the future value of prepared forward price of the underlying asset.

The exact expression of the futures price depends on the underlying assets and presence of yields. If the underlying asset does not generate yields, the price of a futures contract can be expressed as

$$F_{0,T} = S_0 e^{rT} \quad (1)$$

where $F_{0,T}$ is the futures price of a stock today that is to be delivered at time T , S_0 is the spot price today, T is the time until delivery date, and r is the risk-free interest rate for maturity T .

If the underlying assets do generate an average yield, the futures price becomes

$$F_{0,T} = S_0 e^{(r-q)T} \quad (2)$$

where q is the average yield during the life of the contract.

These pricing models can be also viewed from the arbitrage-free principle. In an efficient market, a futures contract is so priced as to preclude arbitrage profit. In other words, investors will be indifferent to spot and futures market to execute their buying and selling of underlying assets because the prices they obtain are effectively the same.

OPTIONS AND OPTIONS MARKETS

An option is a contract that gives the buyer the right, but not the obligation, to buy or sell an underlying asset at a specific price on or before a certain date. The underlying assets for options include financial assets and commodities, such as stocks, bonds, currencies, commodities, oil, corn, and orange juice. The options are traded in options markets across the world. The most prominent options market in the United States is CME in Chicago.

Options fall under two main categories, the call option and the put option. A call option is a right to buy a certain asset by a certain date for a certain price. A put is an option to sell a certain asset by a certain date for a certain price.

People who buy options are called holders and those who sell options are called writers. As in the stock market, the buyers are said to have long positions, and sellers are said to have short positions.

The price at which an underlying stock can be purchased or sold is called the strike price. This is also called striking price or exercise price. This is the price at which a stock will be sold when the option holders opt to exercise the option. All of this must occur before the expiration date.

As a right but not an obligation to purchase underlying assets at a predetermined price, an option has a value, which is called the option premium. Call and

put option premiums depend upon the stock price, strike price, time remaining until expiration (time value), interest rate, and volatility. The most important of these is the strike price of the option relative to the market price of the underlying futures contract. A call option becomes more valuable, and hence the premium becomes greater, as the market price of the underlying asset rises above the option's strike price.

The value in options can be classified into two categories, the intrinsic value and the time value. For call options, the option is said to be in the money if the share price is above the strike price. A put option is in the money when the share price is below the strike price. The amount by which an option is in the money is its intrinsic value. An option may or may not have intrinsic value, but it will almost always have time value. Any option may become valuable in the future and, hence, a liability to the seller, so the time value exists to compensate option sellers accordingly.

Two major types of options are traded in the exchanges, American options and European options. American options can be exercised at any time between the date of purchase and the expiration date. Most exchange-traded options are of this type. European options can only be exercised at the end of their life. They are less flexible than American options. Given their tighter structure, the pricing models to value European options are better developed. For this reason, a larger part of the discussion will be on European option pricing in the following section.

OPTIONS PRICING

There are two major option-pricing models that are widely published in the valuation of options, the discrete model and the continuous model. The discrete model is based on a binomial stock process model, developed by Cox, Ross, and Rubinstein.^[2] The continuous model assumes a geometric Brownian stock price motion, based on the landmark paper of Black and Scholes.^[3] Although both models are rooted in Black and Scholes' conception, we discuss the binomial model first, recognizing its intuitive appeal and pedagogical value.

Binomial Option Pricing Model

In the binomial model, the price of the underlying stock is assumed to take two, and only two, possible values. The extension of the binomial model creates the binomial tree, representing the stock price generating process and the result outcomes.

The option price is then derived from a complementary stock and bond portfolio, based on the non-arbitrage principle. The resulting equation for options

price and the definition formula for risk adjusted probability (1) are given in Eqs. (3) and (4).

$$C = [pC_u + (1 - p)C_d]/e^{rh} \quad (3)$$

$$p = [\exp((r - d) \times h) - d]/[u - d] \quad (4)$$

where C represents the value of the option, p is the risk-adjusted probability of an upward move, $1 - p$ is the risk-adjusted probability of a downward move, u is the multiplier for an upward move, d is the multiplier for a downward move, r is the risk-free rate per period, h is the length of a period, and d is the continuous dividend yield.

Specifically, there are two major steps in binomial modeling. First, the time to expiration of an option is broken down into potentially a very large number of intervals. At each interval, the stock price is allowed to move up or down by an amount calculated using volatility and time to expiration. This produces a binomial distribution tree, or recombining tree, of underlying stock prices. The binomial tree represents all the possible paths that the stock price could take during the life of the option. The process works from the present to future expiration. At the end of the tree, i.e., at expiration of the option, all the terminal option prices for each of the final possible stock prices are known, as they simply equal their intrinsic values.

Second, the option prices at each step of the tree are calculated by working back from expiration to the present. The option prices at each step are used to derive the option prices at the next step of the tree using risk neutral valuation based on the probabilities of the stock prices moving up or down, the risk-free rate, and the time interval of each step. Any adjustments to stock prices or option prices are worked into the calculations at the required point in time. At the top of the tree, we are left with one option price.

The binomial model is easier to understand intuitively. It provides an approximation of the more math-intensive continuous model. Another advantage of the binomial model is that it can be used to accurately price American options, which is not available in the closed form of the more accurate continuous model. This is because with the binomial model, it is possible to check at every point in an option's life (that is, at every step of the binomial tree) for the possibility of early exercise.

Black–Scholes Option Pricing Model

The Black–Scholes option pricing model assumes that the underlying stock price follows a geometric Brownian motion with constant volatility. The change of option price is represented and restricted in a partial

derivative equation, the famous Black–Scholes partial differential equation.^[4,5] The solution to the equation is the Black–Scholes formula for options price. The Black–Scholes formula for the price of call option and put option are shown below respectively:

$$c = S_0N(d_1) - Ke^{-rT}N(d_2) \quad (5)$$

$$p = Ke^{-rT}N(-d_2) - S_0N(-d_1) \quad (6)$$

where

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

S is the price of the underlying stock, x is the strike price, r is the continuously compounded risk-free interest rate, T is the time in years until the expiration of the option, σ is the implied volatility for the underlying stock, and N is the standard normal cumulative distribution function.

The Black–Scholes option pricing formula prices European put or call options on a stock that does not pay a dividend or make other distributions. An example of the application is presented in the next section.

EXAMPLE OF OPTION PRICING BY BLACK–SCHOLES MODEL

Consider a European call option on 100 shares of non-dividend-paying stock XYZ. The option is struck at \$40 and expires in three months ($T = 0.25$). XYZ is trading at \$41 and has 30% ($\sigma = 0.30$) implied volatility. The continuously compounded risk-free interest rate is 0.03. Applying Eq. (5), the call option's market value per share of XYZ can be calculated at approximately \$3.06. Since the call is for 100 shares, its total value is \$306.

REFERENCES

1. McDonald, R.L. *Derivatives Markets*; Addison Wesley: Boston, 2003.
2. Cox, J.; Ross, S.; Rubinstein, M. Option pricing: a simplified approach. *J. Financ. Econ.* **1979**, *7*, 44–50.
3. Black, F.; Scholes, M.S. The pricing of options and corporate liabilities. *J. Polit. Econ.* **1973**, *81*, 637–654.
4. Hull, J. *Options, Futures and Other Derivative*, 5th Ed.; Prentice-Hall: Englewood Cliffs, NJ, 2003.
5. Hull, J. *Introduction to Futures and Options Markets*, 3rd Ed.; Prentice-Hall: Englewood Cliffs, NJ, 1998.

Game Theory

Michael A. Kelley

Department of Social Sciences, Pittsburg State University, Pittsburg, Kansas, U.S.A.

INTRODUCTION

Game theory is a mathematical approach to individual decision making that employs games as paradigms of rational decision-maker interactions. Used to study a variety of social, economic, and political issues, such as arms races, legislative coalition formation, military strategy, regulative public policy, voting behavior, and even aspects of marriage, game theoretic models delineate those decisions actors should make when confronted with competing policy alternatives or decision consequences. Far from seeing games as trivial, game theorists posit that game decisions exhibit the same rules or forces that drive more complex political and other social interactions and outcomes. After all, games, like politics, involve player moves and interactions, selection of strategies with specific consequences, and, at times, coalition formation, all predicated on a desire to maximize individual player interests.

POSITIVE POLITICAL THEORY AND THE RATIONALITY ASSUMPTION

Intrinsic to game theory is the idea that individual decisions are rational, i.e., individuals are engaged in purposeful action directed at ordered goal attainment. Positive political theory asserts that individuals order their goals when engaged in public choice, choosing to maximize their expected utility and preferring to win or gain rather than lose or fail to gain when there are real conflicts of interest. Indeed, enforced policy choices by government are the result of socially expressed, ordered, and calculated preferences for particular policies. The idea of rationality and the resulting regularity in human behavior explicates a positive science of politics that encourages the confrontation of assertion with experimentation and tests of validity.

Game theory, which employs games as abstractions of real-world situations and assumes player rationality, assigns specific mathematical payoffs to decision situations. It delineates the appropriate strategy rational players will pursue given a particular game and given that players maximize gains and limit losses when confronted with a decision or problem. There is the further

expectation that other players in the game will behave similarly in order to maximize their own individual utility, when there are competing policy alternatives. Games, thus, are heuristic devices in that they simplify reality based on particular assumptions in order to provide insights into human conduct.

Game theory involves decision problems, where there are at least two parties, and players must choose among competing decision alternatives with particular consequences, consistently order their preferences based on decision consequences and where the consequential interaction of player choices can be specified mathematically. Example 1 exemplifies these points and illuminates the assumption of rationality.

Example 1

"Polluter versus Regulator"

<i>Polluter</i>		<i>Delay (A)</i>	<i>Comply (B)</i>
<i>Regulator</i>	<i>Intensive (A*)</i>	8; -8 (A*; A)	5; -5 (A*; B)
	<i>Less Intense (B*)</i>	1; -1 (B*; A)	3; -3 (B*; B)

In this situation, the regulator confronts a polluter with a set of new rules. Row A* provides for more intensive regulative monitoring, with positive payoffs (fines collected from polluter) of at least five units, and this is the regulator's dominant strategy in this game (payoff is always higher than another choice, i.e., $8 > 3$ and $5 > 1$); while row B* involves less monitoring and would permit more cheating by the polluter. The polluter, wanting to avoid cleanup costs, must choose between meeting the rules (column B) or choosing a strategy (column A) of doing only what is required after legal action. Polluter's choice is column B in this zero-sum game (what one player gets, the other loses), because the polluter wants to limit losses and assumes that a rational regulator will choose row A* to maximize the fines collected. With this choice, at most, polluter will lose five and could lose only three units, if the regulator fails to act rationally and selects row B*. The rational result would be a payoff of +5

units to the regulator and -5 units to the polluter (A^*B), but polluter avoids the -8 units consequence by pursuing a rational strategy of compliance.

EXAMPLES OF GAME SITUATIONS

Game theory is not only a mathematically based science of decision problems, it also is a classification scheme for games. As surrogates, games must reflect the complexity of real-world decision situations. Games are classified by the number of players ($2 \dots n$) and the types of outcomes (typically, zero-sum and non-zero-sum). Each game in the resulting game classification matrix has its own particular set of rules, which explain expected player choice relative to specified outcomes.

Two-Person, Zero-Sum Games

As seen in Example 1's highly conflictual situation, whatever one player wins, the other player loses, interests are diametrically opposed, player payoffs add to zero, and both antagonists possess perfect information about the consequences of their actions. Exemplified by wartime strategy, decisions in this game are based on the "worst-case assumption" or the idea that each player will seek maximal advantage. Assuming the worst, a *minimax* strategy, where regulator seeks to minimize the maximum amount of harm to interests in the face of Regulator's *maximin* choice of maximizing its minimum gain. In the example, a *minimax* strategy leads to the saddle point or optimal solution of A^*B or 5, -5 (the payoff that is a row minimum and a column maximum), in that this is the best that either player can do, given that each opponent is rational. In a zero-sum, two-person game there is always at least one pair of optimal strategies (if there is more than one set of optimal strategies, players are indifferent to any particular pair), hence, there is always a solution to the game.

Two-Person, Non-Zero-Sum Games

In these games, player interests are not diametrically opposed but rather have some shared interest in avoiding particular outcomes. An oft-used example of this game is the "Prisoner's Dilemma." As Example 2 illustrates, two prisoners charged with the same serious crime are separated and cannot communicate. The prosecutor can convict both of them of a minor crime but must get at least one of them to inform on the other in order to solve a more serious charge. During the interrogation, the prosecutor informs the separated prisoners that the one who informs will be set free, and

the other will spend 15 years in jail. If both cooperate and remain silent, they both spend 2 years in jail, while if they both inform on each other, they will each spend 7 years in jail. Which strategy should be pursued is the salient question, i.e., to maximize one's payoff or a strategy of trusting the other to remain silent. Prisoner 1's maximizing strategy would be such that: $B^*A > A^*A$ and $B^*B > A^*B$, hence, his rational strategy is to inform. Prisoner 2 rationally understands that $A^*B > A^*A$ and $B^*B > B^*A$ and also chooses to inform. Pursuit of individual rationality results in a 7-year prison sentence for each prisoner (B^*B). On the other hand, they would have served only 2 years each (A^*A), if they could trust each other to remain silent. The dilemma exists, because the rational choice of payoff maximization and the fear of the consequences of the other cheating results in a lower payoff than an alternative approach of trust, hence, the "Prisoner's Dilemma." This "game" is often used to describe why arms races occur even when cooperation would provide for a more satisfactory outcome for both parties.

Example 2

"Prisoners Dilemma"

Prisoner 2

		Quiet (A)	Informs (B)
Prisoner 1	Quiet (A^*)	$A^*; A$ 2; 2	$A^*; B$ 15; 0
	Informs (B^*)	$B^*; A$ 0; 15	$B^*; B$ 7; 7

N-Person, Zero-Sum Games

In these games, the zero-sum rules apply, but the distinction is that the winners are members of a coalition whose gain equals the losses suffered by those outside of the winning coalition, and where players can reach private agreements about the division of the spoils (side payments). The game's characteristic function (how much each possible coalition of actors can receive by acting in concert) will determine the game's outcome. Players taking into account the characteristic function will form coalitions that they believe are only large enough to win and which maximize their individual gains. This aspect of a game helps explain why it is so difficult to form and maintain "grand" legislative coalitions (see Example 3), because players can win more of the finite rewards if they form smaller coalitions. Indeed, empirical observation of grand alliance structures in domestic and international politics demonstrates that "grand" coalitions tend to form in times of crisis or discontinuity and then devolve into

Example 3*“Size Principle and Legislative Coalitions”*

In this simple game (where the characteristic function is either 1 or 0), the parties in a legislature want to control the government in order to appoint ministers but must form a coalition in order to win. Legislative strengths of the parties are as follows:

Party A = 45%
 Party B = 20%
 Party C = 25%
 Party D = 10%

The minimum winning coalitions (there is no majority if a member is subtracted) are AB, AC, AD, and BCD. It is possible to further refine the results by noting that A probably prefers AD to the other possibilities, because B and C will be more demanding, given their relative legislative strengths. D also prefers AD, because it will share the win with only one coalition partner or the smallest number of players. The AD coalition is the one likely to emerge, if there are no ideological or information problems that lead to larger than minimal winning coalitions.

competing parties or smaller coalitions when the crisis has passed.

N-Person, Nonzero-Sum Games

This game is most relevant to political and administrative decision making, the central aspect of which is coalition formation, where the benefits of cooperation outweigh those of maximizing individual utility. In these games, where constant communication occurs over many different decisions, cooperation can bring

positive payoffs, while noncooperation provides no increase in individual utility. In the $N = 4$ example (Example 4), the question is, when will an individual support a fee, if the levy will generate a desired public value twice as great as the total costs to the individual players. While each person would prefer to receive benefits at no cost, because there is communication, players are more likely to prefer to win something than pursue an individually rational strategy that would result in no individual benefit. After all, players prefer to win something than forego winning at all. A coalition will form if the players receive a benefit

Example 4*“Community Good”*

Agree	Individual Cost	Total Cost	Community Good	Individual Benefit	Individual Gain	Cell
Yes—4 No—0	\$40 \$0	\$160	\$320	\$80	\$40	A
Yes—3 No—1	\$40 \$0	\$120	\$240	\$60	\$20 \$60	B
Yes—2 No—2	\$40 \$0	\$80	\$160	\$40	\$0 \$40	C
Yes—1 No—3	\$40 \$0	\$40	\$80	\$20	-\$20 \$20	D
Yes—0	\$0				\$0	E

The most likely result is Cell C. In Cell E, no one gains anything; while in D, it is unlikely that a single player will solely finance the public good. On the other hand, there is a benefit to defect (compare individual gain in Cell B and Cell A) from the grand coalition. However, if a player can convince just one other person of the levy's value, there is no cost relative to the received and desired benefit (Cell C). A minimal coalition can be formed to acquire the desired good, provided it is profitable (or does not cost) the players.

or there is no cost associated with receiving the public good.

CONCLUSION

The idea that individuals are rational gives social scientists a powerful mechanism with which to understand various forms of political and social decision making. The theory of games assumes that decisions are not randomly derived but rather are arrived at through a process where each player seeks to maximize positive rewards relative to costs in situations when there are competing perspectives. Games, with their specified outcomes relative to particular decision problems, permit decision-making theorists to look at real decisions in a predictive or evaluative mode. The predictive mode allows for the delineation of a strategy relative to an expressed set of ordered goals, with the presumption that rationality compels particular procedures and results. On the other hand, the evaluative approach infers from actual decisions the goals which must have existed under the rationality assumption to lead to the

particular result. Games are not perfect analogies to real-world decision making, yet the use of games allows us to test to see if decision makers behaved in a purposeful and ordered fashion to maximize individual utilities.

BIBLIOGRAPHY

1. Bartos, O.J. *Simple Models of Group Behavior*; Columbia University Press, 1967.
2. Hamburger, H. *Games as Models of Social Phenomena*; W.H. Freeman and Company, 1979.
3. Rapoport, A. *N-Person Game Theory*; University of Michigan, 1973.
4. Rapoport, A. *Two-Person Game Theory*; University of Michigan, 1973.
5. Riker, W.H. *The Theory of Political Coalitions*; Yale University Press, 1962.
6. Riker, W.H.; Ordeshook, P.C. *An Introduction to Positive Political Theory*; Prentice-Hall, 1973.
7. Groennings, S., Kelley, E.W., Leiserson, M., Eds.; *The Study of Coalition Behavior*; Holt, Rinehart, and Winston, 1970.

Gaza Disengagement: Israel Government Forced Resettlement (2005)

Nitza Nachmias

School of Political Science, University of Haifa, Haifa, Israel

INTRODUCTION

Former Israel Prime Minister Ariel Sharon introduced the plan to “disengage” from the Gaza Strip to the Israeli government in early February 2004, at the peak of international criticism of Israel’s project of the separation wall and of Israel’s military operations in Gaza and the West Bank. The plan received the blessing of the Quartet, a political forum consisting of representative of the US, EU, UN, and Russia, that pledged generous funds for a rehabilitation program of Gaza. The Israeli government assumed responsibility for the proper resettlement of the Gaza evacuees. The termination of Israel’s 40-year occupation of Gaza was expected to greatly benefit both Israel and the Palestinian Authority.

On August 5, 2005, Israeli army and police units launched a three-day unilateral military campaign, relocating about 10,000 settlers (some 2000 households) to temporary homes inside Israel. The Palestinian authority was not a party to the project, but tacitly cooperated and planned to assume full control over the evacuated territory. During the operation, the settlements houses and infrastructure were razed to the ground.

However, the outcome of the Disengagement was very disappointing and disturbing to both Israel and the Palestinian Authority. In August 2007, two years after the pullout, most of the Israeli evacuees have been still living in temporary housing and many families have not found suitable employment. The children especially experienced a hard time and their school achievements have declined. In Gaza, economic and political conditions deteriorated drastically in the wake of the 2006 Hamas election victory that resulted in an international boycott. Clearly, the Disengagement operation that began with good intentions ended with doubtful results.

DISENGAGEMENT AND FORCED RESETTLEMENT DEFINED

This article defines “disengagement” as an act of departure, a release from an attachment and a separation

from people or places of friendliness. “Political Disengagement” implies a withdrawal, a pullback, and a pullout from a political conflict. “Strategic Disengagement” denotes a break off a military action with an enemy.^a The Gaza Disengagement included social, political, and strategic departures. The Israel military lost control over strategically sensitive border, while the settlers lost their homes and saw their social and economic infrastructures ruined. In legal terms the evacuees became “internally displaced persons” (Table 1).

International law distinguishes between people who are “refugees”, namely, people who fled their homes and sought refuge in a foreign country for fear for their lives, and “Internally Displaced Persons” (IDPs), who also fled their homes for fear for their lives but remained within their own borders^[1]. Unlike refugees, IDPs are not protected under the refugee’s convention of 1951 and are left at the mercy of their own government, often the government that caused their displacement. The Israel government appropriated \$2.2 billion for the resettlement of the Gaza IDPs,^b however, the results have been less than satisfactory.

The final definition of the Disengagement that was reached with the mediation of the Quartet was: 1) Israel would transfer to the Palestinian Authority (PA) the full control of the Gaza Strip and 2) would lift the closures of Gaza and the border with Egypt. In return, the PA would 1) stop all anti-Israeli violence from Gaza, and 2) would end the shelling from Gaza of cities in southern Israel. The compromise was perceived to benefit the security and well being of both sides.

STRATEGIC, ECONOMIC, AND POLITICAL DILEMMAS

Israel security concerns focused on the three main issues: 1) border crossings from Egypt to Gaza, 2) border

^aThe author’s interpretation of various sources.

^bOne-third of the aid would help defray expenses associated with the redeployment of Israeli military bases and troops following the pullout from Gaza. The remaining two-thirds would be used to develop Israel’s Galilee region in the north and the Nitzan region in the south, areas where the evacuated settlers will be resettled.^[2] The Israeli Disengagement Law passed on February 16, 2005.^[2]

Table 1 Gaza strip population

Total population	Gaza City (incl. Jews)	Population estimates +10%
1949	80,000	200,000
1967	118,000	356,000
1996		963,000
2005		1,400,000

[From UNRWA data; PA projections (1997); FAFO; www.jewishvirtuallibrary.org.]

crossings from Gaza into Israel, and 3) the smuggling of weapons and fighters from Egypt into Gaza. The U.S. Secretary of State, Condoleezza Rice, was actively and personally involved in finalizing an agreement that would alleviate Israel's security concerns. The compromise was that Egypt and the PA would control the crossing and Israel would review the traffic on video cameras in real time. Israel would also receive lists of individuals and goods using the crossing. Palestinian convoys to the West Bank would require Israeli military escorts. Finally, Israel would continue to control the Gaza coast and skies.

The most contentious issue was the control of the southern Gaza Strip–Egyptian border, a 14 km, 50–100 m width “Philadelphi Route” (an Israeli code name). Israel knew that a complex network of underground tunnels existed and used to smuggle illegal weapons, drugs, and fighters from Egypt into Gaza. Israel rightly questioned the Egyptian military capability to control the strategic Philadelphi Route, namely, its ability to prevent the illegal underground activity. Indeed, the 750 Egyptian troops who were deployed along the border, and the 70 unarmed European Union policemen and border experts who supported them, were unable to effectively fulfill their strategic and tactical missions.^c On September 12, 2005, the Israel Defense Forces (IDF) completed their military pullout and were redeployed behind the Philadelphi Route. Israel erected a heavy fortification system along the border and the movement of people and goods was monitored via video cameras in real time. Regrettably, the tunnel activities have never stopped, and the IDF has been constantly blowing up old and new tunnels along the Gaza border with Egypt. For example, on October 8, 2006, Ynet news quoted Israel intelligence estimates that 7500 terrorists with large quantities of advanced missile hardware have entered Gaza after the Disengagement.^d Israel also claimed that Al-Qaeda

members began to infiltrate Gaza through the underground tunnels.

The PA was equally disappointed from the results of the Disengagement. The Quartet's hopes for an economic revival of Gaza were dashed. The Quartet's special envoy, James D. Wolfenson, former Director of the World Bank, labored incessantly to ensure the peaceful transition of the economic and agricultural assets from the Jewish settlers to the PA. He planned mutual industrial zones, trade, transportation continuity between Gaza and the west Bank, and the peaceful transformation of the settlers' greenhouses, valued at \$U.S. 14 million to the PA. Wolfenson's initial optimism was soon replaced with deep frustration. The economic and political conditions in Gaza deteriorated while Israel did not relinquish its control of the Gaza air and the seaports. Israel argued that as long as the rocket attacks on its southern townships continued and terrorist groups were allowed to operate openly and freely in Gaza, it would not lift the air and sea closures. The victory of the terrorist organization Hamas in the 2006 Palestinian elections was perhaps the final nail in the peace process coffin.

THE UNREALISTIC GOALS OF THE DISENGAGEMENT

The Disengagement plan included an elaborate list of both realistic and unrealistic strategic, economic, and political goals.

1. *Strategic goals*: 1) relocate all the settlers to other areas, 2) remove all IDF outposts and installations from the Gaza Strip and Northern West Bank, 3) reduce the military force manning checkpoints throughout the West Bank, 4) encourage the PA to fight terrorism, 5) continue the building of the West Bank barrier, and 6) continue to control Gaza air and coastal waters.
2. *Political Goals*: 1) revive the 2003 Quartet's Road map for peace, 2) reduce friction between Jewish Israelis and Palestinian Israelis, 3) absolve Israel of legal responsibility for Gazan Arabs, and 4) gain international support of Israel foreign policy.
3. *Economic goals*: 1) renew and increase trade with Egypt and the PA, 2) improve the Gaza and the West Bank economies, e.g., the greenhouses were expected to yield over \$50 million

^cThe “EU Police Mission for the Palestinian Territories” (EUPOL-COPPS) is a group of about 70 advisors who were dispatched to Gaza on January 1, 2006 with a mandate of 3 years (EU13696/05). Major-General Pietro Pistolese of Italy heads the mission.

^dThere now appears to be substantial evidence that al-Qaeda has acquired a limited but growing presence in Gaza. This presence seems to have increased since disengagement.^[3]

per year, and 3) improve the movement of people and goods to and from Gaza.

ACTUAL RESULTS

The Disengagement hardly achieved its goals. First, while Israel did not freeze all settlement activity, the PA never stopped or even reduced the violence against Israel. Second, the economic conditions in Gaza greatly deteriorated, and the Quartet shelved all the plans for the economic reconstruction of Gaza. The 2005 World Bank report showed a 40% decline in the Palestinian GDP, while PCI declined by 26.4%.^c Third, Israel's security was not enhanced. In fact, attacks on Israeli southern townships from Gaza intensified, and underground smuggling activities increased. Israel retaliated by launching air strikes and military land attacks and imposing severe and frequent closure and travel restrictions on Gaza and the West Bank.

Additional unpredicted adverse results were:

1. The final cost of the Disengagement greatly exceeded all estimates. The original budget of 3 billion NIS was revised several times, to 4.3 billion NIS, or about \$1 billion,^[4] and later to about 8–10 billion NIS, or \$2 billion.
2. The resettled families have been experiencing severe social and psychological problems, unemployment, increased juvenile delinquency, high school dropouts, and severe shortages of permanent housings.^[5]

P3: POLITICS, PROCEDURES, PROCESSES

Politics

Prime Minister Ariel Sharon launched his surprising political and strategic initiative on December 18, 2003, in a speech at an academic conference. "The purpose of the Disengagement Plan is to reduce terrorism as much as possible, and grant Israeli citizens the maximum level of security. The process of disengagement will lead to an improvement in the quality of life, and will strengthen the Israeli economy. The unilateral steps will be fully coordinated with the United States."^f The essence of the pullout was 1) to shift wasted resources from Gush Katif to the West Bank settlements, 2) to reduce friction and hostilities in

Gaza, 3) to negotiate a settlement on the West Bank on more favorable terms, and 4) to provide the IDF with a free hand to operate against terrorists in Gaza. Sharon needed the support of the United States to his unilateral plan, and in early April 2004 he traveled to Washington. He returned with four major promises: 1) a letter from President Bush supporting the unilateral Disengagement,^g 2) a tacit promise to back Israel policy on major issues to be negotiated in any future peace agreement, 3) the Palestinian refugees would be settled in the future Palestinian state and Israel would not be responsible for their "right of return", and 4) the United States recognizes Israel security need for a temporary security fence. The Bush letter caused a fury in the Arab world, especially the "right of return" clause and the tacit recognition of the settlements in the West Bank.

Procedures and Processes

The Disengagement plan encountered strong opposition among the leadership of Sharon's own Likud party, who even called for his removal from office. Sharon underestimated the opposition and called for a referendum, on May 2, 2004, of all registered Likud voters. To Sharon's dismay 65% of the 193,000 Likud members voted against the Disengagement. The opponents' arguments were twofold: 1) Gush Katif was necessary to protect Israel's border with Egypt and to prevent terrorist infiltration into Israel, and 2) Israel has a historical/biblical claim to the land. Sharon prevailed, and the Israel Cabinet approved, on June 6, 2004, (14–7) the Disengagement plan with the "yes" vote of the Labor opposition and other smaller parties. The second hurdle was the vote in the Legislature. On October 11, 2004, at the opening of the Knesset winter session, Sharon introduced the Disengagement Law that would appropriate large compensations to settlers' families, businesses, and public institutions. On October 26, the Knesset gave a preliminary approval for the plan with: 67 "yes", 45 "no", 7 abstentions, and 1 member absent. Sharon labored over 6 months to build a coalition of "enemies" namely, Likud; Labor; and the center party Shinui. Using the "carrot and stick" tactics Sharon approved the allocation of \$67 million to the United Torah Judaism party for its religious schools in exchange for five votes; the Labor party received about \$140 million for 19 of its 21 votes; and the secular Shinui party received \$163 million to secure the "yes" vote of its 14 Knesset members. Sharon overcame the final legislative hurdle and on February 16, 2005, the Knesset finalized

^cThe 2005 World Bank report showed a 40% decline in the Palestinian GDP during the first 3 years of the Intifada, while the PCI declined by 26.4%.

^fSharon speech before the Fourth Hertzeliya Conference, organized by the Center for Interdisciplinary Studies, 12.18.2003.

^gPresident letter to Sharon dated April 14, 2004.

Table 2 Cost of disengagement compensations

Payments to settlers	Original estimate	Revised/2005	Total
Homes/loss of livelihood	1.6 billion NIS	2.4 billion NIS	
Business compensations	1.0 billion NIS	1.4 billion NIS	
Public institutions	0.3 billion NIS	0.5 billion NIS	
Average family compensation	1.25 million NIS	1.7 million NIS	4.3 billion NIS

Source: Ha'aretz daily, February 16, 2005.

Note: On September 20, 2005, compensations were increased by 15% to about 2025 million NIS per family. The total cost of the compensations reached almost 9 billion NIS. Additional 135,000 NIS was offered as an incentive for resettlement either in the Galilee or in the Negev.

and approved the Disengagement Law with 59 in favor, 40 opposed, and 5 abstained.^[6] An opposition proposal to submit the plan to a national referendum was rejected (29–72). The Disengagement was scheduled for July 20, 2005 by a vote of 72–39. However, on June 9, 2005, the date for the evacuation was postponed to August 15, so as to not coincide with the Jewish commemoration of Tisha B'Av, traditionally a grief period for the destruction of the two Temples.

The opposition to the Disengagement organized massive anti-Disengagement campaigns, which included massive demonstrations by men, women, and children, who demonstrated along major highways and public places. The most confrontational action occurred on May 16, 2005, when protestors, carrying orange flags, symbol of the opposition to the Disengagement, blocked major intersections and highways. Notwithstanding the strong resistance to the Disengagement, the swift evacuation resulted in very few violent incidents. In total, 20 Jewish settlements: 16 in the Gaza Strip and 4 in Northern West Bank were dismantled and demolished. At the end of the day, 44 day-care centers, 36 kindergartens, 6 elementary schools, 3 high schools, 13 rabbinical Yeshivas, 38 synagogues, and a cemetery were evacuated and demolished. Undoubtedly, the Israeli society sustained an enormous emotional and economic cost.

It should be mentioned that about 66% of settlers, most of them secular, have accepted the compensation deals offered by the government^h and evacuated their homes prior to the deadline of August 15. The remaining hard-core, unrelenting, mostly orthodox families had to be forcefully evacuated. Complicating the evacuation was a group of about 5000 activists who illegally infiltrated the Gaza settlements over the weeks prior to the eviction date, violating the military order of March 17, 2005, prohibiting non-Gaza residents from relocating to that area. Only in a few cases of violent resistance the police had to carry out detentions and arrests.

Throughout the evacuation operation the Palestinians and Israel maintained an undeclared truce. Hamas and the other militant groups withheld their rocket attacks, while Israel refrained from air strikes and targeted assassinations. The joy of the Gazan people was clearly expressed by one resident: "The sun seemed brighter when it rose on the morning of Monday, September 12, minutes after the last Israeli soldiers left Gaza."^[7]

Ironically, Sharon, who had long been one of the architects of Israel's settlement expansion, had doggedly pursued and completed the forced resettlement of the most extreme group of settlers (Table 2).

GOOD INTENTIONS BUT POOR RESULTS

In February 2006, the Israeli Comptroller issued the most comprehensive review of the Disengagement. The review showed that while the intentions were right the results were poor:

1. The Sela bureaucracy was inadequate and lacked proper training.
2. Necessary coordination mechanisms with the security and the military forces were lacking.
3. Needed professional advisors in the field of agricultural relocation were not hired.
4. Social services were inadequate. Social workers were hired on a temporary basis.
5. The initial relocation of the 1750 families to their temporary housing was chaotic. The time needed for the construction of permanent housing was inaccurate. Consequently, hundreds of families were lodged for months in trailer-parks, hotels, and red-roofed mobile ("Car-a-villa") homes.
6. No adequate storage space was prepared for the goods and furniture of the evacuees.
7. The infrastructure of civil services necessary to accommodate the influx of the new settlers, i.e., education, health, etc. was not set up.
8. The appropriations for the resettlement were inadequate.

^hCompensation ranged from \$150,000–\$400,000 per family. Source: The Comptroller's report, February 2006.

9. Unemployment was hardly addressed. "Most of the residents (66%) worked where they lived...all are unemployed today. The rest 33% started their own private businesses."^[8] Additional 5365 billion NIS unemployment benefits were distributed in 2006, but the basic unemployment problem was not resolved.
10. The traumatic effects of the resettlement on children were not addressed. "Violence among the youth is high and rising. Drug abuse, which was negligible in Gaza, is on the rise... Children and youth have an almost psychotic fear of policemen and soldiers."^[8] The government approved \$1 billion additional funds for special assistance.^[9] However, the damage was done.
11. The report concluded that the poor management of the Disengagement traumatized the families and made it all but impossible for them to properly reconstruct their lives.

CONCLUSION

The Disengagement hardly answered the parties' expectations since it missed most of its political and strategic goals. First, the peace process suffered major adverse consequences. Mutual hostilities increased and both Egypt and the PA were either incapable, or unwilling to stop the underground traffic of weapons and fighters into Gaza. Second, Israel's assumption that the pullout would enhance its security was a wishful thinking. The electronically smart surveillance system that Israel had installed along the border to replace its military posts, proved useless. In fact, terrorist activities continued unabashed while the rocket attacks on Israel's Southern cities did not stop either. Third, the Palestinians resented Israel's intermittent military operations in Gaza and its continued control of the Gaza sea and airports. Fourth, the success of the Hamas in the 2006 Palestinian elections and their subsequent declarations denouncing all peace possibilities with Israel created dire economic conditions in Gaza. The United States and the European Commission

withheld economic assistance to the Hamas government, while Israel maintained the closures of the Gaza borders. Tougher restrictions were imposed on the movement of goods and people from and to the West Bank, and prospective income from the greenhouses did not materialized. Fifth, the humanitarian crisis in Gaza accelerated. All the Gaza reconstruction projects planned by the World Bank were shelved. Among them: rehabilitation of 100 km of Gaza roads; building of 2000 new housing units; installing 85 km of new water pipelines; reconstruction of sewerage network to 15,000 households; construction of up to 560 new classrooms; and construction of 45 new health clinics.^[10]

Finally, the physical and emotional cost of the Disengagement to the Israeli society was grossly underestimated. In fact, the entire Israeli population experienced the traumatic effects of the forced resettlement. The Disengagement was not a success story for the Palestinian community either. Perhaps the sole achievement of the Disengagement was the termination of the Israeli occupation of Gaza.

REFERENCES

1. Cohen, R. Foreword. In *The Politics of Forced Migration*; Nachmias, N., Goldshtein, R., Eds.; Publish America: Baltimore, MD, 2004.
2. Bashi, J. *The Marker*, December 22, 2005 (in Hebrew).
3. Keyes, D. *The Washington Institute Peace Watch* #531, December 16, 2005.
4. Israel Government resolution, February 16, 2005.
5. Israel government publication, 19 June, 2006, also Report of the Israel Comptroller General, February 2006, Haaretz Daily, January 3, 2006.
6. Yediot, A. Ha'aretz Daily, October 27, 2004.
7. Omer, M. *The Gaza Disengagement*, Washington Report on Middle East Affairs, November 5, 2005 www.wrmea.com.
8. Glick, C. The Jerusalem Post, November 25, 2005.
9. Israel government website; <http://sela.pmo.gov.il/PMO/Hitnatkut/HomePage.html>.
10. Local Aid Coordination Committee (LACC) report, September 22, 2005.

Gender and Public Administration

DeLysa Burnier

Department of Political Science, Ohio University, Athens, Ohio, U.S.A.

INTRODUCTION

Gender writing and research in American public administration typically begins with a description of the organizational reality facing women in the public sector. Among the facts regularly cited in such a description are that women lack parity with men in the workforce at all government levels. Women are concentrated in the lower levels of the civil service and in occupations designated historically as female; and they are underrepresented in state and federal governments' upper management ranks. Women remain underpaid relative to men, and "glass ceilings" continue to block progress up the professional ladder, despite years of public policies aimed at creating workplace equality.

Explicating the political, economic, and cultural elements of this reality with the aim of changing it has been one task of gender scholars. A second task has been to establish gender as an analytic category within public administration. The gender scholarship produced thus far is an evolving body of methodologically and theoretically diverse work that rests on two seminal texts: Kathy Ferguson's *The Feminist Critique of Bureaucracy* and Camilla Stivers' *Gender Images in Public Administration*.^[1,2]

For many reasons, gender scholarship has been slow to make it into public administration textbooks, journals, and classrooms. This entry's purpose is to bring the gender perspective into focus for public administration scholars unfamiliar with it. First, it will describe briefly feminist/gender theory. Second, it will highlight the contributions public administration gender scholars have made in the areas of history, theory, and practice. Third, it will propose a care-centered conception of public administration.

FEMINIST/GENDER THEORY

Feminist theory can be described variously, but three approaches generally have been used to address the issue of gender difference and inequality. The first approach is liberal feminism, which denies or dismisses the importance of sex-based differences. Perceived sex differences, whether biological or social in origin, provide no valid ground for denying women the rights and privileges accorded to men. This approach relies on the

liberal concepts of procedural justice, rights, and equality. It seeks to expose gender biases, challenge traditional sex roles, and implement institutional and legal reforms until women become fully equal to men.

The second approach, difference feminism, recognizes the importance of equality while embracing the differences that exist between men and women. The issue is not that gender differences exist, but that those qualities and experiences typically associated with women have been devalued. Because of their experiences as childbearers, mothers, and caregivers, and their knowledge of what it means to be powerless, women have developed important capacities to connect, nurture, and empathize with others. Out of these experiences, women have developed alternative "ways of knowing" based on emotion, empathy, and intuition.

The third approach, drawing on postmodernism, centers on the gender differences between women. The dualisms (e.g., male/female) central to the previous two approaches either overlook or downplay both the diversity of women's experiences and the fragmentation of their identities along the lines of age, race, class, ethnicity, and sexual orientation. No single theory, framework, set of values, or public policy solution captures the diverse contexts in which women live their multiple lives. Postmodern feminist scholars prefer accounts that are partial, sensitive to context, analytically diverse, and comparative rather than universal in their claims.

Feminist scholars increasingly have turned toward gender, not women, as the key analytic category in feminist scholarship. The term gender was embraced because it is inclusive of both sexes and highlights how the two sexes exist relationally. Gender is a "constitutive element of social relationships based on perceived differences between the sexes and it is a primary way of signifying relationships of power."^[3] Gender and gender relations are never neutral, but embedded in and shaped by such larger systems of power as culture, economics, and politics. Different historical moments and contexts produce and reproduce a vast array of gender power arrangements that affect how men and women come to understand themselves, each other, and the world around them. Gender power arrangements are embedded in organizations and institutions, and in the language, practices, and values that constitute them.

GENDERING PUBLIC ADMINISTRATION'S HISTORY

Until recently, it was understood that women played little to no role in the discipline's founding. The lone exception was Mary Parker Follett, whose writings were "lost" because they were categorized as part of the scientific principles approach to administration. She has been rediscovered currently as a pioneer organizational theorist whose work on power (power with, not over, subordinates) anticipated the humanistic approach to management. She is now credited for her pathbreaking ideas, that leading and administering are situational and relational practices; that administrative situations should be viewed from multiple perspectives; and that problems should be solved cooperatively.

Stivers' analysis of the "bureau men and settlement women," along with Guy's account of the "amazing Miss Burchfield," reveal how municipal reform organizations and disciplinary institutions were structured in ways that systematically excluded women, and both highlight the gendered nature of the discipline's founding narrative.^[4,5] Women were present, not absent, from the field's founding, but they and their professional contributions were "forgotten" when the narrative was constructed.

Not only did women play key roles in the discipline's founding, but also they along with the men involved in the settlement movement actually developed an alternative construction of administrative reality organized around care for others, improving people's actual living conditions, and "collaborating with those closest to problems to develop and implement solutions."^[4] For these reformers, the guiding metaphor was the home, not business, and they saw their reform work as "municipal housekeeping." Their values, practices, and reform goals formed a clear "feminine" alternative to the "masculine" values, practices, and reform goals that came to shape the discipline today.

GENDERING PUBLIC ADMINISTRATION THEORY

Because feminist theory is a complex, multifaceted body of work, there is no single overarching feminist theoretical framework in public administration. Gender scholars have drawn variously on feminist theory both as a means of critique and as a way to frame feminist alternatives to the bureaucratic and market methods of organizing. Feminist theory has been especially useful in exposing the long-standing belief that public administration's mainstream theories and concepts are gender neutral. In *Gender Images*, Stivers' analysis shows that mainstream theories and concepts (e.g.,

management, expertise, leadership, and virtue) contain "dilemmas of gender," and more specifically documents how they reflect values and practices culturally coded as masculine.^[2] McSwite, in *Legitimacy in Public Administration*, maintains that public administration's gender problem runs deeper than just the reflection of masculine values and practices. Drawing on psychoanalytic theory, McSwite argues that public administration as a discipline has suppressed the feminine since its founding by actively denying any role for emotion, feeling, intuition, collaboration, and process in mainstream administrative theory and practice.^[6]

One of the earliest gender scholars to envision public administration transformed by feminist theory was Ferguson, who proposed that "women's submerged discourse" could be used to build a nonbureaucratic collective life centered on human development and community needs.^[1] In the early 1990s, Hendricks argued for a "woman-centered" administrative reality that "empowers and affirms diversity," values the "caring encounter," privileges contextual accounts, and nurtures the development of individuals.^[7] In the early 2000s, Hutchinson proposed a "multigendered" public administration that reflects, not restricts, the different ways individuals perform their identities.^[8] Most recently, McSwite urged public administration to embrace the "model of administration" found in *Bureau Men, Settlement Women*, which "is based on the dynamic of consciousness in which the male and female functions perform in complementary balance."^[9]

GENDERING PUBLIC ADMINISTRATION PRACTICE

Acknowledging that public administration is an applied field, gender scholars purposely have focused on the need to change administrative practice if administration is to become gender inclusive. Changes would include improving women's numerical representation in bureaucracies; correcting women's underrepresentation in specific government professions and agencies; addressing sexual harassment in the public workplace; creating family friendly public workplaces; and removing barriers to women's promotion through the ranks of public executives. Public personnel practices also would need to change so that the performance of emotional labor, generally done by women, might become fully recognized and rewarded.^[10] Other changes would entail making bureaucracies less hierarchical, more participatory, and democratic; and they would be led by administrators whose models of leadership drew from feminist organizing and women's actual experiences as leaders and managers.

The most significant shifts in practice would come if public administration were to become explicitly care centered. A care-centered public administration would draw on the historical legacy of the settlement women and be grounded in an ethics of care. It would promote a situational, contextual approach to administration and place relational leadership at its center, which calls on leaders to build caring relationships inside and outside the organization. Among its governing values would be empathy, compassion, connection, and commitment.^[11]

If the care perspective is to become an alternative framework for public administration, then the assumptive base for theory and practice in the field would have to change. The traditional assumptive base for public administration is bureaucracy, with its focus on the efficient and neutral delivery of public goods and services. This assumptive base was challenged during the 1990s by those arguing for an alternative one borrowed from the marketplace, which emphasized the private and non-profit sectors as deliverers of public goods and services. The care perspective's assumptive base is relationships, and the obligation to create caring, democratic relationships in the public sector is presumed to be as vital as achieving market or bureaucratic efficiencies in the delivery of public goods and services. Relationships of all kinds would become part of the discipline's scholarly and practical focus, including those involving administrators and clients, administrators and citizens, administrators and elected officials, and administrators and coworkers.

Moving public administration toward care is important in part because it would make the discipline much more gender inclusive, but also because it is very likely that in this century how we care for each other as a society will be the next great domestic policy challenge. Public administration's role in meeting that challenge will depend on whether it takes care as seriously as it has taken bureaucracy, business, and markets.

CONCLUSION

This entry provided an overview of the evolving gender perspective within public administration.

Gender scholars have worked to make women, feminist theory, and gender itself visible within a discipline where the invisibility of gender long has been the norm. Their efforts have produced deeper, more complex readings of public administration's history, theory, and practice, and they have made a strong case for installing the concept of care as the assumptive base for theory and practice within public administration. A next step gender scholars would recommend is to establish a fresh discourse within public administration that legitimizes care narratively by telling the many stories of how care is valued and practiced by people at all levels of government.

REFERENCES

1. Ferguson, K. *The Feminist Case Against Bureaucracy*; Temple University Press: Philadelphia, PA, 1984.
2. Stivers, C. *Gender Images in Public Administration: Legitimacy and the Administrative State*, Rev. Ed; Sage: Newbury Park, CA, 2002.
3. Scott, J.W. *Gender and the Politics of History*; Columbia University Press: New York, 1988; 42pp.
4. Stivers, C. *Bureau Men, Settlement Women: Constructing Public Administration in the Progressive Era*; University of Kansas Press: Lawrence, KS, 2000.
5. Guy, M.E. The amazing Miss Burchfield. *Public Admin. Rev.* **2000**, 60 (1), 6–19.
6. McSwite, O.C. *Legitimacy in Public Administration: A Discourse Analysis*; Sage: Thousand Oaks, CA, 1997.
7. Hendricks, J.J. Women-centered reality and rational legalism. *Admin. Soc.* **1992**, 23 (4), 455–467.
8. Hutchinson, J. Multi-gendering pa: anti-administration, anti-blues. *Admin. Theory Praxis* **2001**, 23 (4), 589–604.
9. McSwite, O.C. Creating reality through administrative practice: a psychoanalytic reading of Camilla Stivers' *Bureau Men, Settlement Women*. *Admin. Soc.* **2004**, 36 (4), 406–426, 423–424.
10. Guy, M.E.; Newman, M. Women's jobs, men's jobs: sex segregation and emotional labor. *Public Admin. Rev.* **2004**, 64 (3), 289–298.
11. Burnier, D. Other voices/other rooms: towards a care-centered public administration. *Admin. Theory Praxis* **2003**, 25 (4), 529–544.

Gender Bias, Ethical Analysis of and in Public Policy

J. J. Hendricks

April Hejka-Ekins

Department of Politics and Public Administration, California State University, Stanislaus, Turlock, California, U.S.A.

INTRODUCTION

The content of this entry defines and discusses gender bias, the theories that have been central in its exposure and analysis, ethical implications, and the effect of gender bias in public policy. The theories discussed are elaborated upon as they affect the public service and its mission. Gender bias is an increasingly recognized topic with significant ramifications for citizens and public-sector employees. There are enormous ethical implications issuing from the way in which the legitimate needs of citizens are disproportionately recognized in policies and disproportionately served programmatically as a result of gender. In this entry, the ethics of care and the postmodern structural critique are presented. Both critiques are highly relevant to public administration. Finally, thoughts on attempts to present alternative feminist policy delivery systems are discussed.

GENDER DEFINED

Although often treated synonymously, gender is to be distinguished from the biological category sex, which is used to denote femaleness and maleness. Both sex and gender have come under increasing scrutiny as categories. They have been subjected to criticism as essential, therefore natural, categories. Gender is a dichotomous typology, which is socially constructed through the many processes that pervade a culture. Through the creation of the social types, female and male, humans are divided into categories, and traits and roles deemed appropriate for males and females are assigned to each. Traditionally, in patriarchal societies, males are assigned traits such as strength and competitiveness, while females are assigned traits such as docility and the capacity to nurture. These traits are assigned differently in various cultures, an example being the hijras of India, men who adopt female dress and behavior.^[1]

GENDER ROLES AND BEHAVIORAL NORMS

Gender roles are the behavioral sets that accompany the assignment of gender to the categories of male or

female. Members of society are socialized to display the appropriate behaviors assigned their gender roles. Although these learned roles may vary somewhat in different contexts, for example, in the workplace or in the home, expectations for both genders are remarkably durable and affect interactions in all contexts—the family, the educational setting, and the workplace—to one degree or another. Given the establishment of contexts as the domains of males, for example, combat and business, or females, for example, the home and professions such as social work and nursing, these contexts take on the behavioral norms associated with an assigned gender. In addition, much work is segregated by gender, which means that structural as well as behavioral norms are gendered.

THEORETICAL CRITIQUE

Gender and sex have been subjected to heavy criticism. Sex as a biological phenomenon is a great deal more plastic than popular culture allows.^[2] This is important to a discussion of gender, as biological determinism is often used to support the essential nature of males and females and the established gender categorizations. Biological determinism is the claim that sex determines behavioral outcomes.

Gender as a construct has been heavily criticized by all schools of feminist thought. One of the most recent and influential schools of thought, postmodern feminist critique, originates in a number of sources, but the work of Jacques Derrida is often utilized in constructing an argument. In Derrida's critique of modernism, reality is first divided into oppositional categories (either/or, black/white, male/female, public/private). One opposition is subsequently elevated and privileged, and one is devalued by comparison.^[3] Using Derrida's method of deconstruction, feminist theorists deconstruct the binary opposition of male/female and the subsequent assignment of these categories to either the public or private arena. Thus, bias enters the picture.

Gender bias can be defined as the privileging of males over females and things masculine over things feminine. Black feminist scholars have revised feminist

criticism to include the unequal positions of women of color. Patricia Hill Collins, for example, has developed a matrix of domination, in which a more complex hierarchical positioning that includes the intersection of gender and race is presented. This intersection would incorporate an analysis of the devaluing women of color experience doubly, due to both white privilege and male privilege.^[4]

STREAMS OF ETHICAL ANALYSIS

As a reaction to gender bias, two streams of ethical theory and analysis have emerged. One stems from the feminine school of thought, which advocates that an ethic of care needs to be valued by society and incorporated into organizational norms and institutional processes. A second stems from postmodern feminist theorists who critique the structure and arrangement of male-dominated organizations. Both schools attempt to eliminate gender bias through different approaches.^[5]

THE FEMININE ETHIC OF CARE

As compiled by Scranton and Rainney, the literature indicates that the feminine ethic of care was first articulated by Carol Gilligan in 1982 as a reaction against what she considered to be the gender-biased theory of Lawrence Kohlberg, who advocated that justice is the primary ethic in decision making. Kohlberg constructed a six-stage model of moral development, based on notions of universality, abstract reasoning, and individual rights, while relegating the value of caring to stage three in his schema. In response, Gilligan criticized Kohlberg for discounting the importance and complexity of the value of caring, asserted that the moral development of women is based on a different voice—an orientation of caring—and proposed a morality of care and responsibility. Other feminine theorists, such as Noddings, Tonto, and Chodorow, critiqued the exclusion of care as an essential aspect of gender bias and advocated for its acknowledgement and primary inclusion in all aspects of social and organizational life.^[5]

THE POSTMODERN CRITIQUE

However, research studies by Kanter, Derry, and Stewart and Sprintall have shown that the ethic of care is not expressed by women in male-dominated organizations. Indeed, because no value differences could be found among men and women, it appears that in order to succeed, women are forced to adapt to the demands

of the gendered organizational structure and culture in which they work. As a result, postmodern feminist ethicists critique male-dominated organizations by showing how the structure of hierarchy and the cultural values of aggression and competitiveness suppress the expression of care by women who strive to be successful in organizations.^[5] Kathy Ferguson maintained that the subordinate position of women in public life rests upon the increasingly bureaucratic nature of modern society. In this description, large-scale bureaucracy as a masculinist form of organization with a propensity toward conformity, control, and anonymity, systematically renders the voices of women as workers and clients silent, as well as any alternative vision of collective life.^[6] The recent work of Camilla Stivers discussed the materially different organizational reality and dissonant professional roles of women relative to men as civil servants.^[7] At the same time, Ferguson and Stivers exposed the contradiction that while bureaucracies in structure, culture, and practice are discriminatory toward women and hamper their full expression, the ethos of serving the public good renders the public workforce and its clientele “feminized” and devalued in the thinking of society.

PRESCRIPTIONS FOR CHANGE ISSUING FROM ETHICAL ANALYSIS

It is not difficult to see how devalued notions of the feminine in society affect the ethic of care within the fabric of organization and society. It is also not difficult to find ways in which to demonstrate alternative approaches to organizational life. Two responses to ameliorate gender bias have emerged from the feminine ethic of care within organizational life to transform male-dominated organizational structure and culture. Feminine theorists advocate for the inclusion of the ethic of care within organizational life to transform male-dominated institutions. For example, Helgeson described how women who reach the top levels of management are more willing to express the care ethic and implement feminine values, while Rosner suggested that women’s management style tends to be more inclusive, more flexible, and more supportive than men.^[5] By legitimizing the feminine care ethic, these theorists are trying to persuade the male-dominated power elite that inclusion of the care ethos in organizational structure and culture will improve work life and organizational effectiveness, while simultaneously uplifting the status of women. The second response by feminist theorists and activists is to reject gender-biased institutions and create feminist organizations that reflect a more organic, decentralized form and culture of values that include respect,

equality, diversity, open communication, collaboration, creativity, and consensus. Ferree and Martin documented a wide range of organizations that have been effective on a global level. Women envision these organizations as embodying a transformative process that provides a moral framework for expressing such values as nurturance, democracy, and empowerment.^[8]

POLICY IMPLICATIONS

One question that arises from these two women's movements to counter gender bias is the extent to which the feminine ethic of care and the creation of feminist organizations affect public policy and the public sector. In actuality, the devaluing of the position of women in the public sector interacts with the inequitable implementation of public policies to continue the concrete effects of gender bias. This actuality rests on a gendered philosophical and structural foundation, which initially excluded women.

GENDER BIAS IN THE FOUNDING DOCUMENT

The United States Constitution as the founding social contract on which public policy in the United States rests has been incisively criticized by MacKinnon for its exclusion of the standpoint of women citizens in the deliberation, articulation, and implementation of policy.^[8] This absence of the voices of women's issues from the binary construction of the socially constructed public spaces of men and the domestic spaces of women, was termed "the doctrine of separate spheres."

THE GENDERED NORMATIVE ORDER

Gendered institutions and constructed realities result in a view of public policy in which the feminine as a devalued category affects the "taken-for-granted" nature of the lived experience of citizens. Many of the gendered ways in which resources are allocated are viewed as normal and not questioned. Assertions of inequitable relations and policy distribution must do battle with the normative views of the established order. These views are legitimized by the institutional belief systems of, for example, religious and economic theories extant in society.

IMPACT OF GENDER BIAS REVEALED

The numerous arenas in which policy analyses detail inequitable distribution of resources by gender as well

as by gender intersecting with race, preclude elaborate or substantial documentation, given the scope of this entry. A broad overview presents a grim and inequitable picture. Occupational and wage distribution exhibits gross inequity. Poverty and welfare are increasingly manifest among women, spawning the term "the feminization of poverty." Sociologist Margaret Anderson indicated that in 1997, 31.6% of all female-headed households lived below the official poverty line.^[9] The callousness of the "taken-for-granted reality" of the American public is illustrated by the common term, "welfare queen." Such terms are not harmless, but affect political realities and the policies consistent with such views, i.e., "workfare." Work that is traditionally performed by women is excluded from that which produces a wage. Women in the workforce continue to earn around 74% the income of men. Workplace environments have been addressed above but must be included when addressing violence against women. The incidence of women's sexual harassment has remained at just below 50% since the policy guidelines were articulated in 1980. Costs to individuals, women as a group, and employers are many, serious, and economically very high.^[10] Domestic violence and incest seem impervious to policies and criminal statutes directed toward them. Women's health and reproduction issues continue to be defined largely by men who dominate science, medicine, and policy-making institutions. In issues such as war and peace, women's concerns are largely missing. While 14% of the officers in the military are women, gender segregation persists.^[9] As a gendered institution characterized by violence, aggression, and force, military values legitimize violence and sexual aggression in war. From the incidence of gender bias, one can only conclude that gendered institutions are to date impervious to efforts to ameliorate them or their effects. To overcome this tendency toward the status quo will take great and enduring effort.

CONCLUSION

One insight seems particularly important to note. Feminist organizations that resist cultural norms of gender bias have proliferated both inside and external to the public service over the past 30 years. They serve the purpose of providing alternative structural arrangements, as examples of what might be desirable to achieve. Though valuable in this way, they do not significantly impact the power inequities in public policy that act to structure and control women's lives. Their successes are striking but in no way create the change that the public is led to believe exists in the average women's daily lived existence. In fact, they often function to address and remediate the most

blatant violent examples of gender discrimination, masking the common realities. This conclusion is not meant to devalue or downplay the heroic and enduring efforts of feminists to act over time on behalf of women. However, until changes in consciousness and power occur, the ill effects of gender bias will remain unabated.

Resources:

Activist Organizations and Research:

Gender Watch: <http://gw.softlineweb.com/>.
 Feminist Majority on Line: <http://www.feminist.org/>.
 V-Day, <http://www.vday.org/index2.cfm>.
 Women of Color and Women Worldwide: <http://www-sul.stanford.edu/depts/ssrg/kkerns/wcolor.html>.

Journals:

Feminista!: <http://www.feminista.com/>.
Hypatia: <http://www.csupomona.edu/~hypatia/>.
Yale Journal of Law and Feminism: <http://www.yale.edu/lawnfem/law&fem.html>.

REFERENCES

1. Ward, M. *A World Full of Women*, 2nd Ed.; Allyn and Bacon: Massachusetts, 1999; 82 pp.
2. Fausto-Sterling, A. *Myths of Gender: Biological Theories About Women and Men*; Basic Books, 1992.
3. Derrida, J. *Margins of Philosophy*; Bass, A. Trans.; University of Chicago Press: Chicago, 1984.
4. Collins, P. *Black Feminist Thought: Knowledge, Consciousness and the Politics of Empowerment*, 2nd Ed.; Routledge: New York, 2000; 18 pp.
5. Scranton, A.; Rainey, M. Gender differences in administrative ethics. In *Handbook of Administrative Ethics*; Cooper, T.L., Ed.; Marcel Dekker, Inc.: New York, 2001; 555 pp.
6. Ferguson, K.E. *The Feminist Case Against Bureaucracy*; Temple University Press: Philadelphia, 1984.
7. Stivers, C. *Gender Images in Public Administration: Legitimacy and the Administrative State*; Sage Publications, Inc.: Newbury Park, CA, 1993.
8. MacKinnon, C. *Feminism Unmodified: Discourses on Life and Law*; Harvard University Press: Cambridge, MA, 1987.
9. Anderson, M. *Thinking About Women: Sociological Perspectives on Sex and Gender*, 5th Ed.; Allyn and Bacon: Boston, MA, 2000; 137, 308.
10. Petrocelli, W.; Repa, B.K. *Sexual Harassment on the Job: What It Is and How to Stop It*, 4th Ed.; Nolo: Berkeley, CA, 1999.

Generally Accepted Accounting Principles (GAAP)

Catherine L. Staples

Department of Economics and Business, Randolph-Macon College,
Ashland, Virginia, U.S.A.

INTRODUCTION

Generally accepted accounting principles (GAAP) are the minimum standards or rules organizations must follow in determining what financial information should be included in the general-purpose financial statements and how the information should be presented. These basic guidelines form the basis around which financial statements are created.

WHO ESTABLISHES GAAP?

The primary organizations responsible for setting accounting standards are the Governmental Accounting Standards Board (GASB), the Financial Accounting Standards Board (FASB), and the Federal Accounting Standards Advisory Board (FASAB). For state and local governments, GAAP is established primarily by the GASB, while the FASAB serves as the main standard-setting body for the federal government. All other entities, including non-governmental not-for-profits, follow guidance established by the FASB.

Of the three, FASB and GASB are considered parallel organizations. That is, they perform the same function, standard setting, but for different types of entities. In addition, oversight responsibility for both boards lies with the same private-sector organization, the Financial Accounting Foundation, whose primary responsibilities include appointing the seven members of each of the two boards and providing considerable financial support for the operating expenses of both.

Federal governmental agencies apply accounting and reporting standards recommended by the FASAB. This relatively new organization was created in 1990 and consists of 10 members, four of whom must come from the Board's sponsors which include the Department of the Treasury, the Office of Management and Budget, the Office of the Comptroller General, and the Congressional Budget Office. The remaining six members are chosen from the public or non-federal sector of accounting. After careful deliberation by the Board, and final approval by the Director of the Office of Management and Budget and the Comptroller General, a standard is issued by the FASAB.

Although the GASB, FASB, and the FASAB are the main standard-setting bodies, other organizations,

such as the American Institute of Certified Public Accountants (AICPA) and the Securities and Exchange Commission (SEC), also provide guidance on accounting and financial issues. This guidance may be considered GAAP if no formal statements have been issued on the topic. Widely recognized industry practices and accounting textbooks are also considered potential sources of GAAP.

OBJECTIVES OF ACCOUNTING AND FINANCIAL REPORTING

The development of GAAP is based upon a foundation of financial accounting objectives or goals. This foundation guides standard setters in their formation of the specific details for a statement and is based upon the primary purpose for the creation of an entity's financial statements. For instance, federal, state, and local governments are concerned with accountability to citizens, while for-profit organizations need to provide information to creditors and investors.

Accountability is the focal point of the accounting objectives for state and local governments. In particular, GASB's Concepts Statement No. 1, *Objectives of Financial Reporting*, purported that the basic function of a set of financial statements is to "justify the raising of public resources and the purposes for which they are used."^[1] Consequently, state and local government financial reports should be used primarily to compare actual financial outcomes with the legally adopted budget, appraise the financial condition and results of operations, and ensure compliance with applicable legal or contractual requirements. Governments must be accountable to several groups of stakeholders, including citizens, legislative bodies, oversight officials, investors, and creditors.

The FASB established accounting objectives with the focal point being the provision of useful information to users. FASB's Statement of Financial Accounting Concepts No. 1, *Objectives of Financial Reporting by Business Enterprises*, described the three main objectives as follows: 1) provide information about economic resources, claims to resources, and changes in resources and claims; 2) provide information useful in assessing the amount, timing, and uncertainty

of future cash flows; and 3) provide information useful in making investment and credit decisions.^[2]

The FASAB issued Statement of Federal Financial Accounting Concepts No. 1, *Objectives of Federal Financial Reporting*, detailing four objectives, which focused on accountability. Financial statements prepared by federal government agencies should be based upon the objectives of budgetary integrity, operating performance, stewardship, and systems and controls.^[3] A federal government agency’s compliance with applicable budgets, laws, and regulations is the focus of the objective budgetary integrity in reporting, while the objective of operating performance suggests the importance of evaluating and reporting an entity’s service efforts, costs, and accomplishments. The remaining objectives also reflect a need for accountability over a government’s operations and its financial systems and controls.

THE HIERARCHY OF GAAP

As shown, a variety of reference sources are available to finance officers, including official pronouncements of the FASB, GASB, and the FASAB, pronouncements and practice guides from other organizations, such as the SEC and the AICPA, and even accounting textbooks. Given this variety, how does a finance officer decide which rule, practice, or pronouncement should predominate? The AICPA, in Statement on Auditing Standards No. 69, “*The Meaning of Present*

Fairly in Conformity with Generally Accepted Accounting Principles in the Independent Auditor’s Report,” provided finance officers with guidelines to use in determining when a set of rules takes precedence over all the others.^[4] The hierarchy that should be followed in the selection process for state and local governments and for non-government organizations is shown in Table 1. An amendment to SAS No. 69, SAS No. 91, *Federal GAAP Hierarchy*, provides a hierarchy for federal governmental entities as illustrated in Table 2.^[5] In all three cases, Level 1 sources, which primarily include official pronouncements of the governing board, take precedence over all other possible GAAP sources. If no Level 1 source is available, Level 2 sources are consulted, then Level 3, Level 4, and, finally, Level 5.

In each case, official pronouncements from the governing organizations take precedence over those issued by other organizations. For instance, state and local governments must follow relevant GASB pronouncements, unless specifically approved by GASB. If no pronouncements exist on the subject, other sources, including FASB pronouncements, can be utilized.

EXAMPLES OF GAAP FOR STATE AND LOCAL GOVERNMENTS

The following sections describe a few of the accounting rules and regulations considered GAAP for state and local governments.

Table 1 GAAP hierarchy summary

State and local governments	Non-governmental entities
1. GASB Statements and Interpretations; AICPA and FASB pronouncements if made applicable to state and local governments by the GASB	1. FASB Statements and Interpretations, APB Opinions, and AICPA Accounting Research Bulletins
2. GASB Technical Bulletins; AICPA Industry Audit and Accounting Guides and Statements of Position if made applicable to state and local governments by the GASB	2. FASB Technical Bulletins, AICPA Industry Audit and Accounting Guides, and AICPA Statements of Position
3. Consensus positions of the GASB Emerging Issues Task Force (when created); AICPA Practice Bulletins if made applicable to state and local governments by the AICPA	3. AICPA Practice Bulletins and consensus positions of the FASB Emerging Issues Task Force
4. Implementation guides published by the GASB staff; widely recognized and prevalent industry practices of state and local governments	4. AICPA accounting interpretations, “Qs and As” published by the FASB staff, as well as industry practices widely recognized and prevalent
5. Other accounting literature, including GASB Concepts Statements; FASB Concept Statements; AICPA Issues Papers; International Accounting Standards Committee Statements; pronouncements of other professional associations or regulatory agencies; AICPA Technical Practice Aids, and accounting textbooks, handbooks, and articles	5. Other accounting literature, including FASB Concepts Statements; APB Statements; AICPA Issues Paper; International Accounting Standards Committee Statements; GASB Statements, Interpretations, and Technical Bulletins; pronouncements of other professional associations or regulatory agencies; AICPA <i>Technical Practice Aids</i> ; and accounting textbooks, handbooks, and articles

(From SAS No. 69, *The Meaning of Present Fairly in Conformity with Generally Accepted Accounting Principles in the Independent Auditor’s Report*.)

Table 2 Federal GAAP hierarchy

1. FASAB Statements and Interpretations plus AICPA and FASB pronouncements if made applicable to federal governmental entities by a FASAB Statement or Interpretation
2. FASAB Technical Bulletins and the following pronouncements if specifically made applicable to federal governmental entities by the AICPA and cleared by the FASAB: AICPA Industry Audit and Accounting Guides and AICPA Statements of Position
3. AICPA Accounting Standards Executive Committee (AcSEC) Practice Bulletins if specifically made applicable to federal governmental entities and cleared by the FASAB and Technical Releases of the Accounting and Auditing Policy Committee of the FASAB
4. Implementation guides published by the FASAB staff and practices that are widely recognized and prevalent in the federal government
5. Other accounting literature, including FASAB Concepts Statements; FASB Concepts Statements; GASB Statements, Interpretations, Technical Bulletins, and Concepts Statements; AICPA Issues Papers; International Accounting Standards of the International Accounting Standards Committee; pronouncements of other associations or regulatory agencies; AICPA *Technical Practice Aids*; and accounting textbooks, handbooks, and articles

(From SAS No. 91, Federal GAAP Hierarchy.)

Fund Accounting

Because accountability is a key component in government financial reporting, state and local governments are strongly encouraged to use funds in the accounting for resources. Used to demonstrate compliance with legal and contractual requirements, a fund is a separate accounting and fiscal entity with a self-balancing set of accounts from which financial statements can be prepared. Because the use of funds provides for more control over resources, most state and local governments utilize more than one fund, depending on the source and use of those resources. Essentially, each fund maintains its own set of accounting records, interacting with other funds whenever appropriate. Further, because each fund maintains a separate set of accounts, each fund also generates its own set of financial statements.

Three main categories of funds exist: governmental, proprietary, and fiduciary. Which fund a government uses is determined by the source and use of the resources involved. Governmental funds are utilized for basic government activities financed primarily from taxes and intergovernmental grants. Proprietary funds are for business-type activities that generate most resources through the collection of user fees. Fiduciary funds account for resources held by the government, or its agent, in trust for a third party.

Resource flows for the basic operations of state and local governments are accounted for in a governmental fund. In particular, the majority of the resources are accounted for in the general fund, a subfund of the governmental funds. Typically, the general fund is the largest and most active of all funds. For some small governments, it could be the only fund utilized. General fund resources that have been legally restricted for specific purposes typically are accounted for as a special revenue fund, another subfund of the

governmental funds. A government that provides services to the public at large and charges a fee for the services could account for the resources in an enterprise fund, a subfund of the proprietary funds. A separate enterprise fund would be used for each “business,” maintaining increased control and responsibility over the fund’s resources. An agency fund, a subfund of the fiduciary funds, may be used to account for sales tax monies collected by a city for the state government. Soon after the monies go into the fund, they are transferred out to the state government.

Modified Accrual Accounting

Because governments are concerned with how resources are consumed, and not the bottom line, the basis for accounting for those resources differs from that of for-profit organizations. For-profit organizations utilize full accrual accounting that requires an entity to recognize revenue when it has been earned, that is, when the related goods and services have been provided. The revenues also must be measurable, i.e., the amount of the revenue is known. In addition, full accrual accounting requires the recognition, or recording, of an expense, whenever a cost is incurred that helps to generate revenue. This provides for an appropriate matching of revenues and expenses in the financial statements and results in a measurement of income for a specified period.

Under modified accrual accounting, revenue recognition follows that of full accrual accounting, with one major caveat. In addition to being earned and measurable, revenue must be available to finance expenditures of the current fiscal period. The GASB Statement No. 33, *Accounting and Reporting for Nonexchange Transactions*, stated “available means collectible within the current period or soon enough

thereafter to be used to pay liabilities of the current period.”^[6] Full accrual accounting only requires a reasonable assurance that the cash will be collected, while modified accrual accounting provides greater assurance that the cash will be available to pay liabilities when they come due. So how does a government interpret “available”? GASB only provided a specific definition of “available” for property tax revenues. Existing standards, in particular, GASB Interpretation No. 5, *Property Tax Revenue Recognition in Governmental Funds*, provided that if property taxes are collected more than 60 days after the end of the fiscal year, they may not be recognized in that fiscal year, because they cannot be used to pay the current liabilities.^[7] They must be recognized in the year of the collection. Because no specific guidelines exist for other types of revenues, some governments have established time periods such as 30 days, 60 days, or longer for their recognition.

The recognition of expenses also differs under modified accrual accounting. For instance, governmental funds following modified accrual accounting recognize expenditures, not expenses. Expenditures are defined as decreases in net current financial resources, such as cash, while expenses are a reduction in overall net assets, such as the depreciation of a building or the use of supplies. Existing standards require the recognition of an expenditure when cash is paid out or when a liability has been incurred for goods or services received. For instance, if a local government buys a new fire truck, whether with cash or on credit, an expenditure is recognized immediately. An asset may be recorded but in a separate fund from the expenditure. A for-profit organization would recognize an asset and then expense a portion of the cost each year for the remainder of the truck’s economic useful life.

The Governmental Reporting Entity

Many state and local governments are affiliated with independent legal entities that provide them with essential services the governments do not provide for themselves. For instance, a local government may determine a need for a wastewater treatment plant. Once it is established, the governing board could choose to operate the plant as an integral part of the government and account for it as an enterprise fund. Recall that an enterprise fund is a government business that charges a fee for its services. Under this approach, any debt entered into by the plant would affect the debt levels of the whole government—not just the treatment plant. If the wastewater treatment plant was established as an independent legal entity, however, and incurred its own debt, the direct financial effect on the primary government may be minimal.

Obviously, a plant treated as an enterprise fund would be included in a government’s financial statements. But what about the independent legal entity? If the entity was established by the government, should the plant be included as part of the government’s financial statements? GASB Statement No. 14, *The Financial Reporting Entity*, was issued to help resolve potential problems of this type, and it specified criteria for determining what economic entities should be included in a reporting entity.^[8]

Specifically, a financial reporting entity consists of a primary government and any component units associated with that primary government. GASB No. 14 defines a primary government as “any state government or general-purpose local government (municipality or county).”^[8] A component unit is a legally separate government for which the primary government is financially accountable or an organization whose relationship with the primary government is such that if it were excluded from the primary government’s financial statements, they would be misleading and incomplete.

Once a component unit has been determined to be part of the reporting entity, GASB No. 14 also provides criteria as to how it should be included in the financial statements. Depending on the relationship between the two entities, a component unit is blended into or discretely presented on the financial statements of the primary government. A very close relationship between the two entities results in a set of blended financial statements. A superficial evaluation of the blended statements would not suggest the existence of a component unit, because the financial information of the component unit is fully integrated into the data of the primary government. The financial data for a discretely presented component unit, however, is reported on the face of the financial statements in a separate column from the primary government’s information with a total column used to indicate the overall effect of the two consolidated entities.^[8]

The Financial Reporting Model

In June 1999, GASB issued what many consider to be one of the most important statements in the history of governmental accounting. GASB Statement No. 34, *Basic Financial Statements—and Management’s Discussion and Analysis—for State and Local Governments*, radically changed the way financial statements for state and local governments are presented.^[9] Previously, governments prepared individual financial statements for each fund category. With the implementation of GASB No. 34, governments still prepare fund financial statements, but government-wide statements also must be generated. Government-wide statements

are designed to give users an overview of the financial position of the government by consolidating the governmental and business-type activities into a statement of net assets and a statement of activities. In addition, these statements must be presented using full accrual accounting, while fund-based financial statements follow full or modified accrual accounting, depending on the fund. Because the governmental financial information is presented using both the full accrual and the modified accrual bases of accounting, a reconciliation between the two is required to be presented.^[9]

GASB No. 34 also dramatically changed how a government reports its infrastructure, which is defined as long-lived capital assets such as bridges, roads, and tunnels that have economic useful lives significantly longer than most other capital assets. Previously, governments were not required to report this information in their financial statements. The adoption of GASB No. 34, however, results in the inclusion of infrastructure on a government's financial statements. Because many governments may not have maintained accurate records of the cost of their infrastructure, GASB No. 34 allows the use of estimates for retroactive reporting of infrastructure. New assets must be reported at actual cost.^[9]

In addition, governments must record depreciation for their infrastructure on the government-wide financial statements. Two options for depreciation exist. First, a government may record depreciation in a manner similar to other capital assets—over a fixed period of time using a method such as straight-line depreciation. Each year, depreciation expense and accumulated depreciation would be recorded for the infrastructure. Or, a government may choose to use the modified approach to infrastructure reporting. Essentially, if a government chooses the modified approach, it chooses not to depreciate. However, the government is committed to the maintenance of accurate, up-to-date inventory records of its infrastructure and is committed to preserving the assets at a prescribed condition level. Any expenditures made to extend the life of the infrastructure would be charged to expense and recognized in the period incurred.^[9]

CONCLUSION

In simple terms, GAAP is a set of rules for reporting financial information. Whether for a governmental, non-profit, or a for-profit entity, financial statements prepared following GAAP provide users with assurances as to the reliability of the information and give preparers some guidelines to follow in the creation of those reports.

REFERENCES

1. GASB Concepts Statement No. 1. *Objectives of Financial Reporting*; Governmental Accounting Standards Board, 1987.
2. FASB Statement of Financial Accounting Concepts No. 1. *Objectives of Financial Reporting by Business Enterprises*; Financial Accounting Standards Board, 1978.
3. Statement of Federal Financial Accounting Concepts No. 1. *Objectives of Federal Financial Reporting*; Federal Accounting Standards Advisory Board, 1993.
4. AICPA Statement on Auditing Standards No. 69. *The Meaning of Present Fairly in Conformity with Generally Accepted Accounting Principles in the Independent Auditor's Report*; American Institute of Certified Public Accountants, 1992.
5. AICPA Statement on Auditing Standards No. 91. *Federal GAAP Hierarchy*; American Institute of Certified Public Accountants, 1999.
6. GASB Statement No. 33. *Accounting and Financial Reporting for Nonexchange Transactions*; Governmental Accounting Standards Board, 1998.
7. GASB Interpretation No. 5. *Property Tax Revenue Recognition in Governmental Funds—An Interpretation of NCGA Statement 1 and an Amendment of NCGA Interpretation 3*; Governmental Accounting Standards Board, 1997.
8. GASB Statement No. 14. *The Financial Reporting Entity*; Governmental Accounting Standards Board, 1991.
9. GASB Statement No. 34. *Basic Financial Statements—And Management's Discussion and Analysis—For State and Local Governments*; Governmental Accounting Standards Board, 1999.

Geographic Information Systems and Public Policy

Akhlaque Haque

*Department of Government, University of Alabama at Birmingham,
Birmingham, Alabama, U.S.A.*

INTRODUCTION

Geographic information systems (GIS) are computer-based information management applications for collecting, storing, managing, retrieving, mapping, and analyzing spatially referenced data. They are powerful tool in facilitating quick and interactive visualization of objects or geographical phenomenon over a computer screen and are useful in solving many planning and management problems. By incorporating space or location into decision-making models and displaying the attributes or characteristics of the space, GIS provide policy analysts with new tools of communicating data through maps.

Although GIS's primary purpose has been to understand the spatial distribution of data in question, it finds its most use in decision making, especially for solving complex problems with nondeterministic outcomes. Because public policy problems are often multidimensional and semistructured, GIS can be a valuable tool in exploring and explaining policy outcomes for a given space. Whereas, quantitative statistical data may reveal nuances between and within variables in a model, a map of the data, on the other hand, concisely communicates spatial distribution, enabling the viewer to better understand patterns and relationships through scenarios (of maps). For example, GIS can be used to explore the crime pattern in a given location by investigating the type of criminal activity and explore the socioeconomic characteristics that attract and/or lead to criminal behavior in that location. Furthermore, the same GIS model for the area can be extended to address policy questions related to race relations, substance abuse, education, or health. As an enabler of sorts, "GIS is both a tool box for advanced analysis and capabilities and a sandbox to express our creativity."^[1] Today GIS is also being viewed as an important group-decision-support tool allowing differing views about a project to be portrayed in a set of maps discussed among stake holders.^[2] The GIS support both exploratory and confirmatory analysis, provide tools for both inductive and deductive approaches, and support both scientific research and the implementation of public policy based on GIS models.

The purpose of this entry is to provide an overview of GIS as it relates to public policy decision making.

The first section describes how GIS works followed by particular examples of GIS in public policy. A section describes why GIS is an important tool for public policy analysis with the follow-up section describing where GIS finds its most use. The final sections describe the implications of GIS in a representative democracy.

HOW GIS WORKS

Conceptually, GIS can be regarded as a decision-support tool that combines the art and science of decision making to represent the real world over a computer screen. The process involves transforming the paper maps into digital format and, using one of the geographical models, mathematically manipulating the three-dimensional earth phenomenon on to a computer screen. The GIS applications boasts in transforming each and every point on earth into an interactively manageable database that allows any location or object on earth on a computer screen that can be digitally manipulated to bring about any possible scenario related to that space. For end users, the most expensive and difficult part of GIS has been obtaining the digitized maps. In recent years the Internet has become the primary resource for such maps. Some of the most commonly used maps are available free of charge from the Bureau of Census and GIS software developers.

The operation of GIS in a nutshell can be described using a schematic diagram. A typical GIS operation as it relates to a data model that is obtained from the real world and expressed in digital format on a computer screen is shown in [Fig. 1](#). For computer processing, the digital information is transformed to machine language (M1, M2).

For example, to represent a geographic phenomenon such as tornado touch-downs, GIS software combines the spatial data (location of tornado touch-downs—latitude and longitude) with the attributes of the spatial location (residential area, forest, highway, river banks, or flood zones) within maps. The final outcome of this mapping scheme, for example, could be the location of tornado touch-downs in low populated residential areas ([Fig. 2](#)). The figure shows tornado

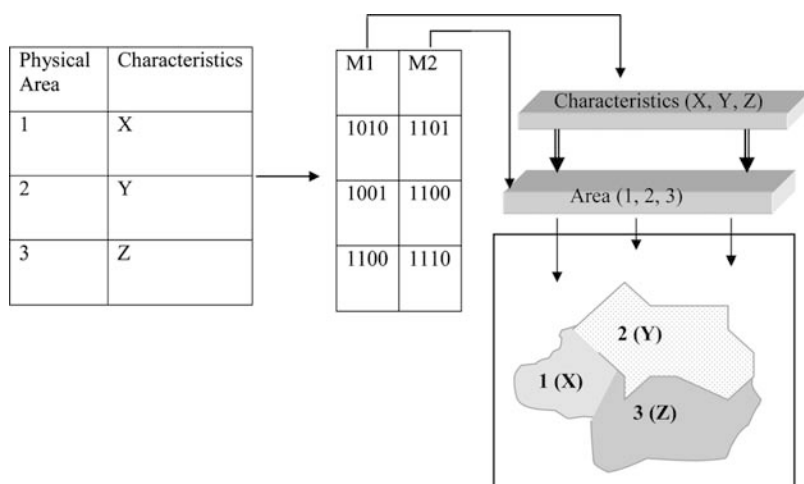


Fig. 1 Schematic diagram of a digital data model of GIS. (From Haque, A. Geographic information systems and decision making. In *Handbook of Decision Making*; Goktug, M., Ed.; CRC Press, 2006.)

touch-downs recorded from May to June 2005 in Eastern Colorado. The base-map shows number of households (darker color showing more households) by census tract along the interstate. Using other census data the map could also be used to reveal information about the people in the affected area, including income level, number of school-going children, age distribution and so on. Such analysis helps disaster management team prepare for recovery and mitigate future catastrophes expected along rural tornado paths. The critical factor is having the right data to show the data model as expressed through the maps.

Once the information is displayed on the screen, decision makers can assess the patterns to further understand the phenomenon and accordingly prioritize areas for future resource allocation and target-based planning. Although GIS technology greatly enhances our decision-making capabilities, it does not replace them. As with any research method tool and more so for GIS understanding the data and the field becomes a necessary condition for the effective use of GIS in policy implementation and planning.

Visual displays generated by GIS software augment decision making by giving added information to data.

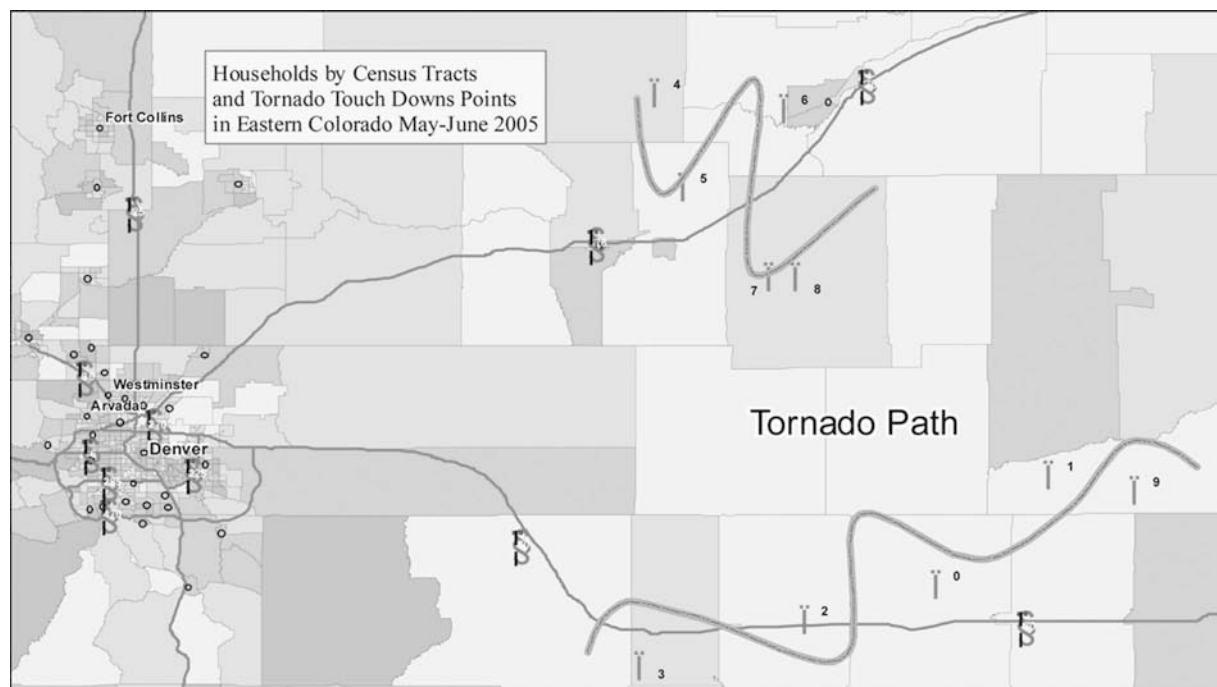


Fig. 2 Tracking Tornado path: touch-down locations and number of households by census tract. Tornado point data. (From <http://www.tornadochaser.net/2005.html>; Map produced using ArcGIS 9.0 by Environmental Systems Research Institute (ESRI).)

The quality of decision making is expected to be better not because GIS software generate more information, but because people can relate to the situation or problem in the real world through visual inspection. For example, data for school children by age and their nutrition level for each city in a county can be mapped to reveal the neighborhoods where the problem of nutritional deficiency among children is prominent. The same analysis can be done for any neighborhood, and the decision maker can relate to his/her own neighborhood. The value of cartographic visualization may indeed add knowledge to traditional approach to decision making.

GIS AND PUBLIC POLICY

According to Bolstad GIS is a computer-based system designed to aid the collection, maintenance, storage, analysis, output, and distribution of spatial data and information.^[3] The foundation of GIS involves both the creation and the maintenance of a relational database. Relational databases are characterized by their ability to combine information from different areas and relate them to each other. The GIS is not viewed as a stand-alone application but as an integrated system that is based on geographic information science and supported by technological innovations made in different disciplines including geography, cartography, photogrammetry, remote sensing, surveying, geodesy, civil engineering, statistics, computer science, operations research, artificial intelligence, demography, and many other branches of the social sciences, natural sciences, and engineering. In fact, the power of GIS lies in its interdisciplinary character and appeal. It provides unprecedented opportunities for collaboration among various governmental, nongovernmental, and private agencies. In addition through interactive visualization and web-enabled GIS, citizens now have the chance to participate in decision making and, at least, visually comprehend how various public policy decisions are affecting them. It is not surprising, whereas in 1992, 40% of the local governments used GIS, in 2003, according to a survey sponsored by the U.S. Department of Interior and conducted by the Public Technology Inc., 97% of the local governments with populations 100 thousand and 88% of those between populations 50 and 100 thousand use GIS technology.^[4] Today, GIS appears as a given tool for emergency dispatch operations, land use planning including zoning, health-care planning, infrastructure planning and tax revenue assessment, political campaigns, reapportionment of congressional districts, and other mandated rezoning operations. A sample list of GIS use by different industries is highlighted in Table 1.

Table 1 GIS by industry type

Industry	Use of GIS
Forestry	Inventory and management of resources
Police	Crime mapping to target resources
Epidemiology	To link clusters of disease to sources
Transport	Monitoring routes
Utilities	Managing pipe networks
Oil	Monitoring ships and managing pipelines
Central and local government	Evidence for funding and policy, e.g., deprivation
Health	Planning services and health impact assessments
Environment agencies	Identifying areas of risk from, e.g., flood
Emergency departments, e.g., ambulance	Planning quickest routes
Retail	Store location
Marketing	Locating target customers
Military	Troop movement
Mobile phone companies	Locating masts
Land registry	Recording and managing land and property
Estate agents	Locating properties that match certain criteria
Insurance	Identifying risk, e.g., properties at risk of flooding
Agriculture	Analyzing crop yields

[From Getting Started with GIS. (n.d.). Who uses GIS and Why. <http://www.gis.rgs.org/10.html> (accessed April 2006).]

Policy Analysis Using GIS

Public policy problems are semistructured—they require thorough exploration before they can be understood and viable solutions can be proposed. Decision makers who have different expertise and goals in finding solutions to the problem can take advantage of GIS tools to interactively communicate and better understand the competing choices available to them. The effectiveness of GIS as a decision-making tool relies heavily on collaboration. Especially, for public policy making, such efforts are greatly enhanced by the capabilities GIS provide for collective decision making. It commonly involves three categories of participants in the decision-making process: 1) stakeholders, 2) decision makers, and 3) technical specialists. The stakeholders or the participants who directly or indirectly will be affected by the decision, make up, what can be called, the *GIS User Group*. The user group

plays a critical role in *formulating* the plan using set of maps showing the decision outcomes. People who are primarily involved in creating the maps, i.e., technical experts who manage and manipulate the data to accurately represent the real world on a computer screen, form the *GIS Support Group*.

The value of GIS in a democracy may not be in its technological supremacy rather in its effectiveness as an intermediary (as a tool) that resolve conflicts and seek for compromising options among competing powers. It can be designed to allow different participants to specify and apply their criteria using the relative weights associated with each model displayed in maps. The solution from each model then can be summarized through a GIS-based data management report that highlights the characteristics of each solution. The collaborative nature of GIS tools has attracted a new group of GIS practitioners who envision GIS as a participatory tool for decision making. Their approach is commonly known as the Public Participation GIS (PPGIS). The PPGIS seek a method within broader GIS scheme to use GIS as an empowering tool by making it accessible to general public and enabling communities to participate in critical decisions in housing, zoning, health, racism, environmental equity, and sustainable developmental projects.^[2] As computers become more widespread and economical, average people is expected to employ GIS technology to amplify their voices as it relates to their own governance. The optimism is clearly due to the communicating capability of GIS through “What You See is What You Get” (WYSWYG) approach.

Although maps are traditionally used for communication (generally understood by a lay person), for policy analysis, visualization techniques will have to take precedence over communication to reveal subtleties in data. In other words, scientific visualization approach must be followed in order to understand the spatial relations that common eye may not reveal. Scientific visualization allows users to interpret, validate, and explore their data in much greater detail. The visualization technique can be described as intuitive science—extracting information for hypothesis generation and policy evaluation. According to Dorling and Fairbairn the science of visualization is “process of learning through the creation and observation of abstract images, providing a method for seeing the unseen.”^[5]

Consistent with Simon’s^[6] view of three stages of decision making, namely 1) gathering intelligence on decision criteria, 2) designing options to choose from, and 3) the choice; decision makers can utilize GIS maps to: a) formulate the question or problem; b) explore and evaluate alternative plan of action (for solving the problem), and c) evaluate a reduced set of competing alternatives. In the absence of a rigorous methodological

“approach to policy analysis, when using GIS, visualization of maps can be used along with the policy cycle.” A policy cycle can be described in four stages: selecting a policy, implementing the policy, understanding the issues, and evaluating the options. In all of the four stages of the policy cycle GIS can be used. “Solutions from each stage can further improve each stages of the policy cycle. The outcomes from each stage can be investigated through a continuous cycle of policy selection and evaluation.

Although the analysis of maps may differ among investigators, a formal process of map generation is in order. The leading GIS software developer Environmental Systems Research Institute (ESRI) recommends simple five steps for mapping:^[7]

1. Frame the question.
2. Select your data.
3. Choose the analysis method.
4. Process the data.
5. Look at the results.

The third step of choosing the map analysis method will differ depending on the level of spatial and cartographic technical knowledge of the user. Some of the more complex policy analysis may require advanced spatial statistical knowledge. However, today most GIS applications have been incorporating spatial statistical techniques for scientific visualization of maps.

IMPLICATIONS OF GIS IN PUBLIC POLICY

The GIS is explicitly employed in the development and implementation and evaluation of public policy, in redistribution of funds, and for identifying communities for target-based intervention. The use of GIS in political reapportionment and redistricting provides powerful evidence of its value as the most efficient tool for that purpose. At the same time, the subsequent abuse of this tool reminds us the dangers and responsibilities of using GIS in a representative democracy (see Refs.^[8,9]). Especially, since September 11, 2001, GIS has taken a new role in providing high-tech GIS surveillance capabilities to the Department of Homeland Security that raises concern about violation of citizen’s rights and civil liberties (see Refs.^[10,11]). Indeed, GIS promises abound so are the challenges. In a more positive side, however, as GIS becomes a common place in the academia including high schools, people will become more aware of the known abuses of GIS and subsequently make GIS users more responsible. In order to reap the benefit of this state-of-the-art technology, it would be foolish to think otherwise. For the next generation of policy analyst in the 21st century this will be

another positive step toward making better policy decisions.

CONCLUSION

Public policy analysis requires a multidimensional approach. No single methodological approach can be best suited for public policy analysis nor can any tool claim to have the magic wand. GIS provides another methodological approach that may have bearing for the future generation of public policy makers and their decision outcomes. Given the built-in communicative power of GIS and proliferation of the Internet, it can have far reaching implications for influencing public policy. One major barrier, however, is the lack of interoperability among GIS and geographic information technologies themselves, and between GIS and other information technology. Moreover, once a GIS project is completed by a governmental agency, it imposes contractual or licensing conditions restricting the use of the data. The reluctance to share the information is in part due to the fear of violating data privacy laws but more for the fear of losing control of the information.^[12] To make best use of this tool, open data sharing policies have to be adopted and GIS should be widely accessible beyond technical experts. In this regard the efforts made by the National Spatial Data Infrastructure (NSDI) through the Federal Geographic Data Committee (FGDC) can have important ramifications for the future of GIS as a public policy-making tool.

ACKNOWLEDGMENTS

I am thankful to Goktug Morcol for his insightful editorial comments on an earlier draft of the manuscript that appears in *The Handbook of Decision Making*; Goktug, M., Eds. CRC Press: Boca Raton,

FL, 2006. I am also thankful to anonymous reviewers for their thoughtful comments and suggestions to improve the manuscript.

REFERENCES

1. Berry, J.K. *Beyond Mapping: Concepts, Algorithms, and Issues in GIS*; GIS World Books: Fort Collins, CO, 1993.
2. Craig, W.J.; Harris, T.M.; Weiner, D., Eds. *Community Participation and Geographic Information Systems*; Taylor and Francis: London, 2002.
3. Bolstad, P. *GIS Fundamentals: A First Text on Geographic Information Systems*; Eider Press: White Bear Lake, MN, 2002; 1 pp.
4. Haque, A. Decision making in geographic information systems. In *Handbook of Decision Making*; Goktug, M., Ed.; Marcel Dekker: New York, Vol. 123, 2006.
5. Dorling, D.; Fairbairn, D. *Mapping: Ways of Representing the World*; Addison Wesley Longman Limited: London, 1997; 102 pp.
6. Simon, H. Rational decision making in business. *Am. Econ. Rev.* **1979**, *69*, 493–513.
7. <http://www.gis.com> (accessed February 2006).
8. Monmonier, M. *Bushmanders and Bullwinkles: How Politicians Manipulate Electronic Maps and Census Data to Win Elections*; University of Chicago Press: Chicago, 2001.
9. Pickles, M. *Ground Truth: The Social Implications of Geographic Information Systems*; Guilford Press: New York, 1995.
10. Monmonier, M. *Spying with Maps: Surveillance Technologies and the Future of Privacy*; University of Chicago Press: Chicago, 2002.
11. Haque, A. GIS, public service and the issue of democratic governance. *Public Admin. Rev.* **2001**, *61*, 259–265.
12. Onsrud, H.J. Identifying unethical conduct in the use of GIS. *Cartography Geographic Inform. Systems* **1995**, *22* (1), 90–97.

Global Ethics and Corruption

Leo W.J.C. Huberts

Department of Public Administration and Communication Science, Vrije Universiteit,
Amsterdam, The Netherlands

INTRODUCTION

Corruption, ethics, and integrity have become important issues in the practice and theory of politics and public administration. This has led to more awareness and knowledge of the ethical or moral dimension of politics and the causes of and solutions for corruption in politics and administration at the national as well as the international levels. Elements of that knowledge and awareness to be discussed include the meaning of concepts, such as corruption, ethics, and integrity; the content of the moral dimension (the values at stake); the causes of (the level of) corruption in different nations; the solutions proposed to curb corruption and to safeguard integrity; and the initiatives by international actors that aim to contribute to that purpose.

CORRUPTION, ETHICS, AND INTEGRITY

Clarity about concepts like corruption, ethics, and integrity is important, certainly when it concerns public debate, policy making, and theory development on an international level.^[1-3]

Public corruption is often defined as involving behavior on the part of officials in the public sector, whether politicians or civil servants, in which they improperly and unlawfully enrich themselves, or those associated with them, by misusing the public power entrusted to them. A brief definition is the abuse of public office for private gain. Corruption, then, is a specific type of violation against the moral norms and values for political and administrative behavior. *Ethics* refer to the collection of values and norms, functioning as standards or yardsticks for assessing the integrity of one's conduct. The moral nature of these principles refers to what is judged as right, just, or good (conduct). Values are principles or standards of behavior that should have a certain weight in choice of action (what is good to do or bad to omit doing). Norms state what is morally correct behavior in a certain situation. Values and norms guide the choice of action and provide a moral basis for justifying or evaluating what we do.

Public integrity denotes the quality of acting in accordance with the moral values, norms, and rules accepted by the body politic and the public. A number

of integrity violations or forms of public misconduct can be distinguished: corruption including bribing, nepotism, cronyism, and patronage; fraud and theft; conflict of interest through assets, jobs, or gifts; manipulation of information; discrimination and sexual harassment; improper methods for noble causes (using immoral means to achieve moral ends); the waste and abuse of resources; and private time misconduct.

In this framework, corruption is a type of integrity violation; however, "corruption" is often also used as the umbrella concept, covering all or most types of integrity violation or unethical behavior.

VALUES AND NORMS TO BE PROTECTED

There are numerous attempts to sum up the basic values that might be seen as constituting public-sector ethics.^[4,5] A well-known ethical framework for public officials was developed in the United Kingdom by the Committee on Standards in Public Life chaired by Lord Nolan.^[4] The Nolan Committee sketched "Seven Principles of Public Life." Holders of public office should decide because of the public interest; private interests or obligations to outside individuals and organizations should have no influence (selflessness and integrity). They should make choices on the basis of merit (objectivity); they are accountable for their decisions and actions (accountability); and they should be as open as possible (openness and transparency). Holders of public office have a duty to declare any private interests and resolve possible conflicts of interest (honesty). They should promote and support these principles by leadership and example (leadership). The Nolan values have been discussed and used in many contexts, for example by the Committee of Independent Experts that reported on fraud, mismanagement, and nepotism in the European Commission of the European Union in 1999 (leading to the resignation of the whole commission in March 1999). Many of the values mentioned can also be found in the codes of conduct of governmental organizations as well as in the United Nations International Code of Conduct for Public Officials. The latter contains a set of basic standards of integrity and performance expected from public officials, with "wide acceptance on the basis of

experience acquired in various countries.” The code includes general principles: public officials shall act in the public interests, function efficiently and effectively, in accordance with the law and with integrity, and shall be attentive, fair, and impartial. The code also contains rules concerning conflicts of interest and disqualification, disclosure of assets, acceptance of gifts and other favors, confidential information, and political activity.

Discussion

An important question is always how important the values actually are for the people working in the organizations with often impressive mission statements and codes of conduct. When politicians tell researchers that honesty and openness are among the values that are most important for their work, this, of course, does not mean these indeed are the central values in actual political decision making.

In the code of the United Nations, values such as efficiency and effectiveness are added to the values more often associated with “moral” values and norms like honesty, selflessness, openness, and accountability. The relationship between these two sets of values has long been a topic in the international political as well as scientific debate. The dominant view in the global discussion nowadays is that state corruption undermines state efficiency and that good governance is a precondition for financial support and developmental aid. On the national level, a comparable view is dominant, although research shows that anticorruption policies and regulation can also harm efficient and credible government.

Another subject to be mentioned is the relationship between the core values of the public and the private realm. How do public- and private-sector values relate to one another? The answer is not an easy one. In the literature, contradictory visions exist. On the one hand, there is the thesis that public and business sector ethics share basic values and norms (as integrity, honesty, and efficiency).^[6] On the other hand, there is the opposite view that stresses that there is a fundamental conflict between the moral foundations of politics and commerce and that organizations will sink in “functional and moral quagmires . . . when they confuse their own appropriate moral system with the other.”^[7] Warnings against the confusion of morals often lead to doubts about practices from the commercial world when these practices are to be applied in the public sector.

EXTENT OF THE PROBLEM IN A GLOBAL CONTEXT

Empirical research by social scientists on the extent of public corruption and fraud is, by definition,

complicated. Corruption is crime without a (recognizable) victim, and the corruptor and the corrupted benefit from secrecy. Most empirical data, therefore, concern criminal cases involving corruption: what has been discovered, investigated, prosecuted, and convicted can most easily be counted. Additionally, information has been collected on opinions on the extent and seriousness of the corruption problem. The most famous research has been done by Göttingen University and Transparency International. Since 1995, a corruption perception index (CPI) is published, based on the perceptions of the degree of corruption as seen by businesspeople, risk analysts, and the general public. Some countries have the reputation of being very corrupt, and as might be expected, most of these are developing countries like Nigeria and Indonesia or transitional (Eastern European) states like the Ukraine. Northern European countries are perceived as being among the less corrupt (Table 1).

The image of countries appears to be rather stable. Once a reputation for corruptness is established, it takes time and efforts to change (Table 2).

Corruption in Business and Government

Corruption research often is research on public-sector corruption. This suggests that the problem is more prominent in government than in business. However, this idea needs to be questioned.

First, it is clear that corruption involves a corruptor and a corrupted. Many bribe givers are business organizations. In business, the possibility of profit is often more important than integrity and ethics, which until recently, was supported by governments of industrialized countries (the costs of bribes have long been tax deductible).

The Transparency International Bribe Payers Survey concerns the willingness of companies from leading exporting states to pay bribes to public officials to win or retain business. It shows that companies from countries like China, South Korea, Taiwan, Italy, and Malaysia are most likely to pay bribes. Additionally, the survey shows that the scores for countries like the United Kingdom, the United States, and Japan are lower than the corruption scores for the countries. Bribing a foreign official seems more acceptable than bribing at home.

Second, survey research shows that corruption experts in higher-income countries differ from their colleagues in poorer countries concerning the prominence of corruption in the public and private sectors. Respondents from the lower-income countries think corruption is much more prominent in the public sector, while experts from higher-income countries find the opposite to be the case.

Table 1 Corruption perceptions index—transparency international, 2001 (10 highly clean—1 highly corrupt)

Country	2001 CPI score
<i>Least Corrupt Countries</i>	
Finland	9.9
Denmark	9.5
New Zealand	9.4
Iceland	9.2
Singapore	9.2
Sweden	9.0
Canada	8.9
Netherlands	8.8
Luxembourg	8.7
Norway	8.6
<i>Most Corrupt Countries</i>	
Bangladesh	0.4 ^a
Nigeria	1.0
Uganda	1.9
Indonesia	1.9
Kenya	2.0
Cameroon	2.0
Bolivia	2.0
Azerbaijan	2.0
Ukraine	2.1
Tanzania	2.2
<i>Other Countries</i>	
United Kingdom	8.3
United States	7.6
Japan	7.1
South Africa	4.8
Brazil	4.0
China	3.5
India	2.7

^aTransparency International adds that the Bangladesh data should be viewed with extra caution.

Source: Transparency International and Göttingen University, <http://www.gwdg.de/~uwvw> (August 2001); also: <http://www.transparency.org/documents/cpi/2001> (27 June 2001).

CAUSES OF CORRUPTION

Research shows that a conglomerate of social, economic, political, organizational, and individual causal factors are important to explain cases of public corruption and fraud in a country.^[2,8,9]

Among the more important political causes are the values and norms of politicians and civil servants and their commitment to public integrity, the organizational quality of the public sector (working conditions,

Table 2 Corruption perceptions index, 1988–2000 (10 highly clean—1 highly corrupt)

	1988–1992	1996	1998	2000	2001
Sweden	8.7	9.1	9.5	9.4	9.0
Australia	8.2	8.6	8.7	8.3	8.5
United Kingdom	8.3	8.4	8.7	8.7	8.3
United States	7.8	7.7	7.5	7.8	7.6
Japan	7.3	7.1	5.8	6.4	7.1
Brazil	3.5	3.0	4.0	3.9	4.0
China	4.7	2.4	3.5	3.1	3.5
India	2.9	2.6	2.9	2.8	2.7
Indonesia	0.6	2.7	2.0	1.7	1.9
Nigeria	0.6	0.7	1.9	1.2	1.0

Source: Transparency International and Göttingen University, <http://www.gwdg.de/~uwvw> (August 2001).

control, and auditing), the relationship between the state and the business sector, and a number of social factors (e.g., the presence of organized crime and the content of social norms and values).

These factors are important to explain the corruption in higher- and lower-income countries, but an ample explanation also has to take into account the differences between these countries. A number of factors related to developmental problems are of crucial importance to lower-income countries.^[10] Economic failure and poverty, poor conditions for human development (schooling, health), and corruption are inter-related. For example, it is clear that low salaries and bad working conditions in the public sector can be disastrous for the possibility and the willingness to behave ethically.

For countries going through a process of privatization, liberalization, and democratization, the period of transition offers extra possibilities for corruption.^[8,10] It takes time to establish a more stable political and economic system that is able to curb corruption.

SOLUTIONS AND THEIR EFFECTIVENESS

It is always necessary to relate anticorruption strategies to characteristics of the actors involved (and the environments in which they operate). There is not one concept and program of good governance for all countries and organizations, there is no “one right way.”^[10,11] There are many initiatives, and most are tailored to specific contexts; societies and organizations will have to seek their own solutions.

Yet, they can learn from the experiences of others and the international search for practices based on serious and well-evaluated programs. There are

examples of organizations and societies that succeeded in radically curbing corruption. Hong Kong's Independent Commission Against Corruption is often presented as an example for other countries. This city managed to curb the corruption that was prevalent in the 1960s.^[11]

Concepts and programs like National Integrity Systems (Transparency International), The Ethics Infrastructure (OECD), and A Framework for Integrity (World Bank) offer numerous methods and institutions to curb corruption and promote integrity.^[1,12–15] The main elements are summarized.

The overall *goals* should include development strategies that yield benefits to the nation as a whole, including its poorest and most vulnerable members, and not just to well-placed elites. Additionally, it is important that public services be efficient and effective and that government function under law, with citizens protected from arbitrariness (including abuses of human rights).

Many institutions and activities are important in the struggle against corruption. Leadership counts—the research is unanimous on this. Embodying the political will to fight corruption, leaders set the tone through their policies and the examples they present. In the political system, parliament as representative of the people and watchdog of political power is at the center of the struggle to attain and sustain good governance and to fight corruption. An independent, impartial, and informed judiciary holds a central place in the realization of just, honest, open, and accountable government. The public service is important for its services as well as for the protection of the public decision-making process (including the system of public procurement, often vulnerable because of its opportunities for corruption). Integrity is a crucial element of professionalism. People should be given a feeling of pride in their work and that deviance undermines the core of the profession.

A powerful anticorruption device also is the simple establishment of sound financial management practices, including a timely and efficient accounting system combined with punctual, professional review by internal and independent auditors. Important institutions are also an effective Auditor-General as a watchdog over financial integrity and the credibility of reported information, an ombudsman who can recommend improvements to procedures and practices and act as an incentive for public officials to keep their files in order, and Independent Anticorruption Agencies to raise awareness among the public, to stimulate prevention, and to detect and investigate corruption cases. The availability of resources, their independence from management, and the availability of mechanisms of transparency and accountability are important factors for the success of these institutions. Of importance

is a judicious balance between positive and negative social control. An extreme accent on control can be extremely counterproductive. An anticorruption campaign in New York City fostered its own pathology of excessive rules and procedures, paralysis of decision making, and the undermining of quality.^[16] To safeguard integrity presupposes that attention is being paid to group and organizational culture (norms, values, and perceptions), including training and education in ethical dilemmas and the development of codes of conduct.

Additionally, it is important that the media and civil society play a role. Information and public awareness (the “right to know”), are linked inextricably to accountability, the central goal of any democratic system of government. The principal vehicle for taking information to the public is an independent and free media. The role for civil society must be to claim and defend its own values and not leave this integral function to those in power.

INTERNATIONAL AND SUPRANATIONAL INITIATIVES

Nowadays, a lot of attention is being paid to corruption and ethics in international fora. Anticorruption congresses and conferences with thousands of participants are no exception. Conventions and treaties have been prepared, signed, and implemented, in international relations, conditions of good governance have become an important topic. Governments are more aware of the importance of their national integrity system. Many global and international initiatives are worth mentioning.

The United Nations' General Assembly Resolution on Corruption in 1997 reaffirmed its concern at the seriousness of the corruption problem and adopted the International Code of Conduct for Public Officials. The resolution recommended that member states use the code as a tool to guide their efforts against corruption.^[10] In 2001, the foundations for a United Nations Convention Against Corruption were discussed in The Hague at the Second Global Forum on Fighting Corruption and Safeguarding Integrity. Within the United Nations, the Development Program approaches the issue of corruption as a governance problem, and it has made the minimization of corruption central to achieving the organization's overall purpose of alleviating poverty and attaining social and people-centered sustainable development.^[10]

In line with political developments, major international financial and economic institutions like the World Bank and the International Monetary Fund have made corruption their topic. The World

Bank changed its policies in 1996, and the cancer of corruption is seen as a crucial problem for economic development. Its “Pillars of National Integrity” are interdependent and include the rule of law, sustainable development, and quality of life. Corruption and fraud have become important topics within bank-financed projects as well as in the country assistance strategies and country lending considerations.^[10,15]

The Organization for Economic Co-Operation and Development, with 29 member states, representing the industrialized world, has made anticorruption initiatives and ethics and integrity policies an important aspect of its work.^[13,14] The concepts of an “Ethics Infrastructure” summarizes OECD’s view with core elements of political commitment, workable codes of conduct, professional socialization mechanisms, an ethics coordinating body, supportive public service conditions (decent working conditions), an effective legal framework, efficient accountability mechanisms, and an active civil society. The OECD supports its pleas with research and analysis on the ethics measures of its member countries. In the international arena, the OECD’s main contribution has been the 1997 OECD Convention on Combating Bribery of Foreign Public Officials in International Business Transactions. The Convention was signed by all member countries and came into force in 1999. Its main achievement is that it made bribing a foreign public official a criminal offense in many countries (while in the past, this bribe was even tax deductible in several of the OECD countries).

Transparency International, based in Berlin, is a non-governmental international organization against corruption with branches in roughly 70 countries. It favors a holistic approach to a national integrity system.^[1,12] Although each country or region is unique in its own history and culture, its political system, and its stage of economic and social development, similarities exist, and experience and lessons are often transferable. Transparency International proposes a National Integrity System as a comprehensive method of fighting corruption. It comprises eight pillars (public awareness, public anticorruption strategies, public participation, watchdog agencies, the judiciary, the media, the private sector, and international cooperation) which are interdependent. Every 2 years, Transparency International organizes the International Anti-Corruption Conference (IACC).

The mentioned initiatives are part of a more general trend toward anticorruption policies, as, for example, is also shown by the Organization of American States (Inter-American Convention Against Corruption), the European Union, the Council of Europe (and its Group of States Against Corruption *GRECO*), the Global Coalition for Africa, and the International Chamber of Commerce.^[10,14,15]

Criticism

Corruption, ethics, and integrity have become more important topics in international and national political, economic, and social forums, but this does not mean there exists an overall consensus on concepts, causes, and policies. Criticism can still be heard on a number of subjects.

First, there is the question of which interests are best served by the anticorruption initiatives.^[2,10] Optimists see it as a win-win situation: the interests of multinational corporations are served as well as the interests of the peoples of developing countries. Others argue that initiatives of institutions like the World Bank and the International Monetary Fund contribute to the development in the direction of a global free market capitalism and that globalization of that type harms the poor and powerless.

Second, there is the contradiction between the general policy frameworks of, for example, the World Bank, and the notion that there is no “one right way.” Countries are confronted with demands coming from the frameworks, and the room to adapt to national circumstances is limited.

This is related to a third comment. How much consensus is possible and desirable on the ethics of politics and administration? Many frameworks suggest that ethics is universal and based on global consensus. Others argue, however, that the frameworks reflect the values and norms of only a small part of the world.^[3,15] More knowledge of the values and norms reflected in non-Western political and social thought and practice would be useful to come to a more precise picture.

CONCLUSION

A last remark concerns our knowledge of the extent of corruption in countries and the effectiveness of anti-corruption methods and strategies. Most proposals and frameworks are based on experiences and knowledge of anticorruption practitioners and not on scientific evaluation. Although scientific knowledge has increased (as journals such as *Public Integrity* and *Crime, Law and Social Change* show), more comparative research will be necessary to come to more reliable conclusions on the effectiveness of strategies and methods. The same applies to our knowledge of the extent and mechanisms of bribery and corruption in countries.

INTERNET RESOURCES

- a. Council of Europe (<http://www.coe.int>) and its program against corruption and organized crime: <http://www.legal.coe.int/economiccrime>;

- and the Group of States Against Corruption: <http://greco.coe.int>.
- b. European Commission (<http://europa.eu.int>) and its codes of conduct: http://europa.eu.int/comm/codesofconduct/index_en.htm.
 - c. Independent Commission Against Corruption of Hong Kong: <http://www.icac.org.hk>; and New South Wales Australia: <http://www.icac.nsw.gov.au>.
 - d. International Monetary Fund: <http://www.imf.org>.
 - e. Organization of American States' (<http://www.oas.org>) anticorruption efforts, including model laws and Inter-American Convention Against Corruption.
 - f. Organization for Economic Co-operation and Development's (<http://www.oecd.org>) anti-corruption program for business and the public sector, and its general library.
 - g. Transparency International (<http://www.transparency.org>), *The TI Source Book*, and, with Göttingen University, the Internet Center for Corruption Research (http://www.gwdg.de/~uww/i-cr_head.htm), including Transparency International's *Corruption Perception Index*.
 - h. The UK Committee on Standards in Public Life: <http://publicstandards.gov.uk>.
 - i. The United Nations (<http://www.un.org>) and <http://magnet.undp.org> for the Programme for Accountability and Transparency of the United Nations Development Programme.
 - j. The U.S. Office of Government Ethics (<http://www.usoge.gov>) on international developments in ethics (programs).
 - k. The World Bank's (<http://www1.worldbank.org>) public-sector program (<http://www1.worldbank.org/publicsector/anticorrupt>) and links.

ACKNOWLEDGMENT

I am very grateful for the assistance and suggestions offered by Professor Carol Lewis.

REFERENCES

1. Doig, A.; McIvor, S. *Transparency International, The National Integrity System. Concept and Practice*. A Report by Transparency International (TI) for the Global Forum II on Fighting Corruption and Safeguarding Integrity; Transparency International: Berlin, 2001.
2. Heidenheimer, A.J., Johnston, M., Levine, V.T., Eds.; *Political Corruption. A Handbook*; Transaction Publishers: New Brunswick, 1989; 1016 pp.
3. Punch, M., Kolthoff, E., van der Vijver, K., van Vliet, B., Eds.; *Coping with Corruption in a Borderless World. Proceedings of the Fifth International Anti-Corruption Conference*; Kluwer Law and Taxation Publishers: Deventer, 1993; 172 pp.
4. Cooper, T.L., Ed.; *Handbook of Administrative Ethics*; Marcel Dekker: New York, 2001.
5. Sampford, C., Preston, N., Bois, C.-A., Eds.; *Public Sector Ethics. Finding and Implementing Values*; The Federation Press: Annandale and Routledge, London, 1998.
6. Lawton, A. *Ethical Management for the Public Services*; Open University Press: Buckingham, 1998.
7. Jacobs, J. *Systems of Survival. A Dialogue on the Moral Foundations of Commerce and Politics*; Hodder & Stoughton: London, 1992.
8. Heywood, P., Ed.; *Political Corruption*; Blackwell: Oxford, 1997; 250 pp.
9. Klitgaard, R. *Controlling Corruption*; University of California Press: Berkeley, 1988.
10. UNDP United Nations Development Programme. *Corruption and Integrity Improvement Initiatives in Developing Countries*; UNDP: New York, 1998; 174 pp.
11. Huberts, L.W.J.C. Anticorruption strategies: the Hong Kong model in international context. *Public Integrity* **2000**, 2 (3), 211–228.
12. Pope, J. *Confronting Corruption: The Elements of a National Integrity System (TI Source book 2000)*; Transparency International: Berlin, 2000; 364 pp. (also at <http://www.transparency.org>).
13. OECD Organization of Economic Co-operation and Development. In *Trust in Government. Ethics Measures in OECD Countries*; OECD: Paris, 2000; 329 pp.
14. OECD Organization of Economic Co-operation and Development. *No Longer Business as Usual. Fighting Bribery and Corruption*; OECD: Paris, 2000; 276 pp.
15. Pieth, M., Eigen, P., Eds.; *Korruption im internationalen Geschäftsverkehr. Bestandsaufnahme. Bekämpfung. Prävention*; Luchterhand: Neuwied, Kriftel, 1999; 735 pp.
16. Anechiarico, F.; Jacobs, J.B. *The Pursuit of Absolute Integrity. How Corruption Control Makes Government Ineffective*; University of Chicago Press: Chicago, 1996; 274 pp.

Governance Networks

Robert Agranoff

School of Public and Environmental Affairs, Indiana University, Bloomington, Indiana, U.S.A.

Michael McGuire

Department of Public Administration, University of North Texas, Denton, Texas, U.S.A.

INTRODUCTION

Governance networks refer to formal and informal structures comprised of representatives from governmental and non-governmental agencies working interdependently to jointly formulate and implement policies and programs, usually through their respective organizations. These networks are a function of “differentiated polities,” characterized by institutional and functional specialization. They bring the non-profit and for-profit sectors together with government in a number of policy arenas, including economic development, health care, criminal justice, human services, information systems, rural development, environmental protection, biotechnology, transportation, and education. The activities of governance are purposeful efforts to guide, steer, control, or manage, in which public and private actors do not act separately but in conjunction, operating as a network. Governance refers to the patterns that emerge from the governing activities of the actors, e.g., coregulation, costeering, coproduction, cooperative management and public-private partnerships.^[1,2]

EXAMPLES OF A GOVERNANCE NETWORK

Three different types of governance networks demonstrate the range of collaborative activities undertaken. First, Sematech, an acronym for semiconductor manufacturing technology, is a public-private development partnership involving the Defense Advanced Research Projects Agency of the U.S. Department of Defense and 14 information technology manufacturers, who jointly invest equal amounts of money in basic research and development designed to restore U.S. leadership in the computer chip industry. Sematech has focused its efforts on a prototype chip fabrication plant that leads to the high-tech equipment that serve as a testing venue for new American precompetitive technologies. The research and development effort improves the competitive stance of all partners. This approach has allowed all of the potentially competing partners to remain involved and invested, and it indirectly serves

the Defense Department’s needs for research and development.

Second, a less formal set of governance networks operate in Michigan in order to create the “Gold Collar” jobs needed to emphasize strategic priorities in advanced manufacturing, information technology, and life sciences. Implemented by its state agency, the Michigan Economic Development Corporation, a number of strategic alliances have been created with the automotive industry, information technology companies, pharmaceutical research centers, medical centers, university research centers, engineering schools, and others. In addition, the state agency staff promotes eight industry cluster roundtables (e.g., automotive, food products, forest products) designed to identify common problems and opportunities, recommend policy and regulatory changes, share and develop technology, promote technical assistance, and forge other new working relationships.

Third, the Applegate Partnership of southern Oregon and northern California is a locally organized watershed governance network that promotes ecologically and financially responsible resource management. The partnership involves industry, conservation organizations, public natural resource agencies, and residents working together on land management efforts through community involvement and education. Applegate’s “signature” project involved the development of a GIS system that integrated BLM, U.S. Forest Service, and county tax lot information.

BACKGROUND AND CONCEPTS

There are many forces that have led to the emergence of governance networks. One important factor is the changing nature of work from labor-based production and services to the integration of knowledge-based symbolic-analytic work, which places greater value on human capital. The basic challenge is that knowledge is specialized and must be integrated collaboratively to solve many problems, a core issue in change management. As a result, government agencies, once thought to be the monopolistic holders of key

information and expertise relating to public issues, now only possess some of the information needed to solve problems.

A second factor is the changing nature of government. The twentieth century was a time of growth of welfare states and, consequently, government agencies and programs at national and subnational levels. The government took on more and more problems and created many new policy arenas. As public efforts grew, however, it became apparent that the government could not garner the resources, investments, expertise, or commitments needed to solve all public problems. Networks involving several organizations became one of a number of collaborative efforts to try to approach some of society's "wicked problems," or challenges that could not be handled by dividing them in simple pieces, in isolation from one another.^[3] A related factor is the idea that government should not only collaborate or partner but also should take on more of a developmental or steering role, promoting, regulating, and encouraging various types of non-governmental activity and operations. This philosophy has, in many fields, unfolded in the 1980s and 1990s, and it has led to greater variety in government–non-governmental organizational interaction.

A number of academic research streams have also confirmed the importance of governance networks. The urban politics work of Clarence Stone on regime theory is perhaps most seminal. In his study of urban power, he concludes that in a fragmented, world where power, resources, knowledge, and the other means to solve problems are dispersed among many individuals and organizations, "the issue is how to bring about enough cooperation among disparate community elements to get things done—and to do so in the absence of an overarching command structure or a unifying system of thought." Governance is the power to combine the necessary elements toward a result, that is, the capacity to assemble and use needed resources for a policy initiative.^[4]

Intergovernmental researchers have also recognized the importance of public and non-public networks, often operating in complex and overlapping fashions,^[5,6] and in many ways, changing the traditional role of governments and their links with non-governmental organizations^[7] and with the various tools and strategies that lead to different public–private configurations.^[8] A study by Radin and associates demonstrates how federal–state–private councils in rural development have recommended many program changes and demonstration approaches in governance.^[9] In the same vein, research on economic development at the state and local levels has demonstrated how these governments enhance their economies by stimulating private sector action, engaging in partnerships with such organizations as chambers of commerce and industry groups, and

working together on developmental policies in human resource development, technology advancement, and global marketing.^[10–12] Finally, research in the environmental policy area also demonstrates that new solutions to such problems as non-point source pollution (e.g., agricultural chemicals), estuary restoration, and forest management can be approached by formally and informally convening government agencies, conservation advocacy groups, industry representatives, land developers, and the scientific community into joint decision bodies.^[13,14]

HOW GOVERNANCE NETWORKS OPERATE

How do governance networks function? [Table 1](#) provides a live example of a formal effort to enhance rural development policy in Iowa. It was called a Rural Policy Academy, because it was part of a program of that name sponsored by the National Governor's Association. A small team guided the effort, led by two state officials, but it also involved several statewide non-governmental organizations involved in agriculture, health, and economic development. Early in the process, the Academy added four state legislators as a means of enhancing its political support. The network brought in dozens of other stakeholders through its work groups in each of the five areas it chose to emphasize strategically. The Academy built upon some 14 existing state rural development programs, using a strategic planning approach that was designed from the start to influence policy and be implementable through state government and several non-governmental stakeholders.

Through a series of retreat-workshops, the Iowa Academy identified the five core areas of emphasis: business development, agriculture and agribusiness, health care, community leadership, and government services and infrastructure. After the key issues were identified, subsequent retreats by the core team reviewed programs, set priorities, and drafted a strategic document. From there, the network was expanded in each of the five strategic areas to include the technical work groups comprised of a broadened group of stakeholders. Those groups ultimately developed a set of policy statements. These draft policy statements were discussed by sets of 25- to 30-person focus groups, who represented even more rural development stakeholders. Finally, the refined proposals were introduced into the state budget and into legislative proposals.

The results of the Academy effort were mixed, because the State of Iowa was facing a budget shortfall during the year of completion. Nevertheless, over a dozen program expansions and new programs were

Table 1 Iowa's rural policy academy: a governance network in action

Network actors	Antecedent programs	Academy aims	Key strategic areas	Sequence of activities	Highlights of results
1. Academy team.	• Rural Enterprise Fund	• Bottom-up rural development	• Business development: agricultural and electrical, non-electrical machinery; 14 industry sectors identified	• Formation of team	• Increases in Agriculture Enterprise Fund to foster linkages with commodity groups, support livestock production policy network
• Iowa Department of Economic Development (DED) Rural Coordinator	• Infrastructure Planning	• Foster interagency and interorganizational cooperation with existing programs—partnerships are key	• Agriculture: new production and marketing technology while addressing food safety and environmental protection; value-adding businesses	• Identification of key strategic issues	• Extension Service project on livestock facility retrofitting
• Director Iowa Department of Management and Budget Representative, Regional Councils	• Tourism Welcome Centers	• Focus on existing problems and future demands	• Health care: maintaining a viable rural health care system through partnerships between the state, private sector, and local leaders	• Workshop retreat: review public and private programs; analyze existing needs and program gaps; set priorities; explore alternative strategies; examine potential coordination efforts	• Increased technology transfer funding
• Director, Iowa Agricultural Development Authority	• Rural Main Street	• State government's role is increasingly that of a facilitator and catalyst for action, rather than that of a direct and sole provider of a service	• Community leadership: assisting rural communities to develop local leaders and organizations designed to improve their growth and services	• Subsequent team meetings to sift issue areas and draft strategic documents	• Enhanced Rural Enterprise Fund
• Iowa Hospital Association Director	• Rural business incubators	• Comprehensive strategy with stakeholder involvement and support	• Government services/ infrastructure: maintain services that are sufficient to promote growth; local governments taking innovative actions to better serve residents and control their own destinies	• Five issue areas are identified and stakeholder workgroups are identified and developed for each area	• Rural leadership program shift to county community leadership clusters and linkages
• Area Development Group	• Rural leadership development	• Creation of an implementation plan		-Second retreat to refine the five issue areas	• Funding for major rural infrastructure assessment

• Iowa Cooperatives Bankers' Association	• Agricultural diversification	• Legislative submission, incorporation in the governors budget, enactment	• Work groups meet to conduct environmental scan and to draft policy statements	• Task Force on "Government of the Future" with the focus on collaborative services and regional institutions authorized
• Federal-State Cooperative Extension Service	• Marketing efforts	• Program development by DED, Department of Public Health, Extension Service, regional centers, and many non-governmental organizations	• Two days of focus group sessions held on each issue area, involving 25 to 30 persons per group	• Funding of plastics technology and graphic arts centers to promote small business networks
• Four legislators, two from each house, each party • Consultant/facilitator (former legislator)	• Office of Rural Health • Regional trade and business centers (16)		• Drafting of final strategy • Introduction of strategic elements into state budget and legislative proposals	• Internet marketing consortium launched • Training program for rural bankers on non-agriculture lending
2. Work group representatives	• Technology transfer foundation			• Entrepreneurship development materials for Iowa schools
• State agency program heads	• Lottery funds for environmental projects			• Establishment of pilot community health system in one rural county
• Federal agency (USDA, SBA, EDA) staff located in Iowa • General farm organizations (FarmBureau, Prairie Fire) • State agricultural commodity organizations • Iowa State University, School of Planning and Design, School of Agriculture, Extension Service	• Quality and productivity enhancement coalition • Road funds to support economic development • Fiber optic telecommunication system			• Public-private task force on retention of health professionals

Source: From Iowa Rural Policy Academy.

ultimately adopted, including enhancing various forms of venture capital, creating several new production networks and consortia, and redirecting other programs. Most important, many of these programs were implemented outside of state government, for example, the pilot rural health system, the Internet marketing consortium, the bankers training, and the rural leadership program. The Iowa Rural Policy Academy demonstrates the concrete actions involved in solving public problems by governance networks.

GOVERNANCE NETWORK MANAGEMENT

Despite the prevalence of governance networks, less is known about how they are managed, because so much of the focus in public administration has been on the hierarchical governmental organization. A few important approaches can nevertheless be advanced. The overriding concern is that networks are not hierarchical and, thus, are not bound by legal authority or by the normal rules and procedures that bind single organizations. Moreover, the importance of knowledge or possession of some other resource, like financing or implementation ability, makes the administrative playing field different than that of a public bureaucracy.

It appears that network actors follow a sequence of activities that is different from the usual planning, organizing, staffing, directing, and so on. The process usually begins with *activation*, or identifying participants and stakeholders and tapping the disparate skills, knowledge, and resources. Then, *framing* follows, which is the process establishing and influencing the operating rules of the network, influencing the governance structures prevailing values and norms, and altering the perceptions of the participants. Next comes the *mobilizing* actions, involving getting the various governance stakeholders to sense a strategic purpose and to establish a common set of objectives. Finally, there are the *synthesizing* activities, that is, the acts of creating the environment and enhancing the conditions for favorable decisions related to the strategic purpose of the network.

Working in a network is behaviorally different than working in an organization. Instead of “motivation,” various forms of shared learning processes are key, which come from interagency task group development and lead to jointly arrived at solutions. Under normal circumstances, decisions are deliberative, creative, and group derived, as organization representatives shed some of their ideologies and personal “baggage” in order to forge redefined problems, a mutual course of action, a multiorganization implementation technology, and diversified methods of resource acquisition. Some argue that these processes are held together by a generally held belief in the overall

purpose of the network, whereas others point to the importance of social capital, which is the stock or built-up reservoir of good will that flows from different interests and organizations working together for mutual productive gain. It is also widely recognized that trust, the obligation to be concerned with others’ interests, is an important ingredient in network behavior.

The presence of collective decision making and shared learning does not mean that issues of power are absent in governance networks. Coequality is a principle that is often professed in networks but is hard to maintain. Beyond the facade of trust and collaboration, stronger partners may be able to take advantage of weaker partners. Some actors sit in positions where their knowledge, financial resources, organization position, or legal authority alters the balance of power. On the other hand, a governance network working together and making reasonable proposals can accrue a great deal of power to create new strategies and policies that carry the support of a wide range of organizations.^[15]

“METAGOVERNMENT”

Operating in governance networks is obviously changing the nature of government organizations, at least in regard to that portion of its activities that relate to policies and programs that are shared with other organizations. Foremost, public agencies, or their representatives, become partners with other organization representatives in examining problems, establishing strategies, and formulating policy responses. Often, the public organization actor serves as the convener, but once the process begins, this person becomes one among many participants. Second, the public agency representative does not have nearly the monopoly or the corner on technical expertise that previous public administrators possessed. Many stakeholders, e.g., scientists, organizational researchers, interest groups, and advocacy groups, bring needed knowledge and information to the governance table. Third, resources are more dispersed. In the past, the government agency possessed the major allocation or appropriation needed to launch a program, and money (and indirectly control) was dispensed through a chain of agencies, public and non-public. Today, the resources are more dispersed throughout the network, as government increasingly tries to use its role in governance to leverage investments by a host of non-governmental organizations.

Fourth, program implementation occurs through many of the same organizations that were involved in formulating strategies and policies. As government has taken on more of a guidance role, and has encouraged non-governmental investments, the carrying out

function is no longer exclusively through an intergovernmental chain of public organizations but involves a variety of grantees, contractees, and collaborating partners. Finally, some analysts have argued that this renders the government agency unimportant, powerless, and a bystander to a series of private actions. Governance networks change the role of government in democratic systems. Government agencies are not, however, marginal players in the multiple organization process. They remain important actors, because they continue to possess a legitimacy to approach public problems and policy solutions, retain important legal authority to set rules and norms, continue to contribute financial resources to programs, and retain some of the needed subject matter and scientific knowledge needed to approach problems. Government agencies are almost always key partners in governance networks.

It is important to note that the notion of democratic accountability may need to be reconceptualized in the light of multiorganizational decision making that lies largely outside of the formal organs of representative government.^[16] There is no obvious principal or agent in network governance and no exigent authority to steer the activities of the network in harmony with elected officials. Interaction between civil servants and representatives of private interest groups, other governmental layers, and implementing organizations makes it hard for representative bodies to influence policy. Network governance may force a shift from concerns with hierarchical accountability to notions of responsibility, responsiveness, and the fostering of democratic ideals. Because an hierarchical control orientation is problematic for collaborative structures, some steps can be taken to maintain the flexibility of collaboration while ensuring accountability.^[17] For example, collaborative processes can be developed as supplemental to, not exclusive of, normal decision-making processes. The activities of networked arrangements could be reviewed by independent, ostensibly objective, sources. Additionally, collaborative decision making can strive toward achieving agreed-upon performance measures that capture the intent of policy objectives, and the home agencies of collaborative decision makers can monitor and evaluate the network's progress toward achieving the desired objectives.

FUTURE OF GOVERNANCE NETWORKS

The future will put these collaborative decision structures at the core of organized efforts. Knowledge is an increasingly important resource, a new capital that resides in human resources or knowledge workers, who will continue to need some form of collective that will

bring them together. Knowledge is non-hierarchical, in that if it is required for a situation, professional performance needs to be applied, regardless of organizational position, social status, or possession of wealth. As a result, portable knowledge application plus rapid access to information can and has led to the disintegration of large-scale organizations into more flexible structures, such as the identified partnerships, joint ventures, alliances, and other governance structures. In the future, several different types of organizations are expected to interlock along these lines. These trends will also accelerate the need for greater study of the new forms of organizing and operating. As Peter Drucker maintains, "Despite all the present talk of 'knowledge management,' no one yet really knows how to do it."^[18] Finally, these structures will increasingly be used to deal with social problems. David Korten concludes that knowledge and reformation have provided powerful new collective intelligence that can be used to master social and institutional discovery and innovation through problem solving.^[19]

REFERENCES

1. Kooiman, J. Social-political governance. In *Modern Governance: New Government-Society Interactions*; Kooiman, J., Ed.; Sage: London, 1993; 1-2.
2. Rhodes, R.A.W. *Understanding Governance: Policy Networks, Governance, Reflexivity and Accountability*; Open University Press: Buckingham, 1997.
3. O'Toole, L.J. Treating networks seriously: Practical and research-based agendas in public administration. *Public Adm. Rev.* **1997**, 57 (1), 46.
4. Stone, C. *Regime Politics*; University Press of Kansas: Lawrence, KS, 1989; 227 pp.
5. Agranoff, R.; McGuire, M. Multi-network management: Collaboration and the hollow state in local economic policy. *J. Public Adm. Res. Theory* **1998**, 8, 67-91.
6. Meier, K.J.; O'Toole, L.J. Managerial strategies and behavior in networks: A model with evidence from U.S. public education. *J. Public Adm. Res. Theory* **2001**, 11 (3), 271-294.
7. Mandell, M. Community collaborations: Working through network structures. *Policy Stud. Rev.* **1999**, 16 (1), 42-64.
8. McGuire, M. Collaborative policy making and administration: The operational demands of local economic development. *Econ. Dev. Q.* **2000**, 14, 276-291.
9. Radin, B.A.; Agranoff, R.; Bowman, A.O'M.; Buntz, G.G.; Ott, S.J.; Romzek, B.S.; Wilson, R.H. *New Governance for Rural America: Creating Intergovernmental Partnerships*; University Press of Kansas: Lawrence, KS, 1996.
10. Clarke, S.E.; Gaile, G.L. *The Work of Cities*; University of Minnesota Press: Minneapolis, 1998.
11. Eisinger, P. *The Rise of the Entrepreneurial State*; University of Wisconsin Press: Madison, 1988.

12. Fosler, R.S. State economic policy: The emerging paradigm. *Econ. Dev. Q.* **1992**, 6 (2), 3–13.
13. John, D. *Civic Environmentalism*; Congressional Quarterly Press: Washington, DC, 1993.
14. Wondolleck, J.M.; Yaffee, S.L. *Making Collaboration Work: Lessons from Innovation in Natural Resource Management*; Island Press: Washington, DC, 2000.
15. Agranoff, R.; McGuire, M. Big questions in public network management research. *J. Public Adm. Res. Theory* **2001**, 11 (3), 295–326.
16. Hirst, P. Democracy and governance. In *Debating Governance*; Pierre, J., Ed.; Oxford University Press: Oxford, 2000; 13–35.
17. Wondolleck, J.M.; Yaffee, S.L. *Making Collaboration Work: Lessons from Innovation in Natural Resource Management*; Island Press: Washington, DC, 2000.
18. Drucker, P.F. The next society. *Economist*. November 3, **2001**, 28, 16
19. Korten, D.C. *Getting to the 21st Century: Voluntary Action and the Global Agenda*; Kumarian Press: West Hartford, CT, 1999; 214–221.

Government Failure: Theory

Thomas S. Nesselin

Urban and Regional Studies, Department of Economics, University of Wisconsin-Green Bay,
Green Bay, Wisconsin, U.S.A.

INTRODUCTION

In *The Wealth of Nations* (1776), Adam Smith presented the basic case for the comparative efficiency of a capitalist market economy. According to Smith, within a competitive economy, an *invisible hand* turns self-interest toward the common good. The modern counterpart of Adam Smith's invisible hand conjecture is the so-called *Fundamental Theorem of Normative Economics*. Using highly sophisticated mathematics, economists such as Kenneth Arrow and Gerald Debreau were able to prove that under certain conditions—albeit highly restrictive—a perfectly competitive economy will result in an efficient allocation of resources.^a

In the twentieth century, however, economists began to investigate seriously under what specific conditions a market economy might fail to achieve efficient resource allocation. Modern textbooks present a fairly standard list of market failures. Markets are argued to fail in the following key ways: 1) monopoly; 2) collective goods (e.g., national defense, the legal system, etc.); and 3) externalities or external effects (e.g., pollution). More recently, the theory of market failure has been extended to include the ubiquitous problem of costly information and incomplete markets.^[2] This broader more refined analysis of market failure suggests a potential role for government to improve the efficiency of, for example, major insurance markets and capital markets. To date, most undergraduate textbooks, however, have yet to incorporate this broader analysis of market failure.

Theories of market failure are without doubt an important analytic framework for helping policy analysts discern potential government policies that may improve the efficiency of the economy. However, the application of market failure theory is an incomplete, and sometimes invalid, guide to efficient public policy making. In brief, for much of the twentieth century, there was a strong tendency by policy analysts to

simply assume away the possibility of government failure. In essence, policy interventions were implicitly assumed to be perfectly implemented by a benevolent, all-knowing, and all-powerful entity called “Government.” This point was stressed initially by members of the Chicago School of Political Economy, such as Coase,^[3] George Stigler, Milton Friedman, and Harold Demetz. Demetz presented a concise but particularly cogent argument in this respect:^[4]

The view that now pervades much public policy economics implicitly presents the relevant choice as between an ideal norm and an existing “imperfect” institutional arrangement. This *nirvana* approach differs considerably from a *comparative institution* approach in which the relevant choice is between alternative real institutional arrangements. In practice, those who adopt the nirvana viewpoint seek to discover discrepancies between the ideal and the real and if discrepancies are found, they deduce that the real is inefficient. Users of the comparative institution approach attempt to assess which alternative real institutional arrangement seems best able to cope with the economic problem; practitioners of this approach may use an ideal norm to provide standards from divergences are assessed for all practical alternatives of interest and select as efficient that alternative which seems most likely to minimize the divergence.

INEFFICIENCIES OF POLITICAL INSTITUTIONS: A FEW ILLUSTRATIONS

Criticism of the market failure paradigm as the sole guide to policy interventions stimulated many scholars, in particular, members of the so-called *public choice* school, to examine the question of government failure. Rather than assuming that political actors generally pursue the public interest or common good, public choice analysts posit that political actors have their own agendas and generally seek their self-interests. They then employ the analytic tools of economics to explain the outcomes of the political process and governmental institutions. The public choice literature has many books and articles on the causes and consequences of a large range of government failures.^[5–11] Often, analysts, in particular textbook authors, simply

^aThe Fundamental Theorem of Normative Economics establishes that a competitive economy will result in a Pareto-efficient allocation of resources. That is, the competitive price mechanism will lead to adjustments such that it is not possible to reallocate resources in any way that will increase the satisfaction of one party without decreasing the satisfaction of some other party.^[1]

present an unorganized listing of what they believe to be important reasons for government failure. A modest improvement in this respect, perhaps, is to relate key government failures to deficiencies in one of the three fundamental structures of all social choice mechanisms, whether market or non-market, namely, the *decision-making* structure, the *information* structure, and the *incentive* structure.

To begin, in a competitive market economy, consumers are able to channel their preferences to producers and make them effective, via their dollar votes in the marketplace. Each consumer, within the limits of their budget, can purchase exactly the quantity and quality of goods and services they desire. Absent market failures, the decentralized decision-making structure of a market economy results in efficiency, i.e., resources are allocated to those lines of production that maximize society's net benefit.

In contrast, the use of a more centralized social decision-making mechanism, namely, collective or political choice mechanisms, often presents major problems in efficiently channeling the preferences of voters and making these choices effective. One method for making social choices is via direct democratic referendum, where the preferences of 50 plus 1 of the voters decide the issue. Clearly, a key problem of a majority decision-rule is that the preferences of a lukewarm and small majority may dominate the preferences of an intense minority. For example, perhaps a majority of voters in a school expenditure referendum are willing to pay a small increase in taxes per year, whereas a large minority may actually desire a significant decrease in taxes. Such a situation may easily lead to an inefficient level of school spending and taxation because of the inability of a majority voting rule to account for the intensity of voter preferences.

Of course, in modern democracies, the bulk of collective decisions are not made by referenda but by elected representatives. Perhaps the central problem of representative government is that the outcomes of the political process often fail to reflect society's general interest, and instead, they reflect the interests of organized groups (i.e., so-called *special interests*). Often, these organized interest groups seek legislation that makes wealth transfers to them at the expense of society in general, the classic example being tariff and other trade barriers protecting favored industries from more efficient foreign competitors.

The basic reasons for inefficient special interest legislation are straightforward. First, for most voters, the cost in time, effort, and money of being well-informed on a large number of public policy issues are large relative to the benefits to them. Consequently, it is rational for most voters to remain uninformed on most policy questions. Obviously, voter ignorance greatly impedes the monitoring of their political

representatives. Second, self-interested political representatives generally place a high priority on getting re-elected in order to continue to enjoy the perquisites of political office. Finally, the influence of special interest groups is greatly increased, because organized interest groups have the means and incentive to be the primary providers of political funding. It is not difficult to see that in political systems where the political funding is paramount to elector success and where the perquisites of political office are exceedingly large, special interest legislation will be especially pervasive.

A second major cause of government failure relates to information deficiencies on the part of voters. Two basic sources of information deficiencies are the following. Again, a fundamental problem is that it is rational for the bulk of voters to be uninformed on most political issues. Greatly compounding this inherent problem is the fact that self-interested politicians have strong incentive to structure policies and programs so as to systematically exaggerate potential benefits while concealing the true social costs. For example, in the United States, the Social Security pension system is financed by a payroll tax on workers and a mandated matching employer contribution. Yet, economists generally agree that workers actually bear the full burden of financing social security, as the employer contribution results in a corresponding reduction in wages in the long run.

A final major source of government failure relates to the large range of incentive problems. Much of government activity is accomplished via government agencies and enterprises. Analysts have long noted that public-sector managers confront a set of incentives that generally undermines efficient performance. In particular, government agencies and firms fail to be disciplined by the threat of bankruptcy or outside takeover, while at the same time soft budget constraints provide incentive for increased expenditure and a lack of cost consciousness. Compounding basic incentive problems is another, often overlooked, cause of public-sector inefficiency noted by Shleifer:^[12] "In other words, state firms are inefficient not just because their managers have weak incentives to reduce costs, but because inefficiency is the result of the government's deliberate policy to transfer resources to supporters."

TOWARD A THEORY OF GOVERNMENT FAILURE

While numerous, rather narrow, articles have been produced examining various aspects of government failure, "few writers have tried to synthesize this literature to construct a theory of government failure that parallels that of the market."^[13] The major exceptions are studies by Wolf, Le Grand, and Vining and

Weimer.^[5–11,13,14] While each of these studies provide valuable insights toward the development of a theory of government failure, to date, no complete and fully adequate theory of government failure exists. Dollery and Worthington provided an excellent concise critique of the government failure literature.^[15] Three key general criticisms made by them are as follows:^[15]

[I]n at least two respects extant public choice theory does not provide policymakers with a conceptual analogue of the market failure paradigm to inform public policy decisions. Firstly, the public choice approach is positive, and not a normative, theory of government processes, and thus cannot provide an idealised or optimal vision of policy intervention against which actual government behaviour can be compared, in the same way as welfare economics furnishes the theory of market failure with a yardstick to evaluate real-world markets. And secondly, public choice theory is premised on the behavioural posulate of *homo economicus* and follows methodological individualism, which may not always capture the full intricacies of politicised policy environments ... [Finally,] effective policy analysis requires a general theory of organisational failure, rather than conceptually distinct theories of market failure and nonmarket failure. This has been recognised by Williamson (1989) who has argued that "... all forms of organization, not just markets, need to be assessed comparatively"

CONCLUSION

What is to be done, then, to aid the policy analyst in making policy choices in a world of imperfect markets and imperfect government? Le Grand's approach appears to be one potentially useful path to follow.^[13] In brief, Le Grand attempted to provide a more general policy analysis by focusing on the three basic ways that governments intervene in a market economy and then assessed these three interventions in terms of efficiency and equity criteria. Le Grand provided a concise and basic analytical discussion of direct government *provision, taxation and subsidy, and regulation*. To be of much use for policy making, however, a more extensive theoretical development is needed. For example, a more precise and refined theoretical analysis of the optimal scope of government and private provision is now available to be incorporated into Le Grand's analytic framework.^[2,16]

Perhaps even more important is the increased production of high-quality empirical research that carefully assesses real-world market outcomes relative to real-world outcomes of various forms of political allocation. As several writers on government failure have noted: The choice between markets and political

allocation is ultimately an empirical question. One prominent example is the rapidly growing, improved empirical research concerning the privatization of state-owned enterprises.^[17,18] Another good example is the improved comparative research of alternative housing systems, where major differences in performance can now be discerned between market-oriented housing systems and various *planning* systems.^[19,20]

REFERENCES

1. Arrow, K.; Debrue, G. Existence of an equilibrium for a competitive economy. *Econometrica* **1954**, *22*, 265–290.
2. Stiglitz, J. *Whither Socialism?* MIT Press: Cambridge, MA, 1994.
3. Coase, R. The problem of social cost. *J. Law Econ.* **1960**, *3* (1), 1–44.
4. Nesselin, T. Housing: the market versus the welfare state model revisited. *Urban Studies* **1988**, *25* (1), 94–108.
5. Gwartney, J.; Stroup, R. *Economics: Private and Public Choice*; The Dryden Press: New York, 1997.
6. Ekelund, R.; Tollison, R. *Economics: Private Markets and Public Choice*; Addison-Wesley: New York, 2000.
7. Wolf, C. A theory of nonmarket failure. *J. Law Econ.* **1979**, *22*, 107–139.
8. Wolf, C. *Markets or Government: Choosing Between Imperfect Alternatives*; MIT Press: Cambridge, MA, 1988.
9. Weimer, D.; Vining, A. *Policy Analysis: Concepts and Practice*; Prentice-Hall: Engelwood Cliffs, NJ, 1992.
10. Mitchell, W.; Simmons, R. *Beyond Politics: Markets, Welfare, and the Failure of Bureaucracy*; Westview Press: Boulder, CO, 1994.
11. Adachi, Y. Inefficiencies in public policies. *J. Comp. Policy Anal.* **1999**, *1* (2), 225–236.
12. Shleifer, A. State versus private ownership. *J. Econ. Perspect.* **1998**, *12* (4), 142.
13. Le Grand, J. The theory of government failure. *Br. J. Polit. Sci.* **1991**, *21*, 423.
14. Vining, A.; Weimer, D. Government supply and government production failure: a framework based on contestability. *J. Public Policy* **1991**, *10* (1), 1–22.
15. Dollery, B.; Worthington, A. The evaluation of public policy: normative economic theories of government failure. *J. Interdiscip. Econ.* **1996**, *7* (8), 27–39.
16. Shleifer, A. State versus private ownership. *J. Econ. Perspect.* **1998**, *12* (4), 133–150.
17. Vining, A.; Boardman, A. Ownership versus competition: efficiency in public enterprise. *Public Choice* **1992**, *73*, 205–239.
18. World Bank. *Bureaucrats in Business*; Oxford University Press: London, 1995.
19. Nesselin, T. *Housing: The Market Versus the Welfare State Model Revisited*; Urban Studies, Glasgow, Scotland, 1988.
20. Angel, S. *Housing Policy Matters: A Global Analysis*; Oxford University Press: London, 2000.

Government Performance and Results Act of 1993: A Retrospective

G.L.A. Harris

Mark O. Hatfield School of Government, Portland State University,
Portland, Oregon, U.S.A.

INTRODUCTION

The quest for government performance and accountability took its most recent form as the Government Performance and Results Act (GPRA) of 1993. There have been similar, numerous attempts to make the federal government in the United States accountable since the 1940s. This selection will explore the metamorphosis of government performance and accountability at the federal level; assess whether or not the federal government has achieved its targeted objectives; and achievement or not, whether the federal government can incorporate lessons learned to improve performance and accountability.

WHAT IS THE GPRA OF 1993?

The GPRA of 1993 (PL-90-481) sets four basic goals to improve: 1) the public's confidence in the federal government; 2) the effectiveness, efficiency, and accountability of federal programs by focusing on results; 3) "... congressional decision-making by providing more objective information on achieving statutory objectives" (p. 4); and 4) the internal management of the federal government.^[1] Thus, the goal of GPRA, according to Committee Chairman Tom Davis (R-Virginia), is to refocus the federal government's attention on performing and accomplishing results rather than on the process.

GPRA mandates that each federal administrative agency develop three plans for submission to Congress. The first is a five-year strategic plan to be updated either every three years or as suggested or required. The second, an annual performance plan, details the shorter term goals and measurements of the five-year strategic plan. Additionally, the necessary resources, including budget and human capital, are delineated in how goals will be achieved. Third, an annual performance report, states whether and how the intended goals were actually achieved.

GPRA strives to increase transparency by aligning government accountability and performance with the budgetary process. According to former President Bill

Clinton, "The law simply requires that we chart a course for every endeavor that we take the people's money for, see how well we are progressing, tell the public how well we are doing, stop the things that don't work, and never stop improving the things that we think are worth investing in."^[2]

GPRA IN THE MAKING

Federal government leaders have given the highest level of government its historic role as a model organization for lower levels of government and even the private sector. As a result, and more actively since the late 1940s, federal policymakers have engaged in a series of administrative reforms to deliver services more efficiently and cost effectively. Previous reforms have come in the forms of the Lodge-Brown Act of 1947 under President Truman's Hoover Commission; the Johnson Administration's Planning Program Budgeting System (PPBS); Management By Objectives (MBO) during the Nixon era; President Carter's zero-based budgeting; and the Reagan Administration's Grace Commission.^[3-6] To that end, GPRA represents one of the largest and latest in a long line of statutes directed at government reform.

President Clinton's National Performance Review (NPR), later in the 1990s, was an ambitious reform in the federal government declaring that "the era of big government is over."^[7] Based largely on principles touted in Osborne and Gaebler's^[8] *Reinventing Government*, the National Performance Review (NPR) embodied goals to improve performance within the federal government. These goals included evaluating the efficiency of agencies; identifying and eliminating waste within those agencies; streamlining the federal civil service system; changing the culture within the federal government; and empowering federal government employees.^[9] Reinventing government springs from the argument that the federal government should be decentralized to unleash the "entrepreneurial spirit" of employees who would, in turn, change the culture of their work environments.^[9] However, government cannot be run like a business because it

is a “fundamentally different” structured sector.^[8] What is needed is better government or one that promotes better governance.

Finally, in a bold attempt to revolutionize how we conduct the business of government, President Clinton signed into law the GPRA on August 3, 1993. The new initiative, under the leadership of the then Vice President Al Gore as the NPR, sought to improve the way that the citizenry received services by instructing federal agencies to solicit feedback from customers to find out what they want, establish standards for the delivery of services, and benchmark “best in class organizations” in the public and private sectors providing a wide variety of services.^[11]

But, unlike previous reforms, GPRA emphasized the output and outcomes of government programs and activities, not processes.^[12] Focus must, therefore, be on what the private sector calls the “the bottom-line.”^[15] NPR served as a major, and probably the most prominent, arm of GPRA by calling for federal agencies to accomplish the following:

- Establish customer standards to improve products and services.
- Establish performance agreements between the President and each federal agency Cabinet Secretary. Agencies must establish contractual agreements and cascade them to subordinate units in the form of performance goals and objectives that are strategically aligned with each organization’s mission.
- Create “reinvention labs” (p. 75) in federal agencies to experiment with ways to streamline processes and reduce or eliminate the amount of controls exercised by agency heads over their organizations.
- Forge performance partnerships between the state and local governments with the federal government by granting greater flexibility to states and local governments in administering federally funded and/or mandated programs, in exchange for more accountability and results.

The Clinton Administration believed that “citizens deserve to receive the highest quality services for their tax dollars.”^[10] GPRA did not call for cost-cutting measures per se, but its simultaneous enactment with NPR occurred with the push for a balanced budget.^[13] It was, therefore, perceived that GPRA served as a continuation of NPR’s goal to do more with less.

The intent was that accountability at the federal level would trickle down to state and local governments as citizens became intolerant of the practice of business as usual. This kind of accountability, it was believed, would be best achieved by benchmarking the best practices of “best in class organizations” within the public and private sectors.^[11]

Reinventing government changed the way that the federal government chose to conduct business. First, a reduction of the federal workforce by 240,000 reduced the size of 13 of 14 federal agencies and rendered more than 2000 field offices obsolete and eliminated some 200 programs.^[7] Second, \$118 billion in savings were derived from changing the manner in which the federal government operates by reducing the costs of purchasing items to reduce the costs of processes for the most frequently used products and services.^[7] Further, four consecutive years of deficits resulted in an accumulated reduction of \$476 billion for the federal government.

Third, service to the public became more customer oriented, rivaling that of organizations within the private sector.^[7] For instance, a survey revealed the superior performance of the Social Security Administration’s 1–800 telephone service. Fourth, the federal government eliminated unnecessary regulations and forged partnerships with other federal agencies.^[7] This move resulted in the elimination of some 16,000 regulations while simplifying another 3,000.

Lastly, states are asked to experiment with reform in such areas as healthcare and welfare.^[7] In essence, government is attempting to improve working with and in the communities that it serves.

EVALUATION OF GPRA TO DATE

The GPRA of 1993 forced accountability for program, policy and agency results by systematically holding federal agencies accountable for achieving program results; federal agencies must focus on results, not the process to increase efficiency, effectiveness and accountability; federal agencies must provide more objective information based upon statutory objectives to improve decision-making by Congress; and federal agencies must improve their internal management.

GPRA, sometimes known as the “stealth law,” came with great expectations given its anticipated accomplishments.^[14] During the hearing, John Koskinen of the Office of Management and Budget (OMB) said, “If GPRA carries through on its promise, I am confident it will ultimately be viewed as one of the most important pieces of legislation enacted in this decade and will change the dialogue between the public, the Congress, and the Executive Branch about government programs and their goals, performance, and funding” (p. 13).

However, Congress acknowledged that one of the difficulties accompanying any major legislation is its implementation and lack of resources in ensuring successful implementation. From the start, problems emerged. Koskinen conceded that approximately seven of the 77 federal agencies originally part of the pilot on

GPRA, dropped out citing overestimation of their capacity to meet GPRA's performance measurement goals. The OMB asked that these agencies justify why managers experienced such difficulties. Nevertheless, Koskinen added, that the OMB and other agencies could learn more from these failures than from successes.

At the same hearing, Johnny Finch of the General Accounting Office (GAO), envisioned five challenges in implementing GPRA.

- Getting the buy-in and ongoing commitment from agency heads. Part of the problem was due to high turnover as many of the agency principals are political appointees. Further, the average tenure of these appointees is two years. GPRA's goal is to sustain commitment throughout the tenure (s) of multiple political appointments.
- Ensuring that participating agencies have the capacity to meet the performance measurement requirements of GPRA.
- Installing change incentives to facilitate a move from process orientation to outcome orientation. GPRA prescribed that a change in behavior required securing and sustaining the commitment of agency heads.
- Incorporating GPRA into the daily operations of agencies. It is believed that many federal agencies are far from identifying and using outcome measures as intended by GPRA.
- Congressional oversight. Effective Congressional oversight is crucial in ensuring agency accountability for outcomes using performance measurement. Further, Congress is a major stakeholder in ensuring that GPRA is correctly implemented.^[15] Therefore, lack of oversight could prove fatal.

Given these challenges, Congress employed a phase-in approach to GPRA. Implementation would be incremental and it is crucial that agencies identify factors that pose the greatest threat to implementation. Some agencies cited environmental factors that were unique to them. Two such environmental factors were the economic expansion and contraction experienced during business cycles and the delivery and provision of certain services through third parties.^[16]

Congress mandated that agencies develop strategic plans by September 1997; submit annual performance plans following the President's budget in the beginning of 1998; and in the fiscal year 1999, begin to report to the President and Congress on their progress towards agency performance in accordance with goals for that period. Strategic plans were first developed for each five-year period in alignment with agency mission, strategic goals and how these goals were to be achieved. Strategic plans were then to be incorporated into the

daily activities of each agency by identifying performance indicators, performance goals and the level of resources needed to meet them. Given the regulatory nature of many agencies such that results would not be apparent for long periods, Congress endorsed the use of intermediate goals. Finally, annual performance reports served as a way to determine whether intermediate goals were being achieved.

A GAO (1997) study of five federal agencies: Occupational Safety and Health Administration (OSHA), the Federal Aviation Administration (FAA), the Food and Drug Administration (FDA), the Internal Revenue Service (IRS) and the Environmental Protection Agency on what enabled yet impeded attempts to focus on results rather than inputs, found the following.^[16]

- GPRA forced agencies to move from a culture of process-orientation to one of results-orientation.
- GPRA enabled agencies to receive training for developing performance measures. Training helped to secure feedback from stakeholders, to include Congress, the public, private industry, other agencies at the federal, state and local levels of government, and various interest groups.
- Some agencies cited a conflict of interest given their diverse clientele compounded by what the agencies deemed as "inconsistent statutory requirements" (p. 13). This made it more cumbersome in determining what the agencies' goals should be.
- As is the case with the IRS, given enforcement responsibilities, federal law prohibits the evaluation of employees on the basis of these results.
- Data collection was challenging in determining what data was needed to improve results. Some agencies, like the FAA, found itself trying to identify negative data to justify the degree of improvements that have been made to avert accidents through the improvement of, say, airport runways. Another difficulty was, given the complexity of factors that may potentially influence an agency's decision, and due to the lack of control over these factors, performance measures were difficult to establish. The long incubation period for achieving results was another frustration for some agencies.
- Some agencies experienced inconsistencies in priorities including some Presidential initiatives vying with new and old agency initiatives and strategic plans between the process-orientation of those priorities and the results-orientation called for by GPRA.
- Limited resources constrained some agencies' capacity to fulfill many requirements of GPRA. For example, OSHA was limited to conducting no more than two studies per year to determine the impact of certain regulations on workplace safety. The FDA, too, found it difficult to transfer

resources from “high priority risks” (p. 17) to develop a statistically credible baseline for measuring industry compliance.

While the 1997 study represented an undertaking that was early in the GPRA process, the initial drop-outs from the pilot of 77 federal agencies and the current five agencies experiences, indicated that GPRA’s implementation was disappointing.

According to a Congressional Research Service (CRS) report,^[17] federal agencies have a long way to go in meeting the requirements of GPRA.^[18] First, analyses by the CRS identified problems from data systems that were too incomplete to evaluate outputs and outcomes; limited resources and strategies to achieve goals; and insufficient coordination with fellow agencies. The CRS recommended that committees oversee and determine whether agency goals align with policy objectives, that performance measurements be specified in GPRA and that committees monitor agencies to ensure that they are measuring and meeting performance goals.

Second, the GAO argued with the CRS that “... progress is needed in how agencies ... set ... a strategic direction, coordinat(e) ... crosscutting programs, and ensur(e) ... the capacity to gather and use performance and cost data.”^[19] Several heads of House Committees also requested that the OMB ask that the federal agencies align their fiscal year 1999 budgets to the goals in their strategic plans. Congress mandated that these performance goals be measured in quantifiable and objective terms with information as to how the results, based on these goals, could be verified and validated. Further, the findings were that a “culture change” (p. 4) was needed together with the Congressional oversight and leadership of the OMB.

Third, the GAO’s evaluation of agency performance plans for fiscal years 2000 and 2001 found that approximately 60 percent of 35 of the fiscal year 1999 annual performance reports did not align funding requests with explicit methods for reaching performance goals.^[20]

The OMB later called for agencies to link budgetary requests and estimates with performance goals.

Christopher Mihm, also of the GAO, stated that while the importance of GPRA is widely known, it has not necessarily been accepted and is far from inculcated into the federal government.^[19]

The move from process-orientation to results-orientation has been difficult to cultivate given the lack of leadership. Many federal agency middle managers perceived that there is no commitment by top management to such goals. This, they say, would require a “cultural transformation” (p. 12).

Senator Fred Thompson (R-Tennessee) complained that after two rounds of performance reports for fiscal

years 1999 and 2000, the results had revealed that there is much left to do in carrying out a cultural shift. Thompson observed that state and local governments had been more successful. He noted that the federal government had not lived up to the goals that continue to burden the system and stifle its ability to change. The senator admitted that federal agencies had progressed in the right direction, but at a very slow pace.

Fourth, the GAO (2001) confirmed some previous findings and ongoing difficulties with implementing GPRA.^[21] These findings concluded that less than half of the managers in at least 11 federal agencies sensed little to no commitment from agency leadership. Less than half of the managers at 25 agencies felt that employees received systematic recognition for helping to develop performance measures that related to agency strategic goals. Less than half of the managers in at least 22 agencies were asked to be accountable for programmatic results with only one agency showing that more than half of its managers felt that they had the decision-making authority to achieve strategic goals. However, 84 percent of agency managers surveyed felt more involved in performance measures. This represents a 12 percent increase over the 1997 GAO study.

According to the GAO (2001), the federal government has failed to realize the importance of human capital in implementing GPRA. Specifically, the government failed to recognize that succession planning could identify talented and qualified personnel for continuity in leadership. Agencies had not made a significant effort to align leadership performance expectations to agency mission. Problems were compounded by the absence of competitive compensation schemes to recruit, reward, motivate, and retain quality personnel. Managers failed to assess agency staffing needs and strategically deploy personnel to meet these needs. Few linked performance expectations with core competencies to achieve results. And, most agencies failed to collaborate through collective bargaining to achieve results.

Fifth, and on the positive side, GPRA is considered to be an excellent tool for helping to identify management capabilities and problems that may adversely impact an agency’s mission.^[21] Additionally, the GAO notes that it is equally important for agencies to identify their capacity to meet performance goals and link budgetary requests to meet results. Further, where there is the potential for overlap or crosscutting of functions, effort should be made to coordinate or engage in partnerships with other agencies to eliminate duplication and fraud, waste, and abuse. The GAO argued that managers must use credible performance data that is verifiable and valid.

The Bush Administration has attempted to bolster accountability and thus GPRA, through the Program Assessment Rating Tool (PART) to link funding

resources to performance.^[22] PART, which employs data from GPRA to help the Executive branch make budgeting decisions, uses a programmatic approach requiring a five-year phase-in implementation period by federal agencies. PART has forced federal agencies to generate performance measures that justify funding. PART assists agencies in focusing on where their most scarce or most important resources are needed to meet strategic mission goals. GPRA and PART are, therefore, viewed as complementary although PART primarily serves as a budgeting tool for OMB and the President. PART represents an exercise that integrates budget decisions with performance results; GPRA focuses on how performance goals are achieved. PART signals a new wave in accountability and performance and helps decision-makers focus on pressing policy issues while considering the costs and risks of doing so. Finally, despite the deficiencies, GPRA has shown some promise.

Maurice McTigue, of George Mason University, said that the impact of GPRA appears to be taking hold.^[22] He cautioned, however, that his optimism is based on only four years of data and the experience of various federal agencies. Further, this information is inconclusive.

- More agencies have become familiar with GPRA and received the attention of senior leadership.
- The most recent round of annual performance reports reflecting the strategic goals for the fiscal year 2002 have shown marked improvement over fiscal year 1997.
- Agencies have become experienced in managing performance measures to produce desired results. Agencies now do a better job of quantifying their services but still have problems translating these measures to determine the benefits to clients.
- As agencies become more competent in implementing GPRA, they improve the performance indicators that measure outcomes.

McTigue recommends that, in concert with GPRA, PART can be used as an evaluation tool to gauge budgetary decisions based on outcomes. However, he cautions that PART should not replace but be used in concert with GPRA to achieve strategic goals. He also views the President's Management Agenda (PMA) as the most recent technique to augment PART in helping federal agency managers recognize that their budgets are at stake and any allocation is directly tied to agency and/or program performance. McTigue believes that full disclosure of this information will help to determine which stakeholders are being best served and use the information to make better resource allocation decisions in the future.

More recently, the GAO (2004) found some evidence of improvement under GPRA.^[23] Interviews

with focus groups and political appointees as well as OMB officials credit GPRA with helping them through the planning and reporting process toward developing results-oriented performance measures and making use of such measures. Although the reviews are still mixed, more federal agency managers, particularly those at the Senior Executive Service (SES) level, believe that the OMB is acknowledging their efforts. By requiring that agencies develop five-year strategic plans from which annual plans and annual performance reports are based, agencies are forced to align goals and objectives to the mission through performance indicators to achieve results. Failure to do so would result in an agency falling short of its annual plan that is reflected in its annual performance report. This, in turn, could result in Congress' decision to decrease funding to that agency.

However, while the agencies have made remarkable strides in acceding to the spirit as well as the letter of GPRA, they still have a long way to go.

Great expectations were associated with the enactment of GPRA. However, as with any major legislation, significant and unexpected growing pains became evident in its implementation.

- Difficulties in securing the commitment of agency principals was only one of the many challenges encountered.
- The goal that federal agencies must focus on results rather than on process has been an ongoing struggle. This required a culture change for those at the helm as well as those directly charged with GPRA's implementation.
- Many agencies were ill equipped to determine their performance measures much less develop five-year strategic plans. Further, agencies failed to align performance expectations with agency missions.
- Federal agencies were to improve their internal management but lacked the resources to ensure successful implementation.
- Yet, many federal agencies have found GPRA to be instrumental.
- The GPRA helps to identify what they can and cannot do as well as the potential impact on agency mission.
- The Bush Administration has recognized the importance of GPRA by aligning agency performance with budgetary decisions through PART.
- There has been evidence that since GPRA, agencies' familiarity and experience is beginning to improve performance measures that agencies develop.

CONCLUSIONS

This article has attempted to unearth the many gallant trials at reforming government by holding it

accountable for performance. From the 1940s through the 1990s, such reforms have attempted to move the federal government ever closer to efficiency and effectiveness. GPRA represents one of these attempts.

Benchmarking, when used appropriately, has proven to be effective. Moreover, when goals are presented in a clear manner and systematically measured and tracked against best practices, government works.^[24] But unless the leadership from which such benchmarks emanate embraces change by preparing the organization for a cultural shift in thinking as well as in doing, implementation will be feigned and doomed to fail. Additionally, if implementation occurs and organizations are ill-prepared, public managers will continue to use outdated and familiar modes of thinking and doing in responding to new challenges. Such is the case with GPRA. When public managers are equipped with the right tools and environment in which to succeed, they prosper. The key is finding these tools.

REFERENCES

1. *What Happened to GPRA? A Retrospective Look at Government Performance and Results*. Hearing Before the Committee on Government Reform, House of Representatives, 108th Congress, 1st Session, Serial No. 108-75; 2003.
2. *Reaching Public Goals: Managing Government for Results. A Resource Guide*, National Performance Review; 1996; <http://govinfo.library.unt.edu/npr/library/papers/bkgrd/cover.html>.
3. McNeil, N.; Metz, H. *The Hoover Report 1953–1955*, United States Commission on Organization of the Executive Branch of the Government (1947–1949); MacMillan Company, 1956.
4. Malek, F.V. *Washington's Hidden Tragedy*; The Free Press; 1978.
5. Carter, W.H.; Meenach, L. Zero-based budgeting and program evaluation. *J. Rehabil.* **1977**, *43* (1), 18–48.
6. Peters, G.G.; McGregor, E.B.; Goldenberg, E.N.; Rosen, B. *The Unfinished Agenda for Civil Service Reform: Implications of the Grace Commission*; The Brookings Institution, 1985.
7. Gore, A. *The Best Kept Secrets in Government*; National Performance Review; Government Printing Office: Washington, DC, 1996.
8. Osborne, D.; Gaebler, T. *Reinventing Government*; Penguin Books, Ltd, 1992.
9. Dilulio, J.J.; Garvey, G.; Kettl, D.F. *Improving Government Performance: An Owner's Manual*; The Brookings Institution, 1993.
10. Keehley, P.; Medlin, S.; MacBride, S.; Longmire, L. *Benchmarking for Best Practices in the Public Sector*; Jossey-Bass Publishers, 1997.
11. Gore, A. *Serving the American Public: Best Practices in Performance Measurement*, National Performance Review, Benchmarking Study Report; 1997.
12. Ingraham, P.W.; Joyce, P.G.; Kneedler, D.A. *Government Performance: Why Management Matters*; John Hopkins University Press, 2003.
13. Radin, B.A. The Government Performance and Results Act (GPRA): hydra-headed monster or flexible management tool? *Public Administration Review* **1998**, *58* (4), 307–316.
14. The Government Performance and Results Act. Hearing Before the Subcommittee on Government Management, Information and Technology of the Committee Reform and Oversight, House of Representatives, 104th Congress, 1st Session, 1995.
15. Groszyk, W. Implementation of the Government Performance and Results Act of 1993, *Performance Management in Government: Contemporary Illustrations*, No. 9, United States Office of Management and Budget, 1996; 71–86, Chapter 4.
16. *Managing for Results: Regulatory Agencies Identified Significant Barriers to Focusing on Results*, GAO/GGD-97-83, Report to Committee on Governmental Affairs; U.S. Senate, and Committee on Government Reform and Oversight, U.S. House of Representatives, U.S. General Accounting Office, 1997.
17. Genevieve, J.K. *Government Performance and Results Act: Implementation and Issues of Possible Concern*, 97-1028, Congressional Research Service (CRS) Report for Congress; 106th Congress, Resources, Science and Industry Division, 1997 (updated March 30, 1999). <http://www.ncseonline.org/NLE/CRSreports/risk/rsk-21.cfm>.
18. *The Results Act: Has It Met Congressional Expectations?* Hearing Before the Subcommittee on Government Efficiency, Financial Management and Intergovernmental Relations of the Committee on Government Reform, House of Representatives, 107th Congress, 1st Session, Serial No. 107-75, 2001.
19. *The Government Performance and Results Act: 1997 Governmentwide Implementation Will Be Uneven*, GGD-97-109, Report to Congressional Committees; U.S. General Accounting Office, 1997.
20. *Seven Years of GPRA: Has the Results Act Provided Results?* Hearing Before the Subcommittee on Government Management, Information, and Technology of the Committee on Government Reform, House of Representatives, 106th Congress, 2nd Session, Serial No. 106-245, 2000.
21. *Managing for Results: Using GPRA to Assist Oversight and Decision making*. GAO-01-872T, Testimony Before the Subcommittee on Government Reform, House of Representatives. U.S. General Accounting Office.
22. *Should We Part Ways with GPRA? A Look at Performance Budgeting and Program Review*. Hearing Before the Subcommittee on Government Efficiency and Financial Management of the Committee on Government Reform, House of Representatives, 108th Congress, 2nd Session, Serial No. 108-144, 2004.
23. *Results-Oriented Government: GPRA has Established a Solid Foundation for Achieving Greater Results*, GAO-04-38, Report to Congressional Requesters; U.S. General Accounting Office, 2004.
24. Holzer, M.; Callahan, K. *Government at Work: Best Practices and Model Programs*; Sage Publications, 1998.

Government–Business Relations in East Asia

Sung Gul Hong

Department of Public Administration, Kookmin University, Seoul, Korea

INTRODUCTION

Government–business relations refer to the enduring institutional relationship between government and business in the economy and society. There can be a variety of forms in shaping the relationship depending on the particular historical and industrial development experiences. Culture and institutional legacies are certainly important in shaping the relationship as well.

East Asian countries, mainly refer to Japan, South Korea, Taiwan, and China in this paper, have some important commonalities in their government–business relations. In addition to the cultural affinities and geographical proximity, they are typical late developers, meaning that the government intervenes the market intentionally to achieve rapid economic and industrial development. Because the economic performance is a direct outcome of business activities that are closely related to government policies, government–business relations are critical in understanding East Asian economic development. Despite the debate on whether the state or market has been the primary engine of development, it is apparent that the government has been very active in promoting particular industries and economic development in general. Yet, there are also important differences in government–business relations mainly due to the diversities in political regimes, ideologies, economic institutions, and political experiences as will be discussed later in this paper.

At the center of the rapid economic development in East Asian countries, there were very strong authoritarian governments. The authoritarian government has been insulated from society, according to which it could effectively control interest group activities, especially repressing the labor union activities while maintaining close ties with the business. By doing so, East Asian developmental states could secure the labor peace, one of many preconditions for late development. As such, the government–business relationship in East Asian countries was something like “corporatism without labor” as Pempel and Tsunekawa once pointed out.^[1] But the close government–business ties could be maintained only when the business continued to cooperate with the government to achieve the preset and widely shared developmental goals. Otherwise, the government might severely punish the business with a variety of policy tools and institutional mechanisms inherent in the authoritarian regime.

Government–business relations, however, has not been constant over time and across industrial sectors. In fact, the very economic success of East Asian countries has been a fundamental source of the changing government–business relationship. As development proceeds, the business becomes more powerful and influential in the capitalist economy. State autonomy has waned over time, and the government became increasingly dependent on the performance of the business. Industry associations, once played the role of close partners to implement government policies, began to make stronger voices to reflect their particular interests in policy formation. In addition, economic affluence provided more opportunities for the mass education. The more the people got educated, they began to resist the political dictatorship that was common in Taiwan and Korea. As such, political democratization was in due course, according to which the demand for redistribution grew rapidly.

Recently, wide application of Information and Communications Technology (ICT) in everyday lives makes political participation much easier than before by reducing transaction cost of political participation dramatically. The development of ICT infrastructure and the electronic government stimulate active political participation, and active interest group representation both through online and offline become prevalent as in the western democracies.

In the following sections, I will begin with the explanation of typical government–business relations in East Asian countries under the broad context of the developmental state. Overall policy environment and institutional mechanisms that allow a dominant position to the government over the business will be analyzed in detail. The impacts of information technologies upon society will be discussed in order to forecast the future of government–business relations in East Asia.

GOVERNMENT DOMINANCE OVER BUSINESS: AN OVERVIEW OF GOVERNMENT–BUSINESS RELATIONS IN EAST ASIAN COUNTRIES

Government and Business in Preindustrial Stage

Before modernization, a strict status system was maintained in all East Asian countries, in which the aristocratic class monopolized political and military power.

Business people were usually viewed as lower than agricultural and/or manufacturing classes although they were much richer than the others. Commercial activities were usually regulated by the government, meaning the government granted privileges to certain people to maintain commercial activities in specific fields and/or areas. In particular, government procurement needs and official trades with foreign countries could be met only by those who were granted the privileges by public officials. Public officials might use discretionary power in designating particular businesses as suppliers to the government. In return for the privilege, businessmen had to pay the political fund which was needed by the aristocracy. This kind of exchange relationship was both formal and informal, and thus it might be a major source of corruption. The growing competition among business groups to acquire the privileges made the bribery to public officials a common phenomenon, which is typically seen in the predatory state as pointed out by Evans.^[2] A close tie between political elites and businesses were maintained for long period of time, sometimes over many generations, and became institutionalized with varying degrees from countries to countries and from time to time. Thus government–business relations in East Asia during the preindustrial era can be characterized by a long-term enduring cooperative relationship based upon the exchange of political protection for business privileges and the ruling fund to the aristocracy.

Industrialization and Government–Business Relations

Ever since the introduction of the capitalist market economy, governments in East Asian countries played leading roles in the economy mainly to catch up the advanced industrial countries. Although there has been much debate about the role of the state in East Asian countries, scholars generally agreed to the argument that the state was critical in achieving the rapid industrialization.

Japan, for instance, industrialized itself after the Meiji Restoration in 1867, and became a colonial power in East Asia. The ruling elites in the early constitutional monarchy needed cooperation from the business to implement their expansionist policies toward Korea and Manchuria. Later it was changed into the military-based totalitarian regime, and maintained very close ties with big businesses called the *zaibatsu*. The government provided them with ample opportunities to maintain lucrative businesses in the colonial Korea and Manchuria. In return, those businesses served as material bases for the war economy since the 1930s. After the World War II, the Japanese government repeated its prewar legacies to revitalize

its devastated economy. The Ministry of International Trade and Industry (MITI), was at the center of the mixed market economy as Johnson argued.^[3] The collaboration between the MITI and the businesses was essential in bringing about the Japanese resurgence as an economic powerhouse in East Asia in the postwar period. Industrial policy and administrative guidance were both formal and informal instruments that enabled the government to exercise power and influence over the business and industry associations. Stable political regime under the ruling Liberal Democratic Party (LDP) also contributed to the continuation of close government–business ties.

Korea resembled the Japanese type of government–business relations partly due to the colonial legacies. Japanese corporate management style and *zaibatsu* type of business structure were introduced in Korea throughout the colonial period (1910–1945), and the founders of virtually all Korean big businesses were former workers of the Japanese firms operating in Korea. Japanese colonial rule and later the Korean War deprived the landed upper class and the ruling class of the opportunity of exercising power and influence over the formation of the capitalist market economy after 1945. In addition, the American Military Government and the subsequent Rhee government privatized most of the assets formerly owned by Japanese government and corporations during the colonial period. Because there were not many people who could run modern businesses, the privatization simply meant a creation of the new capitalist class in the Korean society (for the legacies of Japanese colonial rules, see Ref.^[4]).

The following Park Chung Hee government (1961–1979) was a military authoritarian regime, and undertook aggressive export-oriented policies toward rapid economic development. Every year, export targets for various industries and even those of big firms were set, and individual firms were forced to meet the targets. The heavy and chemical industrialization adopted in the 1970s was the apex of state dominance over the business, which resulted in a rapid concentration of capital. During this period, the government selected six strategic industries (steel, chemical, non-ferrous metal, machine building, shipbuilding, and electronics), and firms in those industries were financed by huge sum of policy loans.

Taiwan shows similar government–business relations although its economy has been more market oriented than that of Korea. Like Korea, Taiwan was a former colony of Japan (1895–1945). Before 1945, Japan promoted agriculture and small businesses, which made Taiwan “the most agriculturally, commercially, and industrially advanced of all provinces of China.”^[5]

When the Kuomintang (the Nationalist Party) entered into Taiwan in 1949, it decided to acquire all

Japanese-owned industrial assets, and made them the material bases for the Party's ruling. The Kuomintang regime emphasized the separation between political and economic spheres to prevent corruption. As a result, major big businesses were directly owned by the state, while small- and medium-sized enterprises were free to compete with each other.

Taiwan's industrialization efforts were similar to those of Korea with much emphasis on agriculture and small businesses. And Taiwan was far more active to invite foreign direct investments than Korea in order to create the vested foreign interests on the island of Taiwan. This was the case because the Taiwanese viewed the security danger was more serious than Korea, where the American troops were stationed.

Taiwanese government adopted a series of 4-year economic plans, which emphasized different aspects such as economic rehabilitation (1953–1956), growth of national income and investment (1958–1961), heavy industry development (1961–1964), and electric appliances and electronics (1965–1968). These plans served as ultimate guidelines of the pathways of the Taiwanese economy for the businesses. Especially, Taiwanese government put much emphasis on developing high-tech industries since the 1970s. Creating the Hsinchu Science-based Industrial Park, the Electronics Research Service Organization, and the Industrial Technology Research Institute were direct government initiatives to invite private investments in high-tech industries. In short, the Taiwanese government was also a typical developmental state, and the government–business relations were cooperative in nature with clear upper hand on the part of the former.

China is the most recent late developmental state, and thus state dominance over the business is still preeminent. China has maintained annual average economic growth rates of 9% since the economic liberation and reform in 1978. In the past, the communist China maintained a planned economy with state-owned enterprises. Since the adoption of the commercial economy with planning (1978–1991) and the socialist market economy (1992–present), many state-owned enterprises were privatized. But key big enterprises are still under the firm control of the state because the government owns significant portion of stocks. Deregulation was the most important cause of rapid economic growth by stimulating free business activities on the part of the private enterprises. China became a member of the WTO in December 2001, which made its economic system more close to global standards. Foreign direct investments to China increased dramatically since then. Since 2006, Chinese foreign direct investment also increased rapidly through M&A in order to acquire advanced technologies and export market at the same time.

Government–business relations in China after the reform and liberalization can be divided into two different periods. The early stage of reform (1979–1990) emphasized the transfer of management under the condition that collective ownership maintained. Enterprises were free to manage themselves to create more profits, and certain portion of the profits could be withheld within the company. This might serve as an incentive for higher productivity. In the 1990s, China adopted the principle of building the socialist market economy. In order to accomplish this, collective ownership of enterprises should be amended (1992–present). Since then, the Chinese government adopted the modern corporate system, meaning that the stock system was introduced to reform the state-owned enterprises. But this does not mean that the ownership of the state-owned enterprises is privatized completely. Rather, the government maintains control over the former state-owned enterprises by holding significant portion of stocks. However, the management is completely separated from the government, and the enterprises can mobilize huge sum of investment funds through the stock market. Recently, M&A is also actively promoted to make globally competitive big businesses. In short, big state-owned enterprises are being transformed into stock-based corporations, while small- and medium-sized enterprises are privatized to allow M&A and other business activities. At the center of these reforms is the strong government which leads the reform toward the socialist market economy.

Sources of Government Dominance Over Business

Government dominance over business has been formed through the historical conditions and circumstances, and strengthened by institutional arrangements that allow the government to perform the roles of the developmental state.

First and foremost, all East Asian governments are strong authoritarian regimes during most of industrialization periods, and maintained stable political regimes. In Japan, the government was not so authoritarian, but the semitenured dominance of the LDP was essential in the collaboration between government and business. This was possible partly due to the strong and long-lasting ruling coalition among the LDP, the agriculture, the business, and financial capital. The Cold War politics and outstanding economic performance of the ruling party also contributed to the political stability of the LDP regime.

Korea and Taiwan shared geopolitical characteristics, which contributed to the formation of the military authoritarian regimes. China also had the Communist Party-dominated strong government. This certainly

contributed to the government dominance over the business.

Second, at the early stage of industrialization, the government was the only organization to push industrialization ahead with necessary knowledge and information. Private businesses were usually small and underdeveloped, and they did not have appropriate information and knowledge for systematic industrialization. Even relatively big businesses could not mobilize necessary funds for investment without the government's repayment guarantee. Thus, the government played the role of patron for the infant stage of industries and businesses.

Third, in all the East Asian countries, the best elites in society filled the government, economic institutions, and other related organizations. Strong economic bureaucracy that was filled with the best elites in society made long-term economic plans and implemented them throughout developmental periods. The government also controlled the banking sector that was critical for credit allocation. The best talented bureaucrats with the power of credit allocation to the private sector could dominate over the business with a series of developmental plans to achieve unprecedented economic and industrial growth.

But it is basically cooperative relationship rather than unilateral coercion. Close social networks among elites were also critical in understanding the government–business relations. In fact, most of leaders' positions in society including the government, business, industry associations, financial institutions, political parties, and the Parliament, were filled with the graduates of a handful of best universities. They usually had intimate interactions, and often helped with each other to accomplish developmental goals.

Economic bureaucrats in Japan, and to a less extent in other East Asian countries as well, also enjoyed the benefits of *amakudari*, meaning that former economic bureaucrats could be appointed in high positions of the private business and/or industry associations. *Amakudari* certainly provided additional communication channel between government and business and making the relationship much smoother.

Fourth, East Asian governments possess effective policy tools to check the business and to force them when it is necessary. Usually, the government exercises its influence based upon mutual understanding and collaboration. But when the businesses do not voluntarily follow government policies, the government has effective policy tools to enforce them. The most powerful one would be the prosecutor's authority. Because the owners and managers of big businesses need to maintain close ties with political communities, most of them are not free from political corruption. Another very effective policy tool is the control of the central bank. In the developmental state, the government

effectively controls the banking sector, and bank credits might be distributed based upon policy purposes. For example, the Korean government controlled the domestic and to a lesser extent, foreign loans, and allocated the available credits based upon the developmental needs. China, on the other hand, possesses another powerful policy tool based upon the large state ownership as described above.

Fifth, in all East Asian countries, industrialization began with the import substitution of the formerly imported products. Export-led industrialization of labor intensive items such as textile, apparel, and footwear, and the deepening of import substitution called the heavy and chemical industrialization began at different times and with varying degrees. Japan in the 1930s and again in the 1950s, and Korea in the 1970s were comparable in this regard, while Taiwan did not take full initiative in promoting the heavy and chemical industries. Foreign direct investments were invited as much as possible, while Korea had to build up a huge foreign debt as well in order to make up the needed investment funds. The industrial policies of this magnitude could not be lasted long had it not been for the autonomy and capacity of the government. A very cooperative government–business relationship with the dominance of the government was a prerequisite for late industrialization.

Sixth, international contexts were also favorable to the government–business relations of this type. Korea and Taiwan after the World War II were in a constant security danger. The security alliance with the United States provided diverse geopolitical advantages such as huge military and economic aids as well as privileged access to the American market. When the security commitments by the United States had weakened in the 1970s, Korea and Taiwan had to find out their own ways to maintain the credibility from outside. Korea went into the heavy and chemical industrialization along with the *Yushin* system, a dictatorship by President Park Chung Hee, while Taiwan tried to invite foreign investments as much as possible in order to increase the vested interests of the foreign capitals on the island of Taiwan. Under this situation, the government must be very strong and autonomous, and perhaps dominant over society and business.

China is an exception in this regard. It has been one of the superpowers in the region and never been in the security danger. Major Chinese reforms occurred after the demise of the Cold War politics. Thus, the Chinese economic and industrial policies could be concentrated on the economy.

In short, government–business relations in East Asian countries can be characterized by the strong dominance on the part of the government over the business. Yet, the government dominance should be supplemented by the close cooperation with the business.

BEYOND THE DEVELOPMENTAL STATE

Sources of Changing Government–Business Relations

The very economic success of East Asian countries became a source of decreasing government's power and influence over the business. As the economy grew, the government tried to build up big businesses considering the economy of scale. Big businesses could develop a worldwide production networks and marketing channels, and they no longer follow the government leadership in business decisions. At the same time, the government began to lose its competitive edge in setting specific targets for industries and even individual firms, simply because the market changes so fast and public officials could not get the necessary data and information to make appropriate decisions as fast as the business does.

As economic affluence was in common, people with higher education increased over time. This stimulated the rise of the middle class who were more liberal and active in social issues. Labor unions, which had been exploited and sacrificed for the fast economic development, no longer sit down without resistance. It was around the mid-1980s when full-scale democratization occurred both in Taiwan and in South Korea. Democratization was a fundamental source of changing the role of government and the government–business relations in East Asia. Providing ruling funds in return for favorable economic policies was increasingly difficult due to the growing demand for more transparent political processes. Political stability, which contributed to capital investments by reducing the uncertainty of the future, was disappeared. The honeymoon period between political elites and business leaders gradually came to an end.

In addition, demand for the redistribution of the accumulated wealth grew rapidly throughout society. Various non-governmental organizations began to criticize the concentration of capital by a few big business groups (e.g., *chaebols* in Korea). The close ties between government and business, which once served as an institutional condition for rapid economic growth, were blamed as a source of corruption.

While the power and influence of the government decreased significantly, the business was relatively free to engage in active rent seeking. They no longer follow what the government wanted them to do unconditionally. In some cases, business interests might not be the same as national interests as Vernon once points out in the case of multinational corporations.^[6]

The demise of the Cold War politics in 1989, and the increasing globalization thereafter also contributed to the diminishing role of the government in the economy and society. The geopolitical advantages that East

Asian countries had enjoyed such as the infant industry protection, the strategic industrial targeting, the most favored nation principle, and so on, were no longer available to them. In fact, industrial targeting, once common in East Asian countries to promote specific industries, was largely prohibited by the WTO as unfair trade practices. They were asked to play the role of the advanced countries, and Korea joined to Organization for Economic Cooperation and Development (OECD) in 1995, officially graduated from the status of the newly industrializing country.

In late 1997, East Asian countries, especially Korea, were severely hit by a financial crisis. The government once again played a dominant role during this period although it had to strictly follow the IMF's guidelines. Many local firms with high debt–equity ratio had to be sold out to appropriate buyers, and structural adjustment in the banking sector was followed. As shown in the process of overcoming the crisis, the developmental state in Korea did not retreat but revived. A full-scale economic liberalization was followed thereafter, due to which the government–business relations increasingly became pluralistic comparable to those in the western societies.

Globalization, Information Technology Revolution, and Government–Business Relations

Information technology (IT) development also stimulates the changing roles of government and government–business relations. As IT develops, the role of the state in the information era is being discussed. Because the IT revolution comes with globalization and neo-liberalism, many advanced countries began to adopt IT as a key tool to reform and downsize the government and public sectors. As a result, small but efficient governments with active citizen participation through the cyberspace become popular phenomena. E-government and e-democracy are now changing the mechanism of governing from government to governance. As a reflection of this trend, a special issue of *Daedalus* in 1995 published a series of articles with the titles such as “diminished nation-state,” “defective state,” and “decline of the nation-state.” Articles appeared in *Daedalus* commonly argue that information society and globalization would substantially reduce the role of the government, while those of the private sector such as private firms, financial capital, non-governmental organizations, and individual citizens are relatively increased.^[6]

The formation of WTO was a signal for the world market with free trade principle. Numerous bilateral free trade agreements are being actively negotiated. Due to the IT revolution, economic transaction can

be taken place anytime and anywhere at light speed. Under this changing business environment, the role of government and government–business relations had to be changed dramatically.

Protection of domestic businesses is neither possible nor desirable. In fact, the separation between domestic and international markets (or firms) is no longer meaningful. Government's policy tends to invite more competition, and should guarantee fair competition among businesses regardless of their national origins. Job creation is the number one concern for most countries, according to which governments try to provide as much incentive as possible to invite more foreign direct investments. Even domestic firms will go out to look for better places for investment, which may result in further loss of valuable jobs. In short, the government would not be a dominant player in the economy any more. More diversified roles of government are expected in the future, and so does the government–business relations. Huge multinational corporations can enjoy relative independence from government, while small and medium enterprises need more assistance and perhaps guidance from the government.

CONCLUSION

East Asian countries are typical late developers and share important commonalities in their government–business relations. At the early stage of development, a politically stable authoritarian government dominated the business to achieve developmental goals. Although there were close ties between government

and business, the government exercised a clear upper hand in their relationship. As development proceeds, however, government–business relationship changed gradually from dominance to partnership, due to the growing middle class and democratization that weakened the autonomy and capacity of the developmental state. But the government in East Asian countries may play more direct and aggressive roles whenever it is necessary as shown in the process of overcoming the financial crisis. Overall, roles of government in the economy and the government–business relations seem to be diversified due to the changing business environment such as globalization and IT revolution.

REFERENCES

1. Pempel, T.J.; Tsunekawa, K. Corporatism without labor? In *Trends Toward Corporatist Intermediation*; Schmitter, P., Lehmbruch, G., Eds.; Sage Publications: Beverly Hills, 1979; 231–270.
2. Evans, P. *Embedded Autonomy: State and Industrial Transformation*; Princeton University Press: Princeton, 1995.
3. Johnson, C. *MITI and the Japanese Miracle*; Stanford University Press: Stanford, 1982.
4. Woo, J. *Race to the Swift*; Columbia University Press: New York, 1991.
5. Wade, R. *Governing the Market*; Princeton University Press: Princeton, 1990.
6. Vernon, R. International investment and international trade in the product cycle. *Quar. J. Econ.* **1966**, *80*, 190–207.

Government-Sponsored Enterprises (GSEs)

Stuart C. Strother

Steven G. Koven

School of Urban and Public Affairs, University of Louisville, Louisville, Kentucky, U.S.A.

INTRODUCTION

Government-sponsored enterprises (GSEs) are large financial institutions that function in secondary finance markets for the purpose of making credit more available to borrowers for home mortgages, agricultural loans, and student loans. Originally established by the federal government, most GSEs are now privately owned, although they still retain many characteristics of a public agency. Most GSEs are restricted by law to function only in the secondary market—they purchase or guarantee loans originated by primary lenders such as banks, thrifts, and mortgage banks. Because they have little direct interaction with the general public, GSEs function under the radar screen, but they are some of the world's largest financial institutions. Three of the major GSEs—the Federal National Mortgage Association (Fannie Mae), the Federal Home Loan Mortgage Corporation (Freddie Mac), and the Student Loan Marketing Association (Sallie Mae)—consistently appear in *Fortune's* list of the 500 largest companies. In this article, we describe the creation of the GSEs, explain how GSEs function, explore the public and private characteristics of GSEs, and note some current issues facing GSEs.

THE BEGINNING OF GSEs

The first GSEs, collectively called the Farm Credit System, were created by the federal government in 1916 to help stabilize financial fluctuations in the farm sector of the economy. The individual enterprises that comprise the Farm Credit System as well as other enterprises are shown in Table 1. The Farm Credit System was designed to “encourage the flow of credit for farm mortgage loans . . . make farm operating loans . . . and lend to agricultural producer cooperatives.”^[1] Farm Credit System institutions functioned as “cooperatives owned and controlled by the borrowers.”^[1] Wild fluctuations in agricultural operations were smoothed out primarily through the organizational structure of the Farm Credit System, rather than through outright government subsidies of farmers.

Housing-related GSEs have their origin in the federal government's post-Depression efforts to rebuild

the national economy. After the collapse of the national housing market, which was characterized by numerous mortgage defaults by borrowers, local bankers grew discouraged from investing in home loans. Government officials were convinced that access to affordable housing and higher levels of home ownership represented key components to a national economic recovery from the Great Depression. To spur such a recovery, the Federal Home Loan Bank System (FHLBS) was created in 1932. The objective of the FHLBS was to provide loans, called “advances,” to financial institutions such as community banks. To secure advances, member banks put up collaterals such as traditional mortgages, agricultural loans, and small business loans. The FHLB system consists of 12 banks. The FHLBS is not publicly traded, but is owned by over 8000 member institutions that hold equity stakes. The increased availability of funds made possible by this arrangement allowed member banks to offer more home mortgages to their customers.

Another GSE, the “Fannie Mae,” was created in 1938 as a latent component of President Franklin D. Roosevelt's New Deal.^[2] The goal of Fannie Mae was to enable local banks to make more loans for purchases of private homes. If private banks would agree to make low interest home loans to individuals, Fannie Mae would purchase these mortgages with federal funds, thereby assuming all of the investment risks. Once private home loans were guaranteed by the federal government, local banks were able to extend credit to individuals without danger of loan default. This arrangement worked well and Fannie Mae's purchases of mortgages became known as the “secondary mortgage market.” With the backing and strength of the federal government behind them, private banks again invested in home loans, resulting in higher rates of home ownership.

As part of President Lyndon B. Johnson's efforts to streamline the federal government during the Vietnam era, Fannie Mae was sold to shareholders and became a privately owned enterprise in 1968. Fannie Mae's mortgage-backed securities (MBS) operations were retained in the newly created federal agency, Government National Mortgage Association (“Ginnie Mae”). Ginnie Mae operates the federal government's MBS programs, which are described below. For three

Table 1 Comparison of major GSEs

Enterprise	Established	Ownership	Key function
Farm Credit System			
Federal Land Banks	1916	Privatized in 1968	Encourage credit for farm mortgages
Federal Intermediate Credit Banks	1923		Make farm operating loans
Banks for Cooperatives	1933		Make loans to cooperatives
Federal Home Loan Bank System	1932	Over 8000 member institutions have equity ownership since 1951	Provide advances (loans) to member institutions (banks, credit unions, and thrifts) so they will offer more home loans
Fannie Mae	1938	Publicly traded private company since 1968	Make home loans more available to individuals by purchasing mortgages from primary lenders in the “secondary market”
Ginnie Mae	1968	Federal agency within HUD	Operate the federal government’s mortgage-based securities program
Freddie Mac	1970	Publicly traded private company since 1970	Same as Fannie Mae
Sallie Mae	1972	Publicly traded private company since 1972	Make student loans more available to individuals by purchasing student loans from primary lenders
Farmer Mac	1988	Publicly traded private company since 1988	Guarantees mortgage-backed securities to make agricultural loans more available

Source: Refs.^[1,6]

decades, Fannie Mae had operated as a monopoly in the secondary mortgage market, but in 1970, Congress created another GSE, the “Freddie Mac.” Freddie Mac conducts the same financial functions as Fannie Mae, including buying loans from primary lenders, advancing funds to lenders, and issuing debt obligations (bonds).

In 1972, Congress created another GSE, the “Sallie Mae,” to provide greater financial support for guaranteed student loans. At that time, student loans were perceived as “small, expensive to service, and generally unattractive” to private lenders such as commercial banks.^[1] Like other GSEs, Sallie Mae’s primary function was to provide a “secondary market,” such that primary lenders would be encouraged to invest in student loans. Sallie Mae eventually expanded into home equity loans based on the idea that families sometimes take out home equity loans to pay for a child’s college expenses.

The Federal Agricultural Mortgage Corporation (Farmer Mac) was established in 1988 to act as a guarantor for MBS issued by private lenders and the Farm Credit System. In general, GSEs have used their access to federal agency credit markets to create greater access to capital for borrowers, and have provided greater stability in housing, agricultural, and student loans. Table 1 provides basic information about the function, ownership, and establishment of major GSEs.

HOW GSEs FUNCTION

Government-sponsored enterprises have three major functions: lending money, raising money, and community investment. With the exception of the Farm Credit System that lends directly to borrowers, GSEs mainly lend money in secondary markets. Government-sponsored enterprises “loan” funds to primary lenders (i.e., banks, thrifts, and mortgage banks) by purchasing the home mortgages or student loans that were issued by the primary lenders. Government-sponsored enterprise also make collateral loans, called “advances,” to primary lenders who can use “any acceptable market-worthy collateral” to secure the loan.^[1] Figure 1 illustrates how GSEs interact with primary lenders and investors.

To make direct loans to individual borrowers, to purchase loans from primary lenders, or to offer collateralized loans, GSEs need capital. In the beginning, GSEs received federal government capital contributions, which were later repaid or forgiven once the enterprise was fully privatized.^[1] Government-sponsored enterprises now raise funds by borrowing money through debt obligations (such as Series MM bonds), by directly selling MBS, or by guaranteeing the MBS issued by private lenders. Almost 400 private companies issue MBS, including mortgage companies (70%), commercial banks (14%), and savings and loan

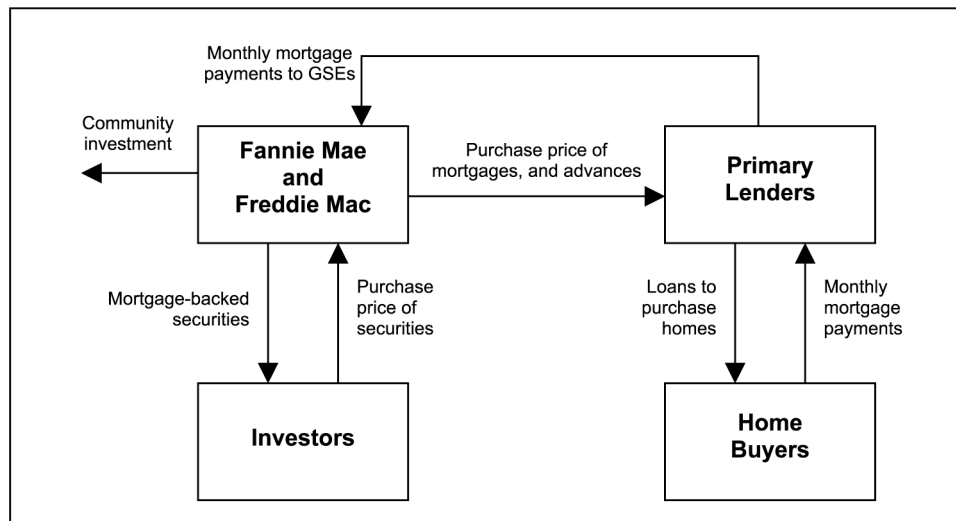


Fig. 1 Interactions between GSEs, primary lenders, and investors. (From Ref.^[2].)

associations (10%).^[3] After these companies have lent money to home buyers in the form of mortgages, a number of mortgages are pooled together and sold to investors in the form of MBS. Once the pooled mortgages are sold, the GSE guarantees payment to investors even if individual home buyers default on their mortgage payment obligations. The 30-year “Ginnie Mae” MBS is a favorite of investors. It can be purchased through any ordinary securities broker for US\$25,000 and has historically yielded 4–6%. Because these are “pass-through” securities, shareholders are paid back their principal, with interest, each month over the life of the security.

In addition to improving communities through increased consumer access to homeownership, GSEs also have contributed through community investment programs. Government-sponsored enterprises have contributed billions of dollars to the affordable housing program and the community investment program. These programs provide housing subsidies for low-income and moderate-income families, and provide funding for community and neighborhood economic development programs.

ARE GSEs PUBLIC OR PRIVATE?

At first glance, GSEs appear to be government agencies, but their core functions and operations have characteristics of both public agencies and private firms.

Notable public agency characteristics of GSEs relate to mission, limitations, leadership, regulation, taxes, access to capital, and quasi-monopoly status. When GSEs were established, their stated mission was to serve a distinctly public purpose—to stimulate and stabilize the economy by providing a secure

secondary market for loans. The congressional charters that established the GSEs also limited them to perform only certain functions. Fannie Mae and Freddie Mac, for example, are “limited to providing a secondary market for residential mortgages” of specified size.^[1] Farmer Mac is limited by its charter to only provide specific types of agricultural lendings. Government-sponsored enterprise leadership also resembles public agencies in that the U.S. President appoints some members to each GSE’s board of directors. Of Sallie Mae’s board of directors, seven are elected by financial institutions, seven are elected by educational institutions, and seven are appointed by the U.S. President.^[1] Government-sponsored enterprises also resemble purely public agencies in that they are not subject to certain regulations such as securities registrations; they are exempt from state and local taxes; and, most importantly, their securities receive preferential treatment due to the implied federal government guarantee. Another public characteristic of GSEs is that they operate as quasi-monopolies—a competitor cannot compete on a level playing field with the GSEs unless they first secured a similar charter from Congress, which is highly unlikely. These public characteristics represent a competitive advantage for the GSEs, making it difficult to assess whether they are efficient providers of financial services, or if they have displaced competitors who are wary of competing against a federally subsidized enterprise.

Notable private characteristics of GSEs relate to ownership and oversight. With the exception of Ginnie Mae, the major GSEs are privately owned corporations. Like other privately owned enterprises, GSE management decisions seek to maximize shareholder wealth. Therefore the balance between the profit motive and social responsibility is a perpetual struggle

for the GSEs. In the 1970s, the HUD Secretary tried to force Fannie Mae to invest more of its resources in subsidizing loans to low-income and inner-city areas. Fannie Mae's supporters claimed that the charter did not require such subsidies, essentially arguing they are not required to give away shareholder money. Government-sponsored enterprises raise money through bond issues like a federal agency, but their decisions to loan money largely mirror a private firm's decision-making process.

CURRENT ISSUES CONCERNING GSEs

Over the past two decades, GSEs have played an increasingly important role in the U.S. economy largely due to changes in consumer investment behavior. Throughout the 1980s and 1990s, consumers shifted their long-term savings from depository institutions (i.e., certificates of deposit at banks and credit unions) into equity investments (i.e., mutual funds and stocks). With their deposits declining, banks sought new sources of money and turned to GSEs, especially the FHLBS.^[4] However, greater reliance on GSEs caused them to come under increased scrutiny and criticism. Government-sponsored enterprises have recently been chastised for wielding too much power, being too politically active, holding too much risk, and lending discriminatorily.^[5]

Those who criticize the GSEs for having too much power claim that the government "subsidies" the GSEs enjoy amount to an unfair quasi-monopoly that prevents healthy competition in secondary loan markets.^[5,6] Another criticism of the GSEs is that they are too politically active. They face only minimal oversight, which perhaps accommodated the June 2003 accounting misdeeds that resulted in the ouster of top executives at Freddie Mac.^[7] Members of Congress have suggested that oversight of the GSEs should be consolidated in one place—the Treasury Department, where the GSEs' complex books are more likely to be understood.^[8] The GSEs have responded to these suggestions by sending an army of high-powered lobbyists (mostly career politicians with little experience in mortgage finance) to Capital Hill to argue on their behalf.^[9]

Another criticism of the GSEs is that they are holding too much risk. In February 2004, Federal Reserve Chairman Alan Greenspan called for actions to prevent the GSEs from further increasing their already US\$3 trillion debt.^[10] Much of this debt comes from the GSEs borrowing money from the U.S. Treasury to fund their purchases in the secondary mortgage market. A related complaint is that GSEs are retaining too much risk in their portfolios by holding on to mortgages rather than pooling them and reselling

them.^[5] In the event a GSE defaults on its own debt obligations, U.S. taxpayers are faced with a bailout much more painful than the savings and loan bailout of the 1980s.

A final criticism of the GSEs is that they have not eliminated discriminatory lending. An empirical study by Myers^[11] concluded that "blacks and other minority group members are more likely than whites to be denied loans because their loans are less likely to be sold on the secondary market to Fannie Mae and Freddie Mac." Therefore the low-income home market is not as well served as it should be. It is likely that the profit motive of the GSEs directs investments away from housing loans for low-income home buyers and toward the more lucrative middle-income and upper-income housing loan markets.

CONCLUSION

Government-sponsored enterprises are large quasi-public financial institutions created by Congress to stabilize specific capital markets, especially the markets for mortgages, agricultural loans, and student loans. The GSEs fulfill this purpose by providing advances to financial institutions and by purchasing mortgages in the secondary market. Government-sponsored enterprises raise funds by issuing bonds and MBS, and by guaranteeing the MBS offered by private lenders. Government-sponsored enterprises have characteristics of both public agencies and private firms. They serve a public purpose, but operate based on private profitability. This hybrid organizational structure allows them to earn lucrative profits in capital markets while enjoying the stability and protection of the federal government's backing. Critics of the GSEs such as Stanton^[1] argue that GSEs have served their public purpose and have concluded their life cycle. When Congress created the GSEs, no workable exit strategy was put in place, and, consequently, GSEs have grown to be powerful financial institutions that will probably continue to play an important role in U.S. capital markets.

REFERENCES

1. Stanton, T.H. *A State of Risk: Will Government-Sponsored Enterprises Be the Next Financial Crisis?* Harper Collins: New York, 1991; pp. 19, 20, 23, 50, 59.
2. Koppel, J.G.S. Hybrid organizations and the alignment of interests: the case of fannie mae and freddie mac. *Public Adm. Rev.* **2001**, 61 (4), 468–482.
3. Ginnie, M. *Our History*; Retrieved on April 7, 2004 from http://www.ginniemae.gov/about/Our_History.asp?Section=About.

4. FHLB. *The FHLBank System: Partners in Community Credit*; 2001, Retrieved on February 4, 2004 from <http://www.fhlbanks.com/Pages/template1.asp?P=131>.
5. Donlan, T.G. Cut fannie and freddie down to size. *Barron's* **2004**, 84 (9), 35.
6. Stanton, T.H. *Government-Sponsored Enterprises: Mercantilist Companies in the Modern World*; AEI Press, 2002.
7. Barta, P.; McKinnon, J.D.; Zuckerman, G. House money—behind Freddie Mac's troubles: a strategy to take on more risk; U.S.-chartered middleman of mortgages transformed into financial highflier. *Wall Str. J.* September 22, **2003**, A1 (Eastern edition).
8. Cochran, J., III. A new watchdog for Freddie and Fannie? *Regulation* Fall, **2003**, 6–7.
9. Murray, A. Fannie, Freddie use political prowess to parry regulators. *Wall Str. J.* June 17 **2003**, A4 (Eastern edition).
10. Serwer, A. Greenspan sounds off: Is it politics or passion? *Fortune* **2004**, 149 (6), 69.
11. Myers, S.L. *The Effects of Government-Sponsored Enterprise (GSE) Secondary Market Decisions on Racial Disparities in Loan Rejection Rates*; U.S. Department of Housing and Urban Development: Washington, DC, 2000; 26 pp.

Governors as Chief Administrators, The Changing Role of

H. Edward Flentje

*Hugo Wall School of Urban and Public Affairs, Wichita State University,
Wichita, Kansas, U.S.A.*

INTRODUCTION

The independently elected chief executive represents a unique American contribution to modern democratic governance. The role of such executives including presidents, governors, and mayors among others in the administration of U.S. governments has evolved dramatically over the course of U.S. history. In the early days of the nation, distrust of government and executive power limited the scope of government and severely restrained the authority granted to elected executives. The progressive movement at the beginning of the twentieth century initiated an enlargement of executive powers, including gubernatorial powers, and this trend has continued in ebbs and flows to the present day. Today, governors—more than any other state officials—are held politically accountable for directing the administration of state government.

BACKGROUND

In its purest form, the role of governor as chief administrator is founded on progressive-era ideals; that is, the public interest in administering state government may best be served through rule of law, hierarchical structures, division of labor, specialized competence, and the insulation of administration from politics.^[1] Rule of law, including written rules and regulations promulgated to interpret the law, assures that administration will be conducted within the bounds of the law. Hierarchical structures unify administration and centralize executive control. Hierarchy also provides the rationale for extending executive control. Division of labor rationalizes executive functions by purpose or process and guides executive reorganization. Specialized competence calls for selection of administrative personnel based on considerations of merit, such as education, experience, and expertise. Insulating administration from politics assures that administration will be conducted free of political influence and corruption by narrow special interests. These ideals provide a model of the “one best way” to conduct gubernatorial

administration and give guidance to the governor’s role as chief administrator.

This model of gubernatorial administration has been applied in state governments at different times and stages, often in fits and starts, throughout the twentieth century. Application of the model received preeminence with the emergence of the modern governorship in the last half of the century. Larry Sabato, who described this transformation of the U.S. governorship as “Goodbye to Goodtime Charlie,” found governors serving between 1950 and 1975 to be “younger and better educated” with preparation for office “more thorough and appropriate” than ever before. Within the executive sphere, these governors were successful not only in “orchestrating constitutional revisions and reorganizations but also in consolidating and fortifying their control of administration.” Most institutional obstacles formerly in the way of governors “have been dislodged and swept away.” Governors gained “appointive powers where it really matters, at the top-levels of state government . . . [T]he strengthening of the executive budget and other planning and management tools have consolidated the control of the chief executive over state administration.” As a result, fewer of this modern new breed of governors “are being defeated because of political and administrative incompetency.”^[2]

Governors in this period became evangelistic in asserting their role as chief administrator of state government. For example, interviews of 15 former governors who served their terms between 1965 and 1979 saw the managerial role to be “the most important facet of the office.” The governors interviewed were emphatic, for example: Edwards of South Carolina: “. . . The management role is really where the meat of the Governor’s office is Without management, you have practically no leg to stand on.” Exon of Nebraska: “To me, that’s number one—being a good manager . . . more important, I think, than programs or anything else.” Rampton of Utah: “. . . if a Governor is going to do his job, he has got to be the manager.” Walker of Illinois: “What counts is management . . . we have to get the Governors to concentrate on

management.” Wollman of South Dakota: “...if a Governor doesn’t devote a high percentage of his time—over half of his time—just to executive management functions, he is really missing the boat.” Each governor interviewed emphasized the governor’s role as manager from his own perspective; none denigrated its import.^[3]

During this period, the same exuberance for the governor’s administrative role permeated the work of the governors’ vehicle for collective action, the National Governors’ Association (NGA). This enthusiasm was transmitted to new governors by NGA through biennial seminars and through *Governing the American States: A Handbook for New Governors*. Over half of the NGA handbook is dedicated to guiding governors on topics such as office organization and management, staff development, selection of cabinet officials, policy and budget development, time management, and reorganization, among other topics. The volume prescribes guidelines for relating to cabinet members, avoiding staff-line conflicts, and keeping records within the governor’s office. The chapter on reorganization states that “reorganization is usually a high priority” and outlines reorganization actions “to increase the Governor’s executive authority and management capacity,” such as expanding the governor’s appointment power, reducing the number of statewide officials chosen by popular election, eliminating fixed terms of office for cabinet members, ensuring appointees serve at the pleasure of the governor, and eliminating boards and commissions that operate state agencies.^[4]

POLITICS OF GUBERNATORIAL ADMINISTRATION

The idealized view of gubernatorial administration shapes the governor’s role as chief administrator but reflects some degree of political naiveté. Governors conduct their administrative role through the exercise of administrative powers, such as the power of appointment, fiscal power, the power of organization, and the power of command. Most state constitutions provide for administrative powers that are *separate*, yet *shared* between the governor and legislature. A governor’s unique constitutional status and independent political base allow the exercise of administrative powers separately from and independently of the legislature. The open-ended constitutional language defining the powers of chief executives gives governors an advantage in the exercise of administrative power. However, administrative powers are shared in the sense that a legislature may also act in the administrative sphere if it so chooses—a situation that makes the exercise of administrative powers competitive and interdependent. Further, as executive powers have

expanded, the legislature’s capacity for oversight of state administration has also increased.

Gubernatorial exercise of administrative powers also immerses a governor in bureaucratic politics. According to one observer, state bureaucracies “tend to be creatures of habit, inflexible, and enmeshed in the process of administering ongoing programs. Around them emerge a complex of interest groups and legislative supporters, all committed to the survival and growth of extant bureaus and their programs.”^[5] These triangular alliances of state bureaus, organized clientele groups, and legislative friends represent potent forces in U.S. politics and a formidable challenge to gubernatorial direction of state bureaucracies. Any governor who takes seriously his role as chief administrator can expect to contest toe-to-toe with the iron triangles of state bureaucratic politics. In sum, state constitutional frameworks and state politics place the governor’s role as chief administrator squarely in the political realm, which may be seen more clearly by examining the primary administrative powers.

POWER OF APPOINTMENT

The power most fundamental to the role of governor as chief administrator is the power of appointment—the power to hire and to fire executive officials. Other administrative powers in large measure flow from the power of appointment. The legislature may share in the appointment powers by enacting legislation that: 1) removes positions from gubernatorial appointment through the creation of independent boards and commissions or independently elected offices; 2) limits gubernatorial removal by placing certain administrative positions under civil service status or giving them terms of office; 3) requires legislative confirmation of gubernatorial nomination to certain positions; 4) establishes specific statutory criteria for appointment to certain positions; or 5) enacts such restrictions in some combination. Constitutional provisions may also restrict gubernatorial powers of appointment, but constitutional and statutory restraints are slowly being removed. In 1955, for example, 337 state officials elected independently from the governor headed state agencies in the 50 states; as of 1994, that number had declined to 260, a decrease of 23%.^[6]

The active exercise of appointment powers also has political ramifications for governors.^a Governors making key administrative appointments based on

^aNumerous references to the political difficulties encountered by individual governors in exercising administrative powers follow and may be found in various sources, including Refs. ^[3,5,7–9].

considerations of specialized competence, managerial skill, education, and experience may cause antagonism within their state party organizations. For example, retaining a competent agency head from a prior administration of the opposing party may offend partisans in the governor's party. Merit-based appointments to cabinet positions may preclude the use of patronage in those departments and, if appointees are recruited from out of state or from outside state service, irritate current state employees. More difficult politically than selection is the removal of key officials. According to one former governor, "Nothing can bedevil an administration more than weaknesses among appointees." According to another, "If they're bad, you get more minuses for their being really bad than you get pluses out of their being good."^[3] Outright firing of a department head in order to change policy direction in a department may unify clientele groups, loyal bureaucrats, and legislative allies to protest gubernatorial action and, if a replacement is not quickly in place, lead to political havoc wrought by a runaway bureaucracy.

FISCAL POWER

The power of the purse—the power of governors to direct or substantially influence levels of state spending and revenue raising—is essential to the governor's role of chief administrator. While state constitutions grant legislatures primary powers over taxing and spending, the fiscal powers of governors evolved significantly during the twentieth century. Through statutory assignment, governors and their appointees now dominate budget making, that is, recommending an annual plan for expenditures, revenues, and year-end balances. Governors may also enforce their fiscal recommendations through use of veto powers—outright veto of appropriations bills, line-item vetoes, or the threat of vetoes. While legislatures jealously guard their control over power of the purse through more detailed line items, appropriation riders, "veto-proof" appropriation bills, and various mechanisms designed to oversee expenditures, governors and their appointees have substantial discretion to shape spending decisions once appropriations are authorized. A number of governors also have statutory authority to exercise spending discretion among line items, impound funds, or make mid-year spending reductions when revenue falls short of projections or for other statutory purposes. Further, state statutes often grant executive agencies "off-budget" fiscal powers, such as raising various fees and charges, accepting non-state funds, and issuing debt, among others, and these powers are susceptible to gubernatorial control or influence.

A governor's struggle with the legislature for control over power of the purse has political consequences. For example, between 1951 and 1975, 21 incumbent governors were evicted from office through elections in which the key issue was a tax initiative of the governor.^[2] Governors can survive a general tax increase, but the survivors list is a short one. On the expenditure side, most studies of state budgetary processes show a contest of competing interests—clientele groups, agency advocates, legislative spokesmen, and other political interests. This contest promotes decision rules for the budget process that do not seriously question existing operations of state agencies. Governors who want to enter this fray and seek substantial fiscal change will be required to expend considerable political capital, and even then, they may be faced with embarrassing budgetary defeats. Another political bugaboo for most governors is the necessity to recommend periodic salary adjustments for state employees that will likely be perceived as too little on one side and too much on the other. Those governors seeking to reward managerial performance may also be stalemated by the egalitarian instincts of legislatures to spread limited salary funds to more employees at the lower end of the salary schedule and reduce compensation for managers.

POWER OF ORGANIZATION

The power of organization—the power to create and abolish offices and to assign and reassign purposes, authorities, and duties to these offices—is critical to the governor's role as chief administrator, for it may be used to confer organizational status and give certain public purposes, programs, and constituencies higher priority and visibility than others. The legislature holds primary authority to create state offices and prescribe duties and responsibilities for such offices and, therefore, shares substantially in the power of organization, but gubernatorial powers of organization are on the increase. In 1955, only two states granted governors such powers, but as of 1994, 23 states granted governors authority to initiate executive reorganization subject to some form of legislative review.^[6] Moreover, within the bounds of state statutes and appropriations, most governors and their appointees have broad discretion in exercising the power of organization, for example, in establishing a cabinet and mechanisms to coordinate major state agencies, reorganizing administrative operations within agencies, creating gubernatorial task forces and study commissions, and assigning supplementary duties of gubernatorial priority to certain offices, among others. The gubernatorial power of organization may be

applied to achieve policy objectives or for purposes of management.

Gubernatorial action in the organizational arena commonly runs into direct and immediate conflict with the iron triangles of bureaucratic politics, and many reorganization proposals can be found in the political graveyards of state government. Most often, exercise of the power of organization will be vigorously opposed by elements in the legislature, by interest groups, from within the bureaucracy, or by a combination of these forces. Such actions will likely be denounced as a “power grab” or as attempts to import the ideas of out-of-state experts and associations. The politics of reorganization are well summarized by the late Robert Bennett, former governor of Kansas:

... during the election campaigns there wasn't a candidate of either party ... that didn't proclaim loud and clear that he was for reorganization, he was going to abolish unneeded activity, he was going to merge ... duplicating departments, he in fact was going to streamline government. In the abstract, it is without a doubt, one of the finest and one of the most palatable theories ever espoused by a modern day politician. But in practice ... it becomes the loss of a job for your brother or your sister, your uncle or your aunt. It becomes the closing of an office on which you have learned to depend for a small portion of your municipal economic sustenance. It becomes the doing away with an activity that is of personal economic benefit to you although it may be of little benefit to others. So there may in many instances be more agony than anything else in this reorganization process.^[10]

POWER OF COMMAND

The power of command—the power of a governor to order the actions of executive officials—is also shared with the legislature and, therefore, varies from state to state.^[11] Most state constitutions grant governors explicit or implied powers in the form of the “supreme executive power,” the power to “faithfully execute the laws,” and the power to “require information in writing from the officers of the executive department.” On the other hand, the authority to create state offices and prescribe the duties and responsibilities of state offices resides with the legislature. Precedents at state and national levels generally authorize the elected chief executive to direct the official actions of any agency head subject to the chief executive's power of appointment. In other words, the discretionary acts of an agency head are subject to command by the governor and enforced, if necessary, through the power of removal. Further, because most state statutes grant agency heads broad authority to

define state programs, specify program purposes, and implement those programs, gubernatorial command may encompass not only the individual acts of agency heads but also their discretionary authority to issue rules and regulations. Gubernatorial orders may take the form of verbal instructions, executive orders, proclamations, reorganization orders, memoranda, and various other policy directives. The legislature may seek to limit the scope of gubernatorial command by prescribing an agency head's duties with more specificity or by removing an agency head from appointment by the governor or through statutory mechanisms providing for legislative oversight of executive rule making.

Gubernatorial commands may generate political liabilities, for example, in issuing directives that narrow program purposes and exclude potential beneficiaries. Directing personnel on a day-to-day basis may be unproductive. According to former governor Pryor of Arkansas, a governor “will spend almost as much time keeping his staff and his cabinet and the people around him happy as he does keeping his constituency happy.”^[3] Governors are also thrust into resolving battles among and between staff and agency personnel that ultimately result in deflated egos. Directing an agency to do something it is reluctant to do may lead to bureaucratic foot dragging and attempts to emasculate the directive. Or, directing an agency to turn down a federal grant may generate interest group pressure to reverse the governor's decision. Another approach to the power of command is to delegate, but delegation too has political impact. A governor who delegates may be viewed as a weak manager by agency personnel and by the public but ultimately will be held politically accountable for problems that occur under delegation.

In sum, the role of governor as chief administrator is conducted in the political arena, and the exercise of administrative power requires an investment of political capital. Although administrative action frequently has more impact than legislative action, governors have come to recognize the political costs of investing time and resources in the administrative role. Says former governor Walker of Illinois: “Nobody cares ... whether you're actually managing state government ... It's not a political plus ... In terms of getting votes or a better image or a better reputation, you'd better spend your time somewhere else.”^[3] Another assessment similarly concludes:

... the substantive success of the Governor's efforts at management improvement could not be translated into votes, for there was little, if any, real political support for effective management. Unlike almost any other area of state policy, there was no major constituency to whom management improvement was a salient

issue. Indeed the situation was quite the contrary, as most organized interest groups pushed for objectives which militated against good management.^[10]

EMERGING TRENDS IN GUBERNATORIAL ADMINISTRATION

At the start of the twenty-first century, new directions in gubernatorial administration are emerging—shaped in part by progressive-era ideals evident from the start of the twentieth century and in part by political realities experienced throughout the past century. Governors today have wider latitude than ever before in determining the character of their governorship and specifically in performing the role of chief administrator of state government. Emerging trends suggest that governors may determine their governorship and administrative role around three distinct patterns, discussed below.^b

First, *a governorship should be balanced in its attention to policy and administration.* In the modern governorship, a governor should attend to two primary and roughly equivalent roles, which are as chief administrator and chief policy maker in state government. As chief administrator, a governor is held politically accountable for the administration of state government and advances those long-standing progressive-era ideals that give order and value to management in state government. Effective gubernatorial management should be conducted through sophisticated delegation to essential components of gubernatorial administration, specifically, gubernatorial staff, executive budget staff, and cabinet officers. A governor should reinforce the core values and distinctive contribution expected of these administrative components. Gubernatorial staff should assure focus on the governor's policy priorities. Executive budget staff should enforce the governor's fiscal policy priorities. Cabinet officers should manage major state functions and represent the administration to the primary constituencies of state government.

Second, *a governorship should focus on policy and political strategy.* In the modern governorship, a governor should focus on policy and political strategy essential to achieving gubernatorial policy priorities. The governor should practice “hands-off” management, and management of state government should be a secondary concern of the governorship. Gubernatorial appointees should be oriented foremost

to gubernatorial policy, and administrative actions should be assessed in terms of their contribution to the policy agenda and political resources of the governor. State bureaucracies are expected to implement public policies.

Third, *an entrepreneurial governorship should commit to reinventing state government through competition and individual choice.* In the modern governorship, a governor should exercise entrepreneurial leadership to reinvent government, take risks in stimulating the economy, and advance free market competition and individual choice. The governor's attention should be focused on steering the state; the rowing functions of state government should be delegated, decentralized, diffused, or privatized. Competition should be injected into the delivery of state services, and those delivering services should be held accountable to their customers. State bureaucracies should be required to compete for service delivery with private and non-profit organizations. Wherever possible, public-private partnerships should be encouraged, and authority for state services should be diffused and decentralized.

REFERENCES

1. Flentje, H.E. State administration in cultural context. In *Handbook of State Government Administration*; Gargan, J.J., Ed.; Marcel Dekker, Inc.: New York, 2000; 80–86.
2. Sabato, L. *Goodbye to Good-Time Charlie*; D.C. Heath & Co.: Lexington, MA, 1978; pp. 90, 107–118, 207–208.
3. Center for Policy Research, National Governors' Association. *Reflections on Being Governor*; National Governors' Association: Washington, DC, 1981; pp. 67, 104–105, 128, 169, 184, 231, 243–244, 266, 284.
4. Center for Policy Research, National Governors' Association. In *Governing the American States: A Handbook for New Governors*; National Governors' Association: Washington, DC, 1978; 204 pp.
5. Connery, R.H.; Benjamin, G. *Rockefeller of New York: Executive Power in the Statehouse*; Cornell University Press: Ithaca, NY, 1979; 153 pp.
6. Beyle, T.L. Enhancing executive leadership in the states. *State Local Gov. Rev.* **1995**, 27 (1), 19–21.
7. Weinberg, M.W. *Managing the State*; MIT Press: Cambridge, MA, 1977.
8. Rosenthal, A. *Governors and Legislatures: Contending Powers*; CQ Press: Washington, DC, 1990.
9. Behn, R.D., Ed.; *Governors on Governing*; National Governors' Association: Washington, DC, 1991.
10. Flentje, H.E., Ed.; *Selected Papers of Governor Robert F. Bennett: A Study in Good Government and Civics Book Politics*; Center for Urban Studies, Wichita State University: Wichita, KS, 1979; pp. 259, 264.

^bThese patterns are summarized from a variety of sources, primarily from Refs. [12–18].

11. Bernick, E.L.; Wiggins, C.W. The governor's executive order: an unknown power. *State Local Gov. Rev.* **1984**, *16* (1), 3–10.
12. Adamany, D. Successful hands-off management. *J. State Gov.* **1989**, *62* (4), 140–146.
13. Andrisani, P.J., Hakim, S., Leeds, E., Eds.; *Making Government Work: Lessons from America's Governors and Mayors*; Rowan and Littlefield: Sanham, MD, 2000.
14. Cox, R., III. The management role of the governor. In *Gubernatorial Leadership and State Policy*; Herzik, E.B., Brown, B.W., Eds.; Greenwood Press: New York, 1991; 55–71.
15. Flentje, H.E. Clarifying purpose and achieving balance in gubernatorial administration. *State Gov.* **1989**, *62* (4), 161–167.
16. Hebert, F.T. Governors as chief administrators and managers. In *Handbook of State Government Administration*; Gargan, J.J., Ed.; Marcel Dekker, Inc.: New York, 2000; 107–126.
17. Osborne, D. *Laboratories of Democracy*; Harvard Business School Press: Boston, 1990.
18. Osborne, D.; Gaebler, T. *Reinventing Government*; Addison-Wesley: Reading, MA, 1992.

Gramm–Rudman–Hollings: Impact on the Federal Budgetary Process

G.L.A. Harris

*Mark O. Hatfield School of Government, Portland State University,
Portland, Oregon, U.S.A.*

INTRODUCTION

Our nation's federal budget deficit looms large at almost \$8.5 trillion, enough to hold each and every American accountable for at least \$27,494 of that debt.^[1] And, given the size of today's federal deficit, in 1985 and 1987 it was this very premise under which the Gramm–Rudman–Hollings (GRH) Act was enacted in a desperate attempt to keep the federal government's deficit, and in turn the national debt, from spiraling out of control.

Today's deficit woes would suggest, that although admirable in principle, the goals of GRH have not been realized. But, according to former Senators Phil Gramm (R-Texas) and Warren Rudman (R-New Hampshire), two of the three sponsors of the Act and for whom the Act was partly named, "Those who call the 1985 Gramm–Rudman–Hollings balanced budget law a bad law because deficit targets have been missed and accounting gimmicks have been used to avoid tough decisions would probably also suggest that religion is a failure because there are a few saints in the world."^[2]

At a time of partisan rancor and an ever increasing federal budget deficit, GRH appeared to represent the best of few alternatives to get the nation's fiscal house in order. The purpose of this article is to describe what prompted the passing of the GRH Act; why the courts got involved on constitutional grounds; the pros and cons of enacting this legislation; and given the significance of the Act, the numerous attempts to revisit the seemingly intractable problem of the federal budget deficit.

TOWARD BUDGET ENFORCEMENT UNDER GRH I

The Balanced Budget and Emergency Deficit Control Act of 1985, known as the GRH I Act (P.L. 99-177), was passed on December 11, 1985 largely due to a political impasse that ensued between President Reagan and the Congress over how to best attack the troubling federal budget deficit.^[3] Accordingly, the Act signaled that 1) the Congressional Budget and Impoundment Act of 1974 failed to address an

increasing federal deficit,^[4] 2) the impasse between the President and the Congress demanded a "statutory framework" to compel resolution (on page 267 of Ref.^[4]), and 3) the intent by Congress was to seize control of budgetary authority from the executive branch.

Although like his predecessor Jimmy Carter, President Reagan's pledge was to achieve a balanced budget, his strategy for doing so was different.^[5] Reagan intended on achieving a balanced budget in accordance with his own commitment to three policy objectives; first, to continue to accelerate the defense buildup; second, to reduce taxes; and third, to force a retrenchment of domestic programs.^[6] Against a Republican Senate and a Democratically controlled House of Representatives that were both concerned about the deficit but diverged markedly on how to reduce it, a compromise in the form of GRH was struck as no lawmaker dared to be seen as being against a balanced budget.

GRH changed the federal budgetary process by mandating the following. First, the Act established specific targets for deficit reduction in \$36 billion increments per year until the budget was balanced or at zero deficit by fiscal year 1991.^[7-9] Second, the Act introduced across the board cuts known as sequestration that would occur in certain categories of outlays (spending). Federal programs such as Social Security, for instance, were rendered exempt from sequestration and half of such cuts would come from certain entitlement programs and some federal retirement programs while the remaining half of the cuts would affect discretionary programs such as defense procurement.^[10] The intent of GRH was that the very threat of sequestration, due to failure to meet targeted budget deficit levels, would force the President and Congress to strike an agreement.^[7]

Finally, specific procedures established spending and revenue limits at the beginning of the budgetary process.^[11] Such constraints were designed to invoke discipline in the budgetary process.^[7] This would in turn neutralize political conflicts, increase the likelihood of adherence to the budgetary process, and expand accountability requiring sequestration reports.^[7,9] Further, GRH provided an incentive for Congress to detail appropriations thus limiting the

discretion of the executive branch over appropriation allocations.^[12] GRH amended its precursor, the Congressional Budget and Impoundment Act of 1974 and instituted procedures that made it difficult for the President and Congress to pass annual budgets that would increase the deficit.

THE ROLE OF THE GENERAL ACCOUNTING OFFICE (GAO) UNDER GRH I

Agreement to GRH was often a very contentious process but clearly marked a period that propelled lawmakers to action given a rising deficit by the 1980s (at 3% of GNP with outlays at 20% of the GNP), an approaching election, vocal concerns by constituents, and the need to break the impasse between the President and the Congress.^[6] The sentiments of Senator Warren Rudman (R-NH), a cosponsor of the Act, reflected the mood of the period when he said that GRH is “a bad idea whose time has come” [on page 237 of Ref.^[5]; on page 85 of Ref.^[10] (as cited by Kowalcky and LeLoup^[13] on page 24)].

But, despite the agreement, GRH signaled the mistrust that existed between the various factions. To achieve neutrality, particularly given the threat of sequestration while ensuring that the threat was not so overwhelming as to undermine the process, GRH called for protection of the deficit but within the President’s budget through the Office of Management and Budget (OMB) and the Congressional Budget Office (CBO).^[6,12] Further, Congress trusted the CBO for its objectivity and reputation as a nonpartisan entity.^[6] However, questions surfaced as to the constitutionality of the CBO giving the President a directive to make certain cuts in the federal budget. To overcome this hurdle, it was thought that the General Accounting Office (GAO), under the Comptroller General, would not only provide political cover and curtail any test of the constitutionality of the Comptroller General’s role in the process but would also give credibility and impartiality to the process before sequestration could occur. GRH mandated that, should the President and Congress fail to reduce the deficit by targeted levels or “maximum deficit amounts,” the process of sequestration would be automatically triggered (on page 10 of Ref.^[6]).

The OMB and CBO would thus base their projections of the federal budget and deficit on reasonable assumptions given the direction of the economy, the impact of these conditions on the federal budget, and the rate of outlays from certain programs in accordance with the policies and/or rules applied to those programs and funding levels. Should the estimate by OMB and CBO result in a disagreement, the accepted estimate would be the average estimate of the two

agencies. The GAO, and thus the Comptroller General’s role, was simply to provide an additional safeguard for the process that triggers sequestration. A joint OMB and CBO report is submitted to the GAO for review by the Comptroller General. The Comptroller General then issues a report that is independent of both the OMB and the CBO reports.

It is thus the estimates contained in the GAO report, once submitted to the President, that triggers sequestration. While it is the Comptroller General’s report that is, in essence, the last say before sequestration, the report must fully explain any discrepancies that exist between his own report and that of the joint OMB and CBO report. GRH intentionally detailed this budget information in terms of where budget cuts are to be made. This, in turn, reduced the discretion of the President to question and/or reallocate cuts.^[6] Additionally, as a fallback, should any of the procedures be rendered unconstitutional, the joint report by the OMB and CBO was to be submitted to the Temporary Joint Committee on Deficit Reduction, not the Comptroller General.^[10] The Temporary Joint Committee on Deficit Reduction was a body comprised of the budget committees from both houses of Congress that would issue a joint resolution. Within 5 days of the receipt of the joint report to both houses, Congress would be required to “certify” sequestration under special rules to vote on the resolution (on page 96 of Ref.^[10]). If the resolution is passed and then signed by the President, this formed the basis for sequestration.

WHY DID THE COURTS GET INVOLVED?

But GRH proved to be one of the most controversial and often misunderstood pieces of legislation. Despite the sponsors’ attempt to allay fears based on constitutional grounds by inserting the GAO and thus the Comptroller General’s review to signify the objectivity of the process and one that would appeal to all parties involved, the Act itself triggered legal action on the very constitutional role of the Comptroller General.

Barely hours following the signing of the legislation, two lawsuits, combined as *Bowsher vs. Synar*, alleging that GRH should be declared unconstitutional,^[14] was argued and heard before a special three judge district court panel^[10] for the following reasons. First, GRH violated delegation of authority where power is vested in the Congress not to delegate this authority.^[6,10] Second, if delegation of authority is constitutional, that delegation to the Comptroller General violated the separation of powers doctrine. And, third, that GRH prevented unionized employees from receiving timely cost of living increases.^[14]

The District Court ruled that the plaintiff had standing to challenge certain provisions of GRH and that broad deficit reduction power to the Comptroller General was unconstitutional. The *Bowsher vs. Synar* decision was automatically appealed under an expedited review to the Supreme Court.^[10] In a 7–2 decision, the Court concluded that sequestration under GRH was unconstitutional and, in writing for the majority on the separation of powers doctrine, Chief Justice Burger wrote:

Congress cannot reserve for itself the power of removal of an officer charged with the execution of the laws except by impeachment. To permit the execution of laws to be vested in an officer answerable only to Congress would in practical terms reserve in Congress control over the execution of laws . . . The structure of the Constitution does not permit Congress to execute the laws; it follows that Congress cannot grant to an officer under its control what it does not possess.

—as cited by *LeLoup et al.*^[10] on page 97 in
Bowsher v. Synar, 1986.

REVISITING FISCAL RESPONSIBILITY UNDER GRH II

Since sequestration under the separation of powers doctrine was invalidated by the Supreme Court, the sponsors of GRH quickly exercised the Act's "fallback procedure" of using the combined OMB and CBO report instead of the GAO report submitted to Congress (on page 22 of Ref.^[6]). Here, the Temporary Joint Committee on Deficit Reduction considered the report. Under joint resolution and expedited procedures, and based upon the contents of the report, the joint resolution goes to the floor of the Congress for a vote. If the contents of the report are agreed to and signed by Congress and the President, the President is obligated to the reductions in the federal budget as per the report. The sponsors of GRH still ensured that the GAO played a role in determining the deficit estimates.^[10] However, it was OMB in conjunction with the CBO estimates that would issue the final report to the President not the GAO as the original version of the Act called for.^[10]

This fallback version of the original GRH legislation, which also encountered stiff opposition in the House and by the President, in a 63–36 vote by the Senate, was enacted and came to be known as the Balanced Budget and Emergency Deficit Control Reaffirmation Act of 1987 (P.L. 100-119) or GRH II.^[10] While this version of the law attempted to address the constitutional flaws of GRH I, it nevertheless bogged down the process because neither the House nor the Senate leaders could agree on how to resolve the constitutional flaws. Further, members

were preparing for the August 15 recess and Congress opted to a temporary extension without deciding on the 1987 amendment.

However, the OMB and CBO estimates predicted what many lawmakers had feared; that the federal deficits for both fiscal years 1986 and 1987 were larger than expected. The CBO and OMB estimates for fiscal years 1986 and 1987 were \$170.6 billion and \$156.2 billion, respectively. However, without the GAO estimate to reconcile this amount, the average deficit was taken at \$163.4 billion. Had the original procedures under GRH I taken place, the deficit requirement before sequestration would have been \$144 billion.

Under GRH II, while lawmakers avoided the across the board cuts, deadlines for enacting the fiscal year 1987 budget were missed. Many referred to the myriad of tactics employed by lawmakers as "golden gimmicks" and "smoke and mirrors" (on page 99 of Ref.^[10]). Congress agreed to a deficit of \$151 billion, below that of the target under GRH II. Further, attempts by the sponsors of GRH to prevent the government from using its borrowing authority unless the procedures under GRH II were complied with, were unsuccessful. Should sequestration be activated, half of the cuts would have come from defense, while the other half would have affected domestic spending. Entitlement programs such as Social Security and Medicaid, for example, were exempt. The Act was signed into law on September 29, 1987.

REAL OR PERCEIVED: THE IMPACT OF GRH ON THE FEDERAL DEFICIT

The Gramm–Rudman–Hollings Acts of 1985 and 1987 (GRH I and II) represented the most far reaching legislative attempt to overhaul the legal and administrative framework of the federal budget since the 1921 Budget and Accounting Act. However, GRH proved to be controversial in many ways, most notably, challenging the separation of powers doctrine, where the role of the Comptroller General in determining the point at which automatic across the board budget cuts or sequestration, would take place. But assessing GRH's impact on the federal budget deficit really depended on whom and to which side of the aisle you belong.

Proponents of GRH believed that, given the sign of the times, the enactment of GRH, or at least legislation similar to it, was inevitable. First, GRH attempted to take the politics out of the budgetary process by levying legal and institutional constraints against increasing the federal deficit.^[12] The Act's goal was to achieve neutrality in the budgetary process. Further, the political climate was such that many believed GRH would provide the political will through sequestration to hold down federal spending.^[7] In fact, some

also believed that GRH would provide the political cover to reduce the deficit by exploring other unpopular alternatives such as increasing taxes and decreasing entitlement programs. And, the threat of sequestration, it was hoped by the sponsors, would serve to force lawmakers to reduce the federal deficit.

Second, GRH may have provided the mechanism to put the brakes on government spending.^[7] In essence, without GRH II, the federal deficit might have been much larger. Some programs, including those for defense, stopped increasing after the 1980s, while in the 5 fiscal years following the passage of GRH II, some programs were actually reduced. Even some domestic programs, such as Medicare, were restrained under GRH II. Accordingly, “the limits imposed by GRH II may have stiffened the fiscal backbone of the political system” (on page 228 of Ref.^[7]).

Third, GRH has had far reaching effects beyond that of the size of the federal deficit and how such reductions would be achieved.^[7] GRH II has impacted how political power is balanced and how legislative priorities are determined. It has resulted in a change in the positions of some programs and committees. Some tax writing committees, for instance, have gained and lost importance. Tax committees have gained by having both spending programs and revenues under their purview. These tax committees, under the Ways and Means and Finance Committees, may have provisions that raise revenues while overseeing entitlement programs that reduce spending. But these committees have also lost under GRH II (i.e., drug initiatives or disaster relief) because decisions on certain tax legislations were out of their hands.

Fourth, GRH II has helped to reposition the priorities of some programs in securing increasing resources for some (i.e., Medicaid) while helping to scale back already weak programs that leave them more vulnerable to budget reductions (i.e., telephone tax extensions).^[7] Fifth, GRH II has changed the legislative process for the budget by requiring that the deficit cannot exceed a prescribed target.^[12] Congress and the President must work together to achieve agreed upon levels of expenditures and revenues toward deficit reduction to avoid sequestration.

Next, GRH II encouraged uniformity in the budget process through the use of coherent and consistent methods and baseline estimates toward each proposed budget.^[12] This was an attempt to eliminate “manipulation of budget authority/outlay ratios” although it was conceivable that program outlays would be increased as a cushion should sequestration take place (on page 638 of Ref.^[12]). This uniformity also applied in the sense that, should sequestration occur, every budget account would be reduced in a uniform manner. Under sequestration, GRH II actually gave the President latitude on where the cuts in defense

should occur as long as he notified the Congress of those changes.^[15]

GRH also built the estimation of technical and economical factors into the budgetary process to ward off unrealistic assumptions.^[15] The definition of baseline estimates was expanded on where budget cuts would occur by allowing consideration of inflation and increased program participation. GRH II strengthened the enforcement procedures of the budget by prohibiting, for instance, the Senate’s consideration of irrelevant information and those items that may serve to increase outlays in the future. Such prohibitions made it more difficult for waivers to be granted by requiring a 60 vote majority. Finally, GRH II called for further reforms of the budgetary process by inserting a pilot program of certain federal agencies of their biennial appropriations.^[15]

Despite the laudatory reviews about GRH, there is no shortage of reasons as to why the legislation was considered a failure by many.

First, under GRH II, the fact that the deficit targets and timetables by which these targets were to be achieved were relaxed, point to what opponents refer to as “gimmicks in the law” (on page 8 of Ref.^[16]). Many such gimmicks, some believed, came in the forms of overly optimistic economic predictions and manipulating pay dates. Some critics went as far as dubbing GRH II “an ineffective and dangerous piece of legislation” (on page 191 of Ref.^[3]).

GRH II established that the President, in concert with Congress, estimates “in his judgment” what the outlays and appropriations for the budget should be to run the federal government (on page 694 of Ref.^[4]). This signaled personal responsibility of the President for the nation’s fiscal soundness. By dictating prescribed budget targets in anticipation of the deficit, this undermined the President as well as his responsibility for the budget. Congress cannot dictate to the President.^[4] In essence, “An executive budget expresses what the President wants” (on page 698 of Ref.^[4]). The President’s budget is only a proposal, therefore, Congress has no right to tell him what to do. And as such, GRH II makes it easy for the President to shirk his responsibility for the federal budget.

Third, GRH II did not control the federal deficit; it created the illusion that it would.^[3] The first term of the Reagan administration marked a period of increasing defense spending and tax cuts. Although during this period of GRH II the deficit did not grow as much, it was still clear that the deficit had not been controlled.

Fourth, the short-term orientation of GRH II made it inappropriate for long-term budget problems.^[4] This 1-year focus promoted gimmicks such as shifting the costs of some programs from one year to another. Doing so forced appropriation bill projections to be unnecessarily delayed for the most recent budget

projections. Such delays made it more likely that, when critical decisions were to be made, the necessary give and take process would be averted because the members of Congress for such debates would be absent as such decisions would likely result in “budget summits” (on page 698 of Ref.^[4]).

Many also believe that budget delays and thus budget projections were held over to deliberately conceal the size of the budget deficit.^[4] And, the budget, and in turn the deficit, were even more skewed because GRH II emphasized expenditures (outlays), not revenues. Accordingly, this is tantamount to “deceit” as a way of concealing the actual size of the deficit (on page 698 of Ref.^[4]).

Sixth, some saw GRH II as a “crutch” that both the President and the Congress used to avoid making difficult decisions about problems in the federal budget (on page 698 of Ref.^[4]). Complying with deficit targets did not address the situation at hand as this enabled the President and Congress from confronting the real federal budget problems. Further, there was less public pressure on the President and Congress to reduce the federal deficit in the belief that GRH II would take care of the problem.

Some contend that GRH II was poorly designed.^[3] While the threat of sequestration was forced to achieve a compromise, many members of Congress found the President’s proposal for budget cuts to be more palatable since 70% of budget outlays (entitlement programs) were exempt from sequestration. The remaining 30% of the budget, or those programs subject to sequestration (discretionary programs), did not induce the same political fallout. As a result, fiscal years 1986 through 1989 saw the deficit targets and timetables relaxed thus avoiding sequestration and realistic deficit control.

GRH II prevented the normal consensus building between political parties and deliberations on tax increases and outlay cuts to take place.^[3] It favored sequestration of outlays while failing to meet the revenue side of the budget. GRH reduced the incentives for the parties to compromise while encouraging illusory budgetary tactics to circumvent sequestration. GRH II, according to Alice Rivlin, “risks destabilizing the economy” (on page 196 of Ref.^[3]). Rivlin believed that constraints imposed by GRH II in setting specific budgetary targets may have served to weaken an economy that should be stimulated.

Uniformity of cuts under sequestration failed to consider those programs that should be expanded and those that should be reduced.^[3] Sequestration appeared to consistently impact the same 30% of programs while insulating the same remaining 70%. And, the basis for such cuts often reflected the politically charged interest of various groups.

Under GRH II, there was an overreliance on deficit estimates by the OMB.^[3] Since the OMB represented

the President, it tended to insert the partisan interests of the executive branch and unduly influenced the legislative process.

Finally, GRH II encouraged overestimation of the budget revenue and thus an underestimation of the actual size of the federal deficit.^[3] This resulted in gaming the system to avoid sequestration, not control the budget by reducing less than is needed to curb the deficit while forcing bigger cuts in the federal budget in future years.

CONCLUSION

The events leading up to the 1985 and 1987 GRH Acts were contentious. It signaled a period ripe for change. Known in its original form as the Balanced Budget and Emergency Deficit Control (GRH I) Act (1985) and its revised version as the Budget and Emergency Deficit Control Reaffirmation (GRH II) Act (1987), GRH represented the best of few alternatives given an exploding federal budget deficit. Besides, at the time, voting against a balanced budget was not what lawmakers wanted as their legacy. It was an impasse that resulted in this key legislation.

GRH overhauled the federal budgetary process by establishing three primary goals. First, the Act created specific targets and timetables by which the federal deficit should be annually reduced. Second, that failure to achieve prescribed deficit levels within required timeframes would automatically result in sequestration. This feature was inherently designed to force the President and Congress to forge an agreement to reduce the deficit. Finally, the legislative and administrative framework established by the Act demanded a disciplined approach to the federal budgetary process. This, the sponsors hoped, would help to neutralize partisan politics.

Reaction to and the impact of the GRH legislation have been mixed. Failure of GRH to bring about a fundamental change in the federal budgetary process and reduce the deficit is not surprising to those critics who see such reforms as mere regurgitations of previous ones. According to Joyce,^[17] budget reforms, as history has proven, is likely to produce much less than is promised or more of the same because there are those who keep looking for a “silver bullet” to this largely intractable problem (on page 36 of Ref.^[17]). Yet, there are others like Havens^[6] who believe that reducing the federal deficit becomes a “political imperative” as long as the solutions are broached in a “balanced” and “fair” manner (on page 23 of Ref.^[6]). Perhaps it was this political imperative that moved the Clinton Administration to not only balance the federal budget but also leave the succeeding Bush Administration with a record \$500 billion surplus. It is conceivable

then that GRH, for all of its shortcomings, might have provided the traction to impose fiscal discipline after all. But, as Senators Gramm and Rudman stated, “Just as we don’t judge the success of religion by the number of saints but by whether or not the world is better off, so should we judge Gramm–Rudman–Hollings”.^[2]

We now then face tension from the pessimism stemming from our past failures at federal budget reform vs. the imperative for reform to achieve fiscal soundness. On one hand, the abject failure of GRH shows the intractability of the problem. A strong Act, created cooperatively with the President and bipartisan Congressional support, could not hold back the political forces leading to fiscal disaster. Yet on the other hand, the growing voices of discontent like the current movement led by former Comptroller General, David Walker, warning of the impending fiscal ruin, point to the importance of trying again for fiscal reform to attain fiscal soundness.

REFERENCES

1. Bureau of the Public Debt: The Debt to the Penny. <http://www.publicdebt.treas.gov/opd/opdpenny.htm>.
2. Gramm, P.; Rudman, W. Deficit reduction: A love story; We made a commitment. The New York Times **1989**, Section A, p. 31, Column 4, October 25.
3. Thelwell, R. Gramm–Rudman–Hollings four years later: A dangerous illusion. Public Admin. Rev. **1990**, *50* (2), 190–198.
4. Fisher, L. Federal budget doldrums: The vacuum in presidential leadership. Public Admin. Rev. **1990**, *50* (6), 693–700.
5. Ippolito, D.S. *Why Budgets Matter: Budget Policy and American Politics*; The Pennsylvania State University Press: University Park, PA, 2003.
6. Havens, H.S. Gramm–Rudman–Hollings: Origins and implementation. Public Budg. Finance **1986**, *6* (3), 4–24.
7. Reischauer, R.D. Taxes and spending under Gramm–Rudman–Hollings. Nat. Tax J. **2001**, 4–24.
8. Ott, A.F. *Public Sector Budgets*; Edward Elgar Publishing Limited: U.K., 1993.
9. Joyce, P.G. Congressional budget reform: The unanticipated implications for federal policy making. Public Admin. Rev. **1996**, *56* (4), 317–325.
10. LeLoup, L.T.; Luck, G.B.; Barwick, S. Deficit politics and constitutional government: The impact of Gramm–Rudman–Hollings. Public Budg. Finance **1987**, *7* (1), 83–103.
11. Rubin, I. *The Politics of Public Budgeting*, 4th Ed.; Chatham House Publishers, 2000.
12. Stith, K. Rewriting the fiscal constitution: The case of Gramm–Rudman–Hollings. Calif. Law Rev. **1988**, *76*, 595–648.
13. Kowalcky, L.K.; LeLoup, L.T. Congress and the politics of statutory debt limitation. Public Admin. Rev. **1993**, *53* (1), 14–27.
14. Elliott, E.D. Regulating the deficit after *Bowsher v. Synar*. Yale J. Regul. **1987**, *4*, 317–362.
15. LeLoup, L.T.; Hancock, J. Congress and the Reagan budgets: An assessment. Public Budg. Finance. **1988**, *8* (3), 30–54.
16. Hahm, S.Deuk.; Kamlet, M.S.; Su, T.-Tsu. The influence of the Gramm–Rudman–Hollings Act on federal budgetary outcomes, 1986–1989. J. Policy Anal. Manage. **1992**, *11* (2), 207–234.
17. Joyce, P.G. The reiterative nature of budget reform: Is there anything new in federal budgeting? Public Budg. Finance **1993**, *13* (3), 36–48.

Grievance Procedures and Administration

Lisa B. Bingham

Tina Nabatchi

*School of Public and Environmental Affairs, Indiana University,
Bloomington, Indiana, U.S.A.*

INTRODUCTION

This chapter describes the basics of the grievance procedure. It provides a discussion of several theoretical perspectives used in grievance procedure research, and reviews literature on grievance initiation, processing, decisions, and outcomes. Then, a brief discussion about grievances and organizational performance is provided. Finally, the grievance mediation, a recent innovation in the grievance procedure is examined.

THE GRIEVANCE PROCEDURE

The grievance procedure has received significant attention from scholars and practitioners as a critical component of the labor–management relationship. There are substantial variations in the scope of the grievance procedure from contract to contract and across the public and private sectors. The definition of a grievance is subject to the terms of the collective bargaining agreement or contract. In its broadest conception, a grievance is any employee complaint about an employment relationship. In a grievance procedure, a grievance is a *written* allegation by employees that management has in some way violated their contractual rights. Grievances generally allege some breach, violation, misinterpretation, or misapplication of language in the collective bargaining agreement, often with regard to wages and hours, discharge and discipline, safety and health, insurance benefits, seniority, leaves of absence, promotions, vacations, management rights, and union rights.^[1–3] Although there is variation across the private and public sectors, the basic issues brought to arbitration in the two sectors are similar. The primary difference between the public and private sectors appears to be with the prevalence of certain categories of grievances. For example, termination, suspension, or disciplinary actions resulting from drug and alcohol use, insubordination, absenteeism and tardiness, and abusive behavior are the highest incidence of cases brought to arbitration in the public sector.^[3]

Although there are significant variants, grievance procedures generally move through a series of steps before increasingly higher management authorities, beginning with informal discussion and the writing of a complaint, and ending with arbitration. The overwhelming majority of grievance procedures culminate in binding arbitration, in which an outside neutral third party issues a final decision on the grievance. The common step structure of the grievance procedure is described below.

Informal Step

The first step of any grievance requires the expression of a complaint. The employee, sometimes with the assistance of the union steward, discusses the complaint with the immediate, first-line supervisor. Most contracts encourage this informal step, because it is generally preferable to resolve all disputes as early or as quickly as possible, because parties' positions tend to harden once a grievance becomes formal and written. In this step, many disputing parties settle grievances during informal discussions between the employee and supervisor, and these never enter the grievance process as a formal, written complaint. Unfortunately, these informal grievances are generally not captured in traditional management information systems; therefore, researchers are limited in discovering the true nature and extent of informal grievances in the workplace.

Step One

If the employee and supervisor cannot resolve the dispute informally, then the grievance is submitted in writing, formally becomes a grievance, and technically enters the grievance procedure. Usually, there are time limits within which employees must act. Once the written grievance is filed with the first-line supervisor, a formal written decision must be issued, again within negotiated time limits. If the employee or union is dissatisfied with the step one response, they may appeal the grievance to step two.

Step Two

At step two, the grievance is generally presented to officials higher in the organization. In the public sector, it might move from the immediate supervisor to a department head, or from a school principal to the superintendent of schools. In the private sector, it might move from the immediate supervisor to a plant manager, or to some other designated representatives who are familiar with the contract and with grievance precedent. The officials hear the complaint and render a written decision. The employee or union can still appeal the decision; however, they must usually file the appeal within a stricter and shorter time frame.

Step Three

If a complaint reaches step three of the grievance process, then it likely has precedent-setting ramifications, major cost implications, or broad application within the firm's operation.^[4] As such, participants with more authority are brought into the discussion. In the public sector, the grievance might reach a board of police commissioners, school board, city council, personnel board, or other elected multimember public agency. In the private sector, it would reach upper management. Again, the responding official renders a written decision within a limited time frame. This is often the step at which parties call in legal counsel or regional union representatives. The union usually has the sole prerogative to determine which grievances will move beyond this step to arbitration and has an internal union grievance committee for making these decisions.

Step Four

The final step in the grievance procedure is usually arbitration, in which a neutral third party hears both sides of the case and renders a written decision to the appropriate management and union officials. This written decision is usually final and binding, although traditionally, the arbitrator is not bound by precedent. Witness testimonies and documents can be presented as arbitration evidence, and the party initiating the grievance (usually the employee or union) typically has the burden of proof.

THEORETICAL PERSPECTIVES ON GRIEVANCE PROCEDURES

There are several theoretical perspectives used in the scholarly study of grievance procedures. Researchers using *systems theory* examine the connections between several variables and the different stages of grievance

procedures, emphasizing a correlation between labor relations and the filing, processing, and settlement and postsettlement outcomes of grievances.^[5–10] *Human resource management* theorists suggest that human resource practices that encourage worker participation in work teams, problem solving, and other high-involvement activities can improve grievance procedures and have a positive affect on organizational performance. *Procedural-Distributive Justice* theorists, applying a research framework first used in courtroom settings, argue that organizational decisions will be more readily accepted if the processes by which they are achieved are perceived to be fair.^[11,12] Research supports the suggestion that perceptions of the fairness of grievance procedures significantly influence employee perceptions and attitudes toward those grievance procedures.^[13,14]

Research based on Hirschman's^[15] *exit-voice-loyalty theory* suggests that the grievance procedure is a "voice" mechanism that allows employees to air complaints and issues to management. The theory holds that when provided a "voice" option, employees will be less likely to "exit," or quit the organization. Accordingly, some theorists suggest that employees who file grievances will be less likely to leave their employers than those employees who do not file grievances.^[16–18] However, empirical evidence for this theory is mixed. In contrast, other scholars using *organizational punishment-industrial discipline theory* suggest that those involved in grievance activities are significantly more likely to be the subjects of negative post-grievance outcomes, such as retaliation. Empirical support for these propositions is garnered by the fact that "fear of reprisal" is the most frequently cited reason for not using available grievance procedures.^[19,20]

DETERMINANTS OF GRIEVANCE INITIATION

There has been extensive research on the determinants of grievance filing. Some researchers have found significant *demographic* differences between grievance filers and non-filers. For example, researchers have found that grievants tend to be younger and less educated and that minorities tend to file more grievances than Caucasians.^[5,21] Similarly, researchers have found that grievance filers are typically younger, with more skilled jobs, higher rates of absenteeism, dispensary visits, and insurance claim filings than non-grievants.^[22] Finally, others have found that grievance filers have significantly lower job satisfaction, more negative attitudes toward their supervisors, greater perceived pay inequity, stronger preferences for worker participation in decision making, and lower satisfaction with their unions than non-grievants.^[23]

Some researchers have examined the relationship between *management factors* and grievance initiation. Research demonstrates that managerial strictness with regard to performance and disciplinary actions is significantly related to grievance filing rates,^[24] as is managerial monitoring of employees.^[25] *Union factors* also appear to have an impact on grievance filings. Research shows that the grievance behavior of union stewards is related to company commitment, union commitment, and job satisfaction,^[26] and the activity and involvement of union officials in grievance procedures can affect grievance filing rates.^[5,24,27] Just as union officials may have an impact on the filing of grievances, so too do grievance filings affect union officials. For example, more favorable grievance outcomes are associated with larger, more supportive constituencies for union stewards,^[28] and grievance handling has a significant effect on the probability of being reelected as a union steward, as well as on the percentage of votes obtained in that reelection.^[29] Finally, the degree of perceived conflict in the *collective bargaining relationship* is significantly related to grievance filing rates.^[28,30] If union-management relations are hostile or confrontational, then grievance procedures may be used more frequently, possibly to increase pressure on management during times of contract negotiation; conversely, a cooperative union-management relationship may have a constructive and productive impact on grievance activity.^[31,32] In short, evidence indicates that grievance filing rates are likely to decline given cooperative union-management relationships.

GRIEVANCE PROCESSING, SETTLEMENT, DECISIONS, AND OUTCOMES

Research regarding grievance processing addresses the efficiency of the grievance procedure in resolving disputes quickly and with finality. As such, an important measure of grievance procedure effectiveness has been the rate of grievance resolution in the steps before arbitration or the intervention of some other third-party neutral. Early settlement of grievances is preferable, because grievances tend to become more difficult to settle once they are written and proceed through higher levels of the grievance process. Factors such as early authorization from upper management and union officials to resolve a grievance, and the written presentation of a grievance as opposed to simply an informal discussion, have been correlated with the earlier settlement of grievances.^[33]

Typically, an arbitrated case can have one of three settlements: 1) the grieving party can win the case; 2) the grieving party can lose the case; or 3) the arbitration can result in a compromise.^[3,26,34] In terms of

the settlements decisions, research demonstrates that the longer a grievance is under consideration at any level of the process, the more likely management is to respond favorably.^[29] Moreover, some assert that employees can typically expect a favorable grievance settlement, because parties involved in a workplace dispute must usually continue their relationship after the dispute is resolved.^[32,35] Research comparing public- and private-sector grievance outcomes confirms the following: 1) public-sector employees tend to “win” grievance cases more frequently than their private-sector counterparts; 2) there tend to be fewer discharge cases in the private sector than in the public sector; 3) more suspension and reprimand cases reach arbitration in the private sector than in the public sector; and 4) of grievance cases that resulted in suspension, fewer public-sector cases resulted in long suspensions.^[36]

Even though employees can usually expect a favorable settlement, important research using the organizational punishment-industrial discipline theory of grievances suggests that grievants can also face significant unfavorable postsettlement outcomes. For example, in the postgrievance settlement period, grievants tend to have lower performance ratings, promotion rates, and work attendance rates, as well as higher turnover rates.^[5] In addition, research has also shown that grievance filers *and* the supervisors against whom the grievances were filed, have significantly higher turnover rates and significantly lower job performance ratings and promotion rates in the immediate postgrievance settlement period than employees who did not file grievances and their supervisors.^[21] In short, evidence suggests that employees and supervisors involved in grievance activities face the potential of more workplace retribution than employees and supervisors who are not involved in grievance activity.

GRIEVANCE PROCEDURES AND ORGANIZATIONAL PERFORMANCE

Historically, one impetus for the development of grievance procedures was to help reduce strikes and work stoppages; yet, there are few studies that systematically compare the strike activities of employees with grievance procedures to those without grievance procedures. A study of the coal-mining industry, an industry with historically high strike rates, suggests that strikes were more frequent in those mines where supervisors were perceived as being unable to effectively handle grievances or where disputes could not be resolved locally.^[37]

More commonly, research in this area links grievance filing rates to various measures of organizational performance. There is evidence that access to grievance

procedures contributes to lower quit rates, which likely helps lower turnover and retraining costs,^[17,38] and that employee turnover rates are lower among unionized firms with access to grievance procedures.^[39] Other research has shown that high levels of grievance activity are associated with reduced levels of productivity.^[39-43] One possible explanation for these findings is that an effective grievance system can help organizations detect current or potential problems and rectify issues and undesirable situations before they develop into larger, more systemic problems.

INNOVATIONS IN GRIEVANCE PROCEDURES: GRIEVANCE MEDIATION

While employment dispute resolution and arbitration are addressed in other chapters in more detail, there are also innovations under way in the collectively bargained grievance procedure. The most notable innovation is the use of grievance mediation. Grievance mediation is an alternative to arbitration and is typically included as an additional step prior to arbitration. In grievance mediation, a third-party neutral assists the union and management parties in negotiating a mutually agreeable and voluntary settlement to the grievance.^[44,45] Whereas arbitration is generally seen by unions and employers to be expensive, time-consuming, and overly legalistic,^[46] mediation has been shown to have greater cost-effectiveness and result in swifter settlements, while simultaneously ensuring the satisfaction of participants.^[32,47,48] Moreover, whereas arbitration is intrinsically adversarial, mediation has more of a cooperative nature that allows greater opportunity for problem solving and collaborative initiatives. Research has generally shown grievance mediation to be an effective alternative to arbitration, resolving more than 80% of mediated grievances, and an efficient alternative to arbitration, resolving grievances more quickly and cheaply.^[32,35,49]

CONCLUSION

This chapter provided an overview of the collectively bargained grievance procedure, its structure, a brief review of theories about its function, major themes in the empirical research about that function, and an introduction to grievance mediation as a recent innovation. The grievance procedure is an essential component of the collective bargaining relationship, one that makes the benefits of industrial democracy accessible to the individual employee. It provides employees with a meaningful avenue to exercise voice with regard to workplace concerns. Researchers have learned much about what factors contribute to grievance initiation,

how procedures can process grievances most efficiently and effectively, and what grievance procedures produce in terms of decisions and outcomes. However, there is still need for considerably more research, particularly on the systemic impact of grievance procedures on organizational performance.

REFERENCES

1. BNA Editorial Staff. *Grievance Guide*, 10th Ed.; The Bureau of National Affairs, Inc.: Washington, DC, 2000.
2. Richardson, R.C. *Collective Bargaining by Objectives: A Positive Approach*; Prentice-Hall, Inc.: Englewood Cliffs, NJ, 1985.
3. Mesch, D.J.; Shamayeva, O. Arbitration in practice: a profile of public sector arbitration cases. *Public Pers. Manage.* **1996**, 25, 119-132.
4. Duane, M.J. *The Grievance Process in Labor-Management Cooperation*; Quorum Books: Westport, CT, 1993.
5. Lewin, D.; Peterson, R.B. *The Modern Grievance Procedure in the United States*; Greenwood Press: Westport, CT, 1988.
6. Knight, T.R. Feedback and grievance resolution. *Ind. Labor Relat. Rev.* **1986**, 39, 585-598.
7. Knight, T.R. Toward a contingency theory of the grievance-arbitration system. In *Advances in Industrial and Labor Relations*; Lipsky, D., Ed.; JAI Press: Greenwich, CT, 1985; Vol. 2, 269-318.
8. Lewin, D. Theoretical perspectives on the modern grievance procedure. In *New Approaches to Labor Unions: Research in Labor Economics*; Reid, J.D., Jr., Ed.; JAI Press: Greenwich, CT, 1983; 127-147. Supplement 2.
9. Lewin, D. Empirical measures of grievance procedure effectiveness. *Labor Law J.* **1984**, 35, 491-499.
10. Peterson, R.B.; Lewin, D. A model for research and analysis of the grievance process. In *Proceedings of the 34th Annual Meeting of the Industrial Relations Research Association*; IRRRA: Madison, WI, 1982; 303-312.
11. Greenberg, J. Looking fair and being fair: managing impressions of organizational justice. In *Research in Organizational Behavior*; Shaw, B.M., Cummings, L.L., Eds.; 1990; Vol. 12, 111-157.
12. Sheppard, B.H.; Lewicki, R.J.; Minton, J. *Organizational Justice*; Lexington: Lexington, MA, 1992.
13. Folger, R.; Greenberg, J. Procedural justice: an interpretive analysis of personnel systems. In *Research in Personnel and Human Resource Management*; Rowland, K.M., Ferris, G.R., Eds.; JAI Press: Greenwich, CT, 1985; 141-183.
14. Fryxell, G.E.; Gordon, M.E. Workplace justice and job satisfaction as predictors of satisfaction with union and management. *Acad. Manage. J.* **1989**, 32, 851-866.
15. Hirschman, A.O. *Exit, Voice, and Loyalty*; Harvard University Press: Cambridge, MA, 1970.
16. Bemmels, B. Exit, voice, and loyalty in employment relationships. In *The Human Resource Management*

- Handbook*; Lewin, D., Mitchell, D.J.B., Zaidi, M.A., Eds.; JAI Press: Greenwich, CT, 1997; 245–259.
17. Rees, D.I. Grievance procedure strength and teacher quits. *Ind. Labor Relat. Rev.* **1991**, *45*, 31–43.
 18. Freeman, R.B.; Medoff, J. *What Do Unions Do?* Basic Books: New York, 1984.
 19. Lewin, D. Dispute resolution in a nonunion firm: a theoretical and empirical analysis. *J. Confl. Resolut.* **1987**, *31*, 465–502.
 20. Boroff, K. Measuring the perceptions of the effectiveness of a workplace complaint procedure. In *Advances in Industrial and Labor Relations*; Sockell, D., Lewin, D., Lipsky, D., Eds.; 1991; 207–235.
 21. Lewin, D.; Peterson, R.B. Behavioral outcomes of grievance activity. *Ind. Relat.* **1999**, *4*, 554–576.
 22. Labig, C.E.; Greer, C.R. Grievance initiation: a literature survey and suggestions for future research. *J. Labor Res.* **1988**, *9*, 1–27.
 23. Allen, R.E.; Keaveny, T. Factors differentiating grievances and non-grievants. *Hum. Relat.* **1985**, *38*, 519–534.
 24. Labig, C.E.; Helburn. Union and management policy influences on grievance initiation. *J. Labor Res.* **1986**, *7*, 269–284.
 25. Kleiner, M.M.; Nicklesburg, G.; Pilarski, A. Monitoring, grievances, and plant performance. *Ind. Relat.* **1995**, *34*, 169–189.
 26. Dalton, D.R.; Todor, W.D. Antecedents of grievance filing behavior: attitudes/behavioral consistency and the union steward. *Acad. Manage. J.* **1982**, *25*, 158–169.
 27. Bemmels, B.; Reshef, Y.; Stratton-Devine, K. The roles of supervisors, employees, and stewards in grievance initiation. *Ind. Labor Relat.* **1991**, *45*, 15–30.
 28. Meyer, D.; Cooke, D. Economic and political factors in formal grievance resolution. *Ind. Relat.* **1988**, *27*, 318–335.
 29. Meyer, D. The political effects of grievance handling by stewards in a local union. *J. Labor Res.* **1994**, *15*, 33–52.
 30. Dastmalchian, A.; Ng, I. Industrial relations climate and grievance outcomes. *Ind. Relat.* **1990**, *45*, 311–324.
 31. Ury, W.; Brett, J.; Goldberg, S. *Getting Disputes Resolved: Designing Systems to Cut the Cost of Conflict*; Jossey-Bass Publishers, Inc.: San Francisco, CA, 1989.
 32. Feuille, P. Dispute resolution frontiers in the unionized workplace. In *Workplace Dispute Resolution: Directions for the Twenty-First Century*; Gleason, S.E., Ed.; Michigan University Press: East Lansing, MI, 1997; 17–55.
 33. Davy, J.A.; Stewart, G.; Anderson, J. Formalization of grievance procedures: a multi-firm and industry study. *J. Labor Res.* **1992**, *13*, 307–316.
 34. Mesch, D.J.; Dalton, D.R. Arbitration in practice: win, lose, or draw? *Hum. Resour. Prof.* **1992**, *44*, 37–41.
 35. Feuille, P. Grievance mediation. In *Employment Dispute Resolution and Worker Rights in the Changing Workplace*; Eaton, A.E., Keefe, J., Eds.; Industrial Relations Research Association: Champaign, IL, 1999; 187–217.
 36. Mesch, D.J. Grievance arbitration in the public sector: a conceptual framework and empirical analysis of public and private sector arbitration cases. *Rev. Public Pers. Adm.* **1995**, 22–36, Fall.
 37. Brett, J.M.; Goldberg, S.P. Wildcat strikes in bituminous coal mining. *Ind. Labor Relat. Rev.* **1979**, *32*, 465–483.
 38. Spencer, D.G. Employee voice and employee retention. *Acad. Manage. J.* **1986**, *29*, 488–502.
 39. Ichniowski, C.; Lewin, D. Grievance procedure and firm performance. In *Human Resources and the Performance of the Firm*; Kleiner, M.M., Block, R.N., Roomkin, M., Salsburg, S.W., Eds.; Industrial Relations Research Association: Madison, WI, 1987; 159–193.
 40. Ichniowski, C. The effects of grievance activity on productivity. *Ind. Labor Relat. Rev.* **1986**, *40*, 75–89.
 41. Katz, H.C.; Kochan, T.A.; Gobeille, K. Industrial relations performance, economic performance, and the effects of quality of working life efforts: an interplant analysis. *Ind. Labor Relat. Rev.* **1983**, *37*, 3–17.
 42. Katz, H.C.; Kochan, T.A.; Weber, M. Assessing the effects of industrial relations and quality of working life efforts on organizational effectiveness. *Acad. Manage. J.* **1985**, *28*, 509–527.
 43. Norsworthy, J.R.; Zabala, C.A. Worker attitudes, worker behavior, and productivity in the U.S. automobile industry, 1959–76. *Ind. Labor Relat. Rev.* **1985**, *38*, 544–557.
 44. Moore, C. *The Mediation Process*, 2nd Ed.; Jossey-Bass Publishers, Inc.: San Francisco, CA, 1996.
 45. Connors, E.K.; Bashore-Smith, B. Employment dispute resolution in the United States: an overview. *Can.-U.S. Law J.* **1991**, *17*, 319.
 46. Goldberg, S.B. The mediation of grievances under a collective bargaining contract: an alternative to arbitration. *Northwest. Univ. Law Rev.* **1982**, *77*, 270–315.
 47. Goldberg, S.B.; Brett, J.M. An experiment in the mediation of grievances. *Mon. Labor Rev.* **1983**, *106*, 23–30.
 48. Brett, J.M.; Goldberg, S.P. Grievance mediation in the coal industry: a field experiment. *Ind. Labor Relat. Rev.* **1983**, *37*, 49–69.
 49. Feuille, P. Why does grievance mediation resolve grievances? *Mediat. J.* **1992**, *8*, 131–146.

Groups and Their Properties

Claus Langfred

John M. Olin School of Business, Washington University, St. Louis, Missouri, U.S.A.

INTRODUCTION

Groups are an inherent part of human social organization and have relatively recently been heavily incorporated into modern organizations in the form of work groups and teams. Groups have many important features that affect their performance and dynamics, and while they are prized for the performance benefits they have compared to the same number of individuals, they can also be a source of inefficiency, conflict, and errors.

As a general rule, effective groups require three things. First, the group must have the knowledge, skills, and abilities (KSAs) necessary to perform the task. Not every member of a group needs to have all of the KSAs for the group to succeed, but the group collectively must have them. Second, motivation and willingness to expend effort toward the group's goal is necessary. If group members are not interested in performing the task, it matters little whether they are able to or not. The converse is also true: no matter how motivated or enthusiastic members of a group are, the group will not succeed without the necessary KSAs. Third is the communication and coordination necessary to carry out the task. Many groups fail in this third category and often do not understand why. These are the three minimum requirements for groups to be effective and successful.

DEFINITION AND TYPES OF GROUPS

A work group (or team) is a collection of individuals who see themselves, and are seen by others, as a social entity and who are embedded in one or more larger social systems. They are interdependent (to some degree) in their tasks and often manage relationships with people external to the group. While most agree with the basic definition by Hackman,^[1] some researchers feel that "team" connotes more than "group," describing a group with relatively more commitment and cohesion, in addition to a greater task focus.

Teams can be subdivided into various categories and typologies, such as parallel, project, cross-functional, management, and others. While such distinctions separate them on the basis of task, it is not clear that they distinguish groups from teams. As a general rule, the terms group and team can be used

interchangeably with little loss of meaning. This article focuses on the small group or team, typically ranging in size from as little as three or four people up to as many as 20 people.

POPULARITY AND PREVALENCE OF GROUPS

The use of work teams has rapidly grown in the United States, with an estimated half of all employees currently working in team situations. Work groups and teams have become critical to organizational effectiveness, and their use has become pervasive in modern organizations. The use of teams and groups ranges from the production floor to the executive suite and covers activities from heart surgery to fire fighting and from dog food manufacturing to space exploration. Teams are used by social service agencies, in the military, and in hospitals, and in schools and universities. It is difficult to identify a field of human endeavor in which groups and teams have not become commonplace. Even in a relatively specialized and individual field like academia, we find groups represented in permanent and temporary committees, departments, and for administrative tasks.

With the increased adoption of groups and teams by a wide range of firms and organizations, and as the use of groups and teams in organizations continues to increase, researchers and practitioners focused their attention on understanding the dynamics, design, and effectiveness of teams. As the use of teams has increased, so has the variety of different team designs. The use of self-managing, empowered, or autonomous groups and teams has increased dramatically in recent decades. Of Fortune 1000 firms, about two-thirds use self-managed work teams, and practically all use employee participation groups of some sort.

CHARACTERISTICS OF GROUPS

Group size is one of the most obvious of group characteristics, and it affects groups in many ways. The ideal group size is between four and eight people. When a group becomes too large, it is more difficult to effectively coordinate the behavior of group members, and the group starts to divide into smaller subgroups

and cliques. Larger groups are more susceptible to various motivational problems, such as free riding and social loafing. If a group is too small, on the other hand, many benefits of groups can be lost, and conflict and polarization are more likely.

The structure of a group can vary widely and refers to the way a group is organized and how various positions in the group relate to one another. Well-established groups will have a much more developed and rigid structure than newly formed groups. Characteristics like roles, status, norms, cohesion, and interdependence are part of a group's structure.

Diversity refers to the variation among group members on a variety of dimensions, such as background, education, nationality, ethnicity, age, gender, etc. If the task of the group involves decision making, problem solving, or creative tasks, diversity allows the group to perform better and with greater accuracy and quality. However, if the group's task is highly interdependent and requires considerable coordination, then diversity can interfere with its ability to perform effectively. While a group with little diversity, and hence few differences, can coordinate and communicate effectively, it cannot be as effective at problem solving and decision making. The benefits or costs of diversity are thus dependent on the task and other characteristics of the group.

Roles are the expected behaviors associated with a given position within a group. Members of a group will generally occupy different social roles. Some roles are formal and may be externally imposed, while others are less formal and may emerge over time. Group members may hold multiple roles, and roles may even be switched or interchanged within a group. Different social roles are usually associated with different degrees of status, which affect group dynamics insofar as members with higher status may have more influence or may be less susceptible to group norms or pressure.

Group norms are the general rules and expectations for appropriate behavior that are accepted by the group. Norms can be explicit, or they can be implicit and emerge over time. Norms generally develop quickly and often have a strong influence on member behavior. Groups are effective at enforcing behavior to conform to the norms in the group, particularly in cohesive groups. The more cohesive a group, the better it is at enforcing conformity to group norms.

Cohesion is the extent to which group members feel a part of the group and wish to remain in the group. Generally, the more cohesive a group, the more loyal and committed group members are to the group. Cohesion can be good or bad, in that high cohesion in a group means it is good at enforcing norms, regardless of what those norms are. If the group is interested in and motivated toward high performance, this is desirable, but it is quite detrimental if the group

is motivated toward other goals, interfering with high performance. Effective teams generally are characterized by high cohesion combined with performance-oriented norms.

Interdependence is the degree to which coordination between individuals in the group is necessary for successful task attainment. Depending on the nature of the task, interdependence can be described as pooled, reciprocal, or sequential.^[2] The interdependence in a group dictates the extent to which coordination and communication are required in the group and is one of the defining and most crucial characteristics of a group. A highly interdependent group with high coordination requirements clearly is different from a group in which close coordination among group members is not necessary for group goals to be realized.

While leadership has long been known to exert a strong influence on group effectiveness, few general conclusions can be drawn about leadership. Depending on organizational contexts—such as the immediate task environment, resources, time pressures, interdependence, diversity, personalities and attitudes of group members, and other factors—different leadership styles may be appropriate to the situation. Under some conditions, an empowering and inclusive leadership style may yield the best results, whereas under other conditions, a directive and dictatorial leadership style may be most effective.

Autonomy has received a lot of attention in recent years, as the prevalence of self-managing teams, autonomous work groups, and empowered teams has increased dramatically. Granting autonomy to groups can lead to better group performance, because groups and group members are closer to the task and, thus, are able to make more accurate task-related decisions than a more “distant” supervisor. However, the increased coordination and communication required for group decision resulting from increased autonomy can also lead to process losses, so while the potential of autonomy is great, the implementation of it is difficult. Autonomy also raises the issue of striking a balance between the amount of autonomy that individual group members have and the amount of autonomy the group as a whole has—another difficult management decision.

Finally, an important characteristic of a group is the stage of its life cycle. Groups, like people, often have a specific life span. They are created at some point, they go through a period of learning and maturation, and they become mature groups. While some groups replenish themselves indefinitely through member turnover, others may end. Where a group is in its life cycle also has implications for dynamics and performance. Young groups are busy developing norms and communication systems and will not be able to perform as well as mature groups. A mature group, on the other hand, might be stable and cohesive but may suffer

from problems of conformity, polarization, and other biases in decision making and problem solving.

BENEFITS OF GROUPS

Some of the benefits of groups are obvious, as a group of people can accomplish things that the same number of individuals cannot. Given effective coordination, groups achieve many feats that are completely outside the scope of the same number of individuals. Groups with high degrees of task interdependence particularly demonstrate this, such as a military assault team or a surgical team involved in a complicated heart procedure.

As groups make decision or solve problems, they are generally more accurate than the average group member and often as accurate as even the best group member. The group decision is much more than just a simple aggregation of individual decisions, involving active discussion and evaluation of ideas, viewpoints, and alternatives. This process of information sharing and interaction, and even conflict and disagreement, generates better decisions.

Often, disagreements lead to novel solutions or alternative viewpoints that increase the quality of decisions. Having a dissenting minority opinion in a group has been found to cause more complex and rich discussion of issues. In this sense, limited conflict in a group contributes to a higher quality of decisions. By the same logic, the amount of diversity in the group also leads to more creative outcomes, which can be further facilitated by a wide variety of existing group decision-making techniques.

Groups, through such characteristics and cohesion and strong norms, can also increase commitment and motivation to a task, such that a group may expend a lot more effort and energy on a task than the same number of individuals. In military organizations, morale and esprit de corps is deliberately encouraged and reinforced so that units will be more motivated and will not give up on difficult tasks, where otherwise, individuals might lose motivation. This is one of the greatest strengths of groups, the social identification that can be generated, and is a method by which members of a group may be much more motivated toward a goal than a comparable number of individuals.

A more “hard-wired” variation of this motivational effect that membership in a group may have on individual members is the phenomenon of social facilitation. Social facilitation is the effect that the “mere presence” of others has on performance.^[3] When other people are present, individuals tend to perform better or work harder. The theory is more complex, and disagreement exists as to the mechanisms and details of the phenomenon, but it is indisputable that the

presence of other group members affects the effort level of an individual.

Another aspect of the strong social interactions that can improve group and team performance over that of the same number of individuals is trust. Along with strong cohesion and norms, highly effective teams are characterized by high levels of trust within the team and toward the team leader or supervisor. In the context of a team, particularly a highly interdependent team, trust is much more likely to develop than in an independent group of individuals.

Thus, a stable and cohesive team consisting of relatively interdependent people, who share performance-oriented norms and trust one another, and who possess the KSAs, motivation, and communication skills to accomplish their goal, will be far more effective than the same number of individuals working toward the same goal. In some cases, the individuals will not be able to accomplish the goal, and in others they will simply perform at a much lower level.

POTENTIAL HAZARDS OF GROUPS

While a variety of benefits to groups are outlined above, there are also aspects of group dynamics that can hurt performance. It is important to note that in the balance, the benefits of groups far outweigh the hazards (hence, the widespread use in modern organizations), but some hazards exist and can lead to poor performance and dysfunction. Understanding and recognizing these limitations is therefore important in order to avoid them. Some benefits of groups, like strong group identity and competition, can become potential problems if not managed, and these are discussed as well.

One of these potential limitations is process loss. Because groups need to coordinate the activities of members (depending on the amount of task interdependence in the group), inefficient coordination and communications can harm performance. In the extreme case, performance can be extremely poor, and process loss can outweigh any benefits of the group, as group members spend too much time trying to communicate and coordinate rather than working.

Groups can also suffer motivation losses, among individual group members or the group as a whole. Social loafing is a phenomenon where individuals decrease their effort level as the size of the group increases.^[4] The effect is strongest when tasks are simple and additive or when individual effort is not easily identifiable but can be minimized when individual effort is accurately identifiable or when tasks are sufficiently involving.

Another motivational problem that occurs is free riding, which occurs when a member of a group exerts less effort and lets others in the group do more of the

work. Free riding is somewhat more deliberate than social loafing and occurs when people perceive their efforts to be dispensable. In other words, if an individual believes that the group will succeed regardless of his efforts, he will be less likely to work hard. In general, groups with high task interdependence are less susceptible to extensive free riding, as are groups in which individual effort is more easily identifiable.

Free riding and social loafing can lead to other motivation losses, namely, those based on equity theory. If other group members perceive that some in the group are free riding off their efforts, they tend to lessen their effort level to reduce the perceived inequity of “carrying” the free rider. This loss of motivation has been called the “sucker effect”^[5] and is a conscious and deliberate behavior. It is important to note that much as it appears that free riding can lead to a downward performance spiral, it is unlikely in cohesive groups with strong performance norms. It is more likely in groups of a temporary nature or in groups in which norms and trust have not fully developed.

Conformity effects are powerful in groups.^[6] In general, people tend to conform to the group (even when they know it is wrong) when they face a unanimous group, the judgment is difficult, they value or admire the group, their deviation is easily identifiable, or they are scared. On the other hand, people conform to the group less when they have great confidence in their expertise on the issue, they have high status in the group, they are strongly committed to their initial view, or they do not like or respect the group. This is generally a good thing, in that conformity allows groups to be unified and decisive in their action, but it has a potential downside in that creativity can be stifled, and group decision making can be superficial or even erroneous.

Such pressure to conform or not to express a different view is the central component of the groupthink phenomenon,^[7] in which valuable information is ignored or not even raised during discussion. The best examples of groupthink are in foreign policy, such as the Bay of Pigs invasion, Argentina’s decision to invade the Falkland islands, or the escalation of the Vietnam War under Lyndon Johnson, suggesting that a crisis atmosphere contributes to the problem.

Groups can also become more polarized in opinions or behavior than the individual group members were originally. This effect occurs when group members share an underlying agreement. When this is the case, group discussion often leads to group consensus around a more extreme viewpoint than the one originally held by the group members. Group polarization is sometimes also discussed as *risky shift* in groups, in which group discussion leads to riskier decisions than individual group members’ decisions. Again, like conformity, this can be a good thing as well as a bad thing, depending on circumstances.

The risk of disagreement and conflict is, of course, a risk in any social group. While a solitary individual working on a task has no potential for conflict, groups, by definition, run the risk of conflict. Even though some disagreement and dissent contributes to better and more complex decisions, too much disagreement can obviously lead to conflict that interferes with the coordination and stability in the group, and ultimately harms performance. While a potential threat, a stable group with members that possess good communication skills is relatively immune to such destructive levels of conflict.

Groups that exist in an environment with other groups, which is the case in most organizations, also suffer from a number of perceptions and behaviors that appear to be ingrained in humans. One of these effects is in-group favoritism and out-group discrimination. Group members tend to favor their own group and discriminate against other groups, in terms of allocating resources or rewards. This effect, called the minimal group paradigm, can even be triggered by creating groups based on insignificant and arbitrary differences^[8] and is, thus, quite powerful when applied to existing groups that differ on significant and salient dimensions. Individuals perceive members of their own group to be more varied and individually distinct (in-group heterogeneity) and perceive members of other groups as being more similar to one another and having less variation (out-group homogeneity). It is worth noting that this is again only a *potential* hazard. Having groups with a strong identity and commitment to the group over other groups is generally a positive trait that will contribute to cohesiveness and trust.

Groups can also become quite competitive and aggressive toward one another, leading to intergroup conflict, especially when competing for scarce resources or when incompatible goals exist.^[9] Many groups in large organizations will be competing with other groups for resources, rewards, and other forms of recognition, and conflict can intensify effects such as perceptions of out-group homogeneity. When managed, this is a great source of motivation and energy for groups, and group identification combined with competitiveness can spur greater performance. Some of the greatest team performance is created in the context of intense competition with other teams. Very beneficial to organizations, this only becomes a potential problem if it is allowed to spin out of control and become destructive and non-performance oriented.

CONCLUSION

Groups are a pervasive and inescapable aspect of life in modern organizations, and understanding and managing them is crucial to overall organizational performance. While groups can easily outperform the same

number of independent individuals, it is important to be aware of potential limitations, related to process and coordination, as well as to motivational and social psychological effects, that can interfere with the considerable benefits of groups and teams.

REFERENCES

1. Hackman, J.R. Designing work for individuals and for groups. In *Perspectives on Behavior in Organizations*; Hackman, J.R.; McGraw-Hill: New York, 1983; 242–256.
2. Thompson, J.D. *Organizations in Action*; McGraw-Hill: New York, 1967.
3. Zajonc, R.B. Social facilitation. *Science* **1965**, *149*, 269–274.
4. Latane, B.; Williams, K.; Harkins, S. Many hands make light the work: the causes and consequences of social loafing. *J. Pers. Soc. Psychol.* **1979**, *37* (6), 822–832.
5. Kerr, N.L. Motivation losses in task-performing groups: a social dilemma analysis. *J. Pers. Soc. Psychol.* **1983**, *45*, 819–828.
6. Asch, S.E. Opinions and social pressure. *Sci. Am.* **1955**, *193* (5), 31–35.
7. Janis, I.L. *Victims of Groupthink*; Houghton Mifflin: Boston, 1972.
8. Tajfel, H. *Human Groups and Social Categories: Studies in Social Psychology*; Cambridge University Press: Cambridge, England, 1981.
9. Sherif, M.; Harvey, O.J.; White, B.J.; Hood, W.R.; Sherif, C.W. *Intergroup Cooperation and Competition: The Robbers Cave Experiment*; University Book Exchange: Norman, OK, 1961.

Health Care Decision Analysis, Alternatives in

Krishna S. Dhir

Campbell School of Business, Berry College, Mount Berry, Georgia, U.S.A.

INTRODUCTION

Health care professionals routinely face complex decisions. There are a number of factors that contribute to increasing complexity in practically all aspects of health care professions. Research knowledge is growing rapidly. Technological sophistication is increasing. Health care organizations have become more complex. Laws are not simple. The challenge of decision making in the health care industry is becoming more and more daunting. Adding further complexity to health care decision-making process is the fact that physicians are no longer the sole arbitrators of patients health care provided to the patients. The process has become democratized, and even politicized, with the entry of other parties, such as health insurers, regulators, politicians, lawyers, ethicists, family members, and indeed, patients, in the decision-making arena. These parties are not necessarily adequately trained to tackle the problems that confront them. They may not understand the dynamics involved in regard to the various factors to be taken into account. Often the relationship between actions taken and desired outcomes is uncertain. All options available to the decision maker may be poorly specified or understood. Therefore, unassisted, decision makers may make poor decisions. In this article we review some of the theories and frameworks available for the analysis of decisions, which could assist the decision-making process. However, before we explore the alternatives available to decision makers, we further explore the nature of clinical decision making.

THE NATURE OF CLINICAL DECISION MAKING

Subjectivity plays an important role in practically every aspect of medical practice, including the tasks of diagnosing, assessing, and treating. Although clinical as well as laboratory techniques continue to become increasingly sophisticated and precise, the integration of the information obtained by health care professional remains a subjective process. Clinical judgment has its limitations. To make a decision based on judgment, the health care professional processes social information. This information is obtained from

the environment and could come from a variety of sources. The decision maker must integrate this information. The cognitive image provides a representation of the environment based on the professional's past training and experiences, essentially predisposing the person to interpret, integrate, and respond to the information in predictable ways. This adds to the complexity of the interaction between the clinician and the patient. Each may bring very different sets of beliefs, values, experiences, and perspectives to the decision. Additionally, the information may be imperfect or incomplete. The outcome may be uncertain. The value placed by the patient on the outcome may not be clear. In light of all these considerations, clinical judgments are probabilistic. The process of clinical judgment is fundamentally a covert process. Individuals seldom describe their judgment process accurately. Ordinarily, the only means of explaining judgments are introspection and guessing at reasons. These explanations are generally incomplete and misleading. Subjective reporting is fallible. Judgments are thus inaccurately reported and are observed to be inconsistent. Identical circumstances do not always lead to identical judgments. Observations of inconsistencies may send an observer looking for hidden motives or incompetence on the part of the decision maker.^[1] Robust evidence from randomized controlled clinical trials often does not support clinical decisions by health care professionals. Therefore, many clinical decisions have to be made on the basis of limited evidence, even when robust information is absent.^[2,3] Additional considerations impact the decision-making process of health care professionals. The pressure on them to take account of both costs and effectiveness of alternative treatment strategies is mounting. Technological and pharmacological innovations have generated a range of treatment options where hitherto options were limited. For instance, kidney inefficiencies may be treated either by drugs, or dialysis, or through transplant. Hypertension may be treated with beta-blockers or diuretics. Patient may prefer increased quality of life to longevity, or vice versa.

Fortunately, a number of theories and frameworks are available that offer considerable promise to health care professionals. Approaches based on these frameworks deal with the difficulties described above by

formalizing the manner in which the health care decision problems are structured, and making explicit the manner in which they are then analyzed. If the parameters of one's judgmental policy orientations could be identified, then we might model the process of that person's judgment and develop support systems to aid clinical judgment. This process may be explored by posing the following questions:

1. What is being judged? This refers to the clinical or medical criteria being assessed.
2. What factors influence the individual's judgments? Answers will identify variables that influence the professional's assessment of the criteria. These factors should have mutually exclusive characteristics or properties.
3. What relative emphasis or weight does the individual put on each of these factors? This question refers to the relative weights assigned to the factors the professional takes into account in making the clinical assessment. One source of disagreement between health care decision makers arises from the fact that different weights are likely to be attached by different decision makers to the factors.
4. How does the health care professional integrate the information regarding each factor to arrive at an overall judgment? Identification of the mathematical relationship, which describes the dependence of the overall judgment on the factors considered, is important. The relationship between each factor and the overall judgment may be linear or non-linear, and the contribution of each factor to the overall judgment may be positive or negative. The nature of dependence of the overall judgment on each factor is referred to as that factor's function-form.
5. How consistent is the health professional in making judgments? An individual may make different judgments about the same situation on different occasions. At least two characteristics of the judgment task are known to affect consistency: task complexity and task uncertainty. Studies show that consistency is lower when the judgment task is complex and when the task requires the use of non-linear rather than linear function forms. Even when a decision maker intends to use a specific judgment rule as defined by a specific set of factors, relative weights and function forms, his or her judgments or assessments may deviate from those suggested by that rule. In the absence of explicit and immediate feedback on judgment, the decision maker may be unaware of the degree to which the actual judgment or assessment deviates from that intended.

Consistency is different from accuracy. A decision maker can be accurate but inconsistent, consistent but inaccurate, both consistent and accurate, or neither. Consistency has to do with the reliability of the decision maker in executing the intended judgmental policy as defined by a specific set of factors, relative weights and function forms. When deviation between the actual judgment and that intended is low, consistency is greater. On the other hand, accuracy has to do with the validity of the policy itself—whether the judgmental policy executed (the specific set of factors, relative weights and function forms) is indeed the one that was to be executed.

ALTERNATIVE APPROACHES

Among the various theories and frameworks for analysis of the decision-making process available to the health care professionals are the following: Decision Theory and Multiattribute Utility Theory,^[4] Behavioral Decision Theory,^[5] Analytic Hierarchy Process,^[6] Prospect Theory,^[7] Social Judgment Theory,^[1,8,9] Information Integration Theory,^[10] Attribution Theory,^[11] Conflict Theory,^[12,13] Constraint Theory,^[14] Bazerman's bargaining and negotiation framework,^[15] and Fuzzy Decision Theory.^[16] While Decision Theory evaluates unidimensional utility functions over single attributes, Multiattribute Utility Theory seeks to maximize the overall utility derived from a set of criteria or attributes. They both are prescriptive in nature. Both these theories assume rational behavior. Decision making is reduced to knowing the various alternatives available to the health professional for choice, the probability of the respective alternatives occurring, and the worth or value of the outcomes of the respective alternatives to the decision maker. The decision is made by mathematically discounting the expected value to be realized from each alternative by the probability of realizing them. The probability may be objectively known through historical or empirical observations, or be subjectively estimated.^[4] Behavioral Decision Theory is based on the subjective expected utility concept, and like Decision Theory and Multiattribute Utility Theory, attempts to prescribe rational decisions. It attempts to explain the less-than-optimal behavior of decision makers in psychological terms.^[5] Analytic Hierarchy Process provides a unique mathematical approach for the analysis of multiattribute decision making that does not require the direct assessment of utilities or probabilities as demanded by decision theory and behavioral decision theory approaches. It describes a decision problem in terms of linked hierarchical layers of considerations. Paired comparisons among factors considered are made with respect to a consideration of

the next highest level of hierarchy to yield a final set of ranking among the various options.^[6] Prospect Theory, which is sometimes also referred to as the psychological decision theory, moves beyond description of decision-making process to explanation and prediction of decision behavior. It attempts to explain why decision makers deviate from rational prescriptions. It argues that decision makers psychologically transform probabilities of outcomes into decision weights or emphasis, and value of outcomes into psychological or perceived worth, before indicating their preference or making their choice. It seeks the cognitive sources of departure from the criteria of rationality and attempts to make decision makers aware of their own errors and biases.^[7] Social Judgment Theory, which evolved from the work of Brunswik,^[9] explores interpersonal conflict arising from different judgments and lends itself to applications in policy analysis. The “social” element in social judgment theory is derived from the interaction it affords among two or more parties through improved communication.^[1,8] Information Integration Theory combines a psychological theory of measurement with a theory of information integration to produce a coherent expression of the subjective nature of human judgment. It seeks to discover the metric form of the cognitive algebra employed in cognitive activities, e.g., is the information integrated in one situation by one algebraic principle, and in another situation by another algebraic situation?^[10] Attribution Theory draws from Gestalt psychology, and focuses on one’s explanation of one’s own behavior and the behavior of others. It views people as observing events in their environment and seeking causes of events by evaluating available information. It does not seek models of individual judgment or decision.^[11] Conflict Theory relates to the group decision-making theory of groupthink. It provides a descriptive theoretical perspective on how people handle decisional conflicts where potential outcomes are of some consequence to the ones making the decision. It accounts for the effect of psychological stress on group decision-making behavior.^[12,13] The more comprehensive Constraint Theory seeks to explain how cognitive (time, knowledge, ability, personality, attitudes, etc.), affiliative (need for recognition, acceptance, conformity, obedience, etc.) and egocentric (personal motives, prestige, stress, self-esteem, and other emotional needs, etc.) constraints influence the decision-making process, often resulting in suboptimal outcomes.^[14] Bazerman’s adaptation of Prospect Theory to bargaining and negotiation is also of interest, especially in view of increased democratization of the decision-making process in the health care industry. He extends the application of the Prospect Theory to situations where multiple parties jointly make decisions to resolve conflicting interests.^[15] Fuzzy Decision Theory assumes

that human judgment and decision making is based on a complex system whose elements were fuzzy, yielding imprecise, even vague, measures of information. It deals with possibility and believability (e.g., can the event happen) instead of probability (e.g., will the event happen).^[16] Cooksey,^[8] on pages 26–54, offers detailed discussions of these theories, and also of some others not discussed here, in the context of their applicability to the analysis of judgment and decision.

CHOOSING A FRAMEWORK

Health care professionals encounter a variety of decision-making scenarios. When confronting problems that require coherence, they would likely prefer to project their clinical judgment competence in terms of logical, mathematical, or statistical rationality. In such cases, they are likely to benefit from analytical frameworks that are primarily mathematical or econometric in approach, such as those offered through Decision Theory, Multiattribute Utility Theory, and Analytic Hierarchy Process, with their emphasis on prescription of how rational decisions should be made; Behavioral Decision Theory and Prospect Theory, which explore *why* decision makers depart from rational prescriptions; and Bazerman’s approach on bargaining and negotiation that is philosophically based on rational prescription. The objective in the health care profession, however, is not always to prescribe an approach to clinical health care professional. When confronting problems that demand clinical judgment, the health care professional may benefit more from an *explicit description* of his or her own judgment process rather than a *prescription* of what the judgment process ought to be. In such situations, frameworks offered through Social Judgment Theory, Information Integration Theory, Conflict Theory, and Constraint Theory may be of particular benefit. Social Judgment Theory is particularly useful where the objective is to obtain a model of individual judgment process, which may then be used to develop judgment and decision aids. Attribution Theory does not do this. Attribution Theory generally assumes that the person making the judgment does not influence the cognitive processes underlying causal judgments. The Information Integration Theory has generally stayed away from applications. The framework offered by the Social Judgment Theory is also easy to use. The methodology involved yields relatively straightforward and mathematically clean results that are easily understood by decision makers. Further, literature reports that the simple additive models of decision making based on regression analysis seem to be at least as accurate as the more complex models and are preferred by many users with respect to most criteria of desirability.^[1]

CONCLUSION

In this article we explore the complex nature of clinical decision making. We identified the various factors that contribute to the complexity of the clinical decision-making process. Information is usually incomplete. Clinical decision-making process is becoming increasingly democratized. The various decision makers bring their own beliefs and values to the decision problem at hand and are predisposed to interpret, integrate, and respond to information in predictable ways. The judgment process itself is fallible and covert. Self-reporting of judgmental policies is inaccurate. However, it is possible to describe a decision maker's judgmental policies underlying the decision-making process in terms a set of parameters. These are identified.

Various theories and frameworks available to the health care decision maker to analyze the decision-making process are discussed. Some of these facilitate logical, mathematical, or economic rationality. These tend to be prescriptive, offering what rational decisions might be taken, or explaining why the decision taken may differ from rationality. Others facilitate a description of the decision-making process as it is, avoiding a prescription. Yet others facilitate analysis of the decision-making process when the information available is imprecise and vague, and the decision maker also brings to the decision-making process his or her own perceptions of possibilities or believability. Suggestions are offered as to how the health care professional might choose an approach to analyze the decision-making process from among the alternatives available.

REFERENCES

1. Dhir, K.S. Analysis of clinical judgment. In *Handbook of Health Administration and Policy*; Kilpatrick,

- A.O., Johnson, J.A., Jr., Eds.; Marcel Dekker: New York, 1999; 557–569.
2. Detsky, A.S.; Naglie, G.; Krahm, M.D.; Naimark, D.; Redelmeier, D.A. Primer on medical decision analysis: Part 1—getting started. *Med. Decis. Mak.* **1997**, *17*, 123–125.
3. Tavakoli, M.; Davies, H.T.O.; Thomson, R. Decision analysis in evidence-based decision making. *J. Eval. Clin. Pract.* **2000**, *6* (2), 111–120.
4. Keeney, R.L.; Raiffa, H. *Decision with Multiple Objectives: Preferences and Value Tradeoffs*; John Wiley and Sons: New York, NY, 1976.
5. Edwards, W., Ed.; *Utility Theories: Measurements and Applications*; Kluwer Academic Publishers: Boston, MA, 1992.
6. Saaty, T.L. *The Analytic Hierarchy Process: Planning, Priority Setting, and Resource Allocation*; RWS Publications: Pittsburgh, PA, 1990.
7. Kahneman, D.; Tversky, A. Prospect theory: an analysis of decision under risk. *Econometrica* **1979**, *47*, 263–291.
8. Cooksey, R.W. *Judgment Analysis: Theory, Methods, and Applications*; Academic Press: San Diego, CA, 1996.
9. Brunswik, E. *Perception and the Representative Design of Psychological Experiments*, 2nd Ed.; University of California Press: Berkeley, CA, 1956.
10. Anderson, N.H. *Methods of Information Integration Theory*; Academic Press: New York, 1982.
11. Kelley, H.H. The process of causal attribution. *Am. Psychol.* **1973**, *28*, 107–128.
12. Janis, I.L.; Mann, L. *Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment*; The Free Press: New York, 1977.
13. Janis, I.L. *Groupthink*, 2nd Ed.; Houghton Mifflin: Boston, MA, 1982.
14. Janis, I.L. *Crucial Decisions: Leadership in Policymaking and Crisis Management*; The Free Press: New York, 1989.
15. Bazerman, M.H. *Judgment in Managerial Decision Making*, 3rd Ed.; John Wiley and Sons: New York, 1994.
16. Zadeh, L.A.; Fu, K.-S.; Tanaka, K.; Shimura, M., Eds.; *Fuzzy Sets and their Applications to Cognitive and Decision Processes*; Academic Press: New York, 1975.

Health Care Decision-Making Behavior, Emerging Paradigms of

Krishna S. Dhir

Campbell School of Business, Berry College, Mount Berry, Georgia, U.S.A.

INTRODUCTION

In this article we examine some of the decision-making processes found in the health care industry. We review a set of paradigms to assist the understanding of emerging health care decision-making behaviors. The concept of paradigms should be differentiated from the concept of models. A model is an empirical, sometimes idealistic representation of a particular process in the real world. However, a paradigm embodies a distinctive and more or less coherent set of general ideas and principles, whether intellectual, ideological, ethical, or otherwise, which serve to make sense of and provide guidance for the understanding of the process.

Decision-making processes in the health care industry are becoming increasingly complex. In this day and age of exciting medical developments, it is easy to conjure up images of patients going to hospitals routinely for replacement of various parts of their anatomy, no matter how critical their function. However, new medical technologies are complicating clinical practice. The emergence of technologies that make heart, liver and other organ grafts, transplants and prosthetic replacements feasible demands reconsideration of definitions of life and death. Difficult decisions, many with consequences not well understood or not previously encountered, have to be made. As we come to grip with these, new paradigms are evolving to assist in the understanding of corresponding behaviors. Some of the emerging paradigms of health care decision-making behavior are discussed below. Specifically, we present a discussion of ethical, courageous, and virtuous decision-making behaviors.

ETHICAL DECISION-MAKING BEHAVIOR

An ethical dilemma is characterized by a requirement of 1) a choice to be made between equally desirable or equally unsatisfactory alternatives; 2) an assignment of different or competing priorities and responsibilities to alternatives; or 3) solving a problem that has no satisfactory solution.^[1] The dilemma may emanate from the decision maker's system of values, principles, or sense of duty, and may be exacerbated by uncertainties

of outcome or consequences resulting from the choice, assignment, or solution.

In the health care industry, the Hippocratic Oath is among the earliest representation of an ethical decision-making behavior paradigm. Nevertheless, some writers have observed that in America medical ethics may not be getting the attention it needs,^[2] and several medical schools have abandoned the practice of new students taking the Hippocratic Oath. Others see reason for hope in the transition from the personal ethics of oaths to the communal professional ethics of codes, and applaud the "revolutionary" transition from personally interpreted "gentlemanly" ethics sworn to by individual practitioners to collaboratively interpreted professional ethics.^[3] Through past few decades, the interest in the teaching of *ethical* decision-making behavior in the professions has been growing. Epstein has offered a sweeping review of the development of professional ethics and corporate social policy.^[4]

Practically every aspect of medical practice involves ethical considerations. These include issues pertaining to balancing between morbidity and mortality, the development of treatment methods, cultural contexts, obtaining informed consent, and whistleblowing. Today, health care decision makers are forced to cope with unprecedented legal, socio-political, economic and other issues, such as having to decide who gets an expensive treatment and who does not; and whose treatment gets funded by the grants for experimental treatment and who must pay from his or her own resources. They must determine whether societal resources of sizeable amounts are better spent on development of a mechanical heart or on a prophylactic public health program.

New technologies take physicians, insurers, regulators, politicians, ethicists, family members, patients, and indeed the entire society, through uncharted territory. Not all consequences of new technologies are known. Bates highlights the ethical issues of consent to surgical treatment, linking the issue with law, conditions that render the consent invalid, and implications of poor handling of the consent process.^[5] Weijer points to the need for placebo-controlled trials for novel surgical intervention.^[6] However, what should serve as the

placebo control remains controversial. Furthermore, while blinding is a desirable feature of a randomized controlled trial, it is not indispensable. Efforts to preserve blinding become unduly burdensome on patients, as in the case of trials comparing chemotherapy and radiation therapy for cancer treatment. Typically, placebo drug trials are viewed as a therapeutic intervention. In surgery trials, could sham surgery be viewed in the same way? This poses an interesting challenge for ethical analysis of risk.

Medical science is based on a strong scientific bias that demands clinical trials to mirror laboratory experiments in which variables are manipulated one at a time to facilitate attribution of effects to cause. In reality, clinical care and human response to disease are simply too complex. While the physicians' decisions are based on the state-of-the-art knowledge of medicine, it is also by necessity based on assumptions of the cause-effect relationships between biological factors, and between intervention and consequences. What further complicates matters is that physicians are no longer the sole arbitrators of health care provided to the patients. Other parties, such as health insurers, regulators, politicians, lawyers, ethicists, family members, and indeed patients, have a say in the matter. These parties have now entered the arena that was once the sole concern of physicians. Democratization of the decision-making process has added to the complexity of the emerging paradigms of ethical decision-making behavior in the health care industry. For instance, most adult Americans continue to favor the right to euthanasia and physician-assisted suicide, while lawyers, politicians, and ethicists continue to have difficulty in dealing with their implications. To deal with ethical issues emerging from innovation in medicine and surgery, health care decision makers are increasingly referring to what Agich calls the regulatory ethics paradigm.^[7] The paradigm holds that deviations from standard care involves experimentation requiring the application of a set of procedures designed to assure the protection of the rights and welfare of the subjects of research. This paradigm regards innovative treatment as a departure from standard or accepted treatment, irrespective of whether the accepted treatment is effective or a burden. Innovative treatments are regarded as suspect and questionable until they are framed in a research protocol with prerequisite informed consent mechanisms. The paradigm requires that to be ethically defensible: 1) a research protocol must be prescribed to test the innovative treatment, based on scientific research methods; 2) formal mechanisms of informed consent must be implemented; and 3) professionally competent review board must review the prescribed protocol. Unfortunately, in practice "the complex processes characteristic of clinical innovation are often not reducible to

a scientific protocol... The demand that a clinical trial be undertaken in a field of medicine undergoing rapid and dynamic development can actually thwart innovation."^[7] The recent revision of the *Declaration of Helsinki* recognizes that with informed consent of the patient, the physician should be free to apply unproven or new preventative, diagnostic or therapeutic strategies where proven methods are non-existent or ineffective, if in the physician's judgment these strategies offer hope of saving life, alleviating suffering, or restoring health.^[7]

COURAGEOUS DECISION-MAKING BEHAVIOR

Ethical decision-making behavior may or may not demand courage. Similarly, courageous decisions may have little to do with ethics. Generally, courageous behavior is "characterized by efforts to be productive, make contributions, and help others and results in a sense of personal integrity and thriving."^[8] It is recognized as a complex phenomenon that is promoted and sustained by various interpersonal and intrapersonal factors. In the health care industry, courageous decision-making behavior has received little empirical scrutiny although anecdotal accounts of courageous behavior by patients exist.^[9] Psychologists such as Deutsch and Kohut have reflected on its nature.^[9,10] Deutsch pointed to the need for the use of actual social situations, especially ones of conflict and crisis, as a kind of laboratory for the study of social process.^[11] He concluded that courage is manifested when the inner conviction (regarding the situation relevant to the courageous act) exceeds the perceived punishment potential; it is not manifested when the reverse holds true. He also differentiated between non-conformity, independence, and courage. Philosophers such as Rorty^[12] and Walton,^[13] and theologians such as Tillich,^[10] too, have speculated about the nature of courage.^[8] Unfortunately, little attention has been given to the role of courage in the behavior of health care professionals as they innovate, discover, and practice their professions. In recent decades, however, attempts have been made to understand courageous behavior among patients with long-term health concerns.

Finfgeld conducted a meta-interpretative study of courage among individuals with long-term illness based on two extant works in psychology, and four in nursing.^[8] A meta-interpretative study is one in which findings from a number of qualitative works are synthesized to yield a deeper interpretation of the phenomenon of interest. The goal of such a study is to enrich the theoretical basis of a concept. This study offers a paradigm for courageous decision-making behavior in the health care industry. Finfgeld's meta-study included studies of courage in: 1) a middle-aged

woman with terminal cancer; 2) hospitalized adolescents with a variety of long-term health concerns; 3) sexual assault victims, individuals with severe physical disabilities, and people who had contracted human immunodeficiency virus or had acquired immunodeficiency syndrome; 4) chronically ill elderly; 5) young adults; and 6) middle-aged adults.^[8]

Through the meta-study, Finfgeld found that the ability to be courageous develops over time and requires efforts to fully accept reality, solve problems based on this discernment of reality, and push beyond ongoing struggles.^[8] Becoming courageous involves a lifelong learning process. The learning process begins in youth and continues through efforts of coping with persistent threats to well being. Being courageous is preceded by perceived threat such as uncertainty, loss, personal limitation, and powerlessness, helplessness, lack of control, pain, and embarrassment. The ability to be courageous is acquired through an ongoing bi-directional progressive-regressive process in which disappointments and frustrations can hinder or reverse the evolution of courageous behavior. Throughout its ongoing emergence and manifestation, courage lies on a continuum of behaviors, with non-courageous behavior at one end, courageous behavior at the other end, and different degrees of coping between the two ends. Courageous individuals take responsibility and are productive, and push themselves beyond coping.

The meta-study revealed that struggling to fully accept and comprehend a threatening situation is an essential component of being courageous. Comprehensive acceptance and understanding are enhanced by active acquisition of information, and helping others with similar problems. With full awareness, issues of well being, such as threats, struggles, and even death, are reframed and perceived to be manageable challenges. As a result of living their lives courageously, individuals experience an altered self-perception; enhanced sense of personal integrity, self satisfaction, and pride; and a sense that life has been lived well. Health care providers facilitate this process by demonstrating competence and communicating effectively. Outcomes of being courageous include personal integrity and thriving in the midst of normality. Thriving includes a feeling of power; vitality, zest; and joy and involvement in life.

Finfgeld's meta-study concluded that courage is not limitless, and the process of becoming and being courageous is dependent on intrapersonal and interpersonal factors.^[8] Cuff suggests that courage is based on a motivating purpose or mission in life.^[14] Finfgeld indicates that values promote and sustain courage by rendering powerful goals and expectations. Individuals strive to be fearless, avoid cowardice, help others, and live up to personal expectations. Hope is another factor that helps to instill courage. Patients hope that their

conditions will not deteriorate, if not improve. They also hope that if conditions do worsen, they would experience peaceful death. Individuals with ability to be courageous demonstrate a positive attitude, going beyond the norm to help others and contribute to the society. Behavior not demonstrated by them include persistent anger and pessimism, hopelessness, relinquishing of responsibility for self-care, risk avoidance, ignoring the welfare of others, and substance abuse.^[8]

VIRTUOUS DECISION-MAKING BEHAVIOR

In the evolving literature on ethical decision making one can discern emergent interest in *virtuous* decision-making behavior.^[15,16] Virtuous behavior is demonstrated by the courage to act toward protecting human welfare even with incomplete information about the potentially significant cost to the decision maker. Ethical behavior need not necessarily be virtuous or courageous. It may very well be motivated by self-interest. Unethical behavior can also demand courage. However, virtuous behavior is both ethical and courageous. Various models suggested for the study of ethics are predominantly rule-based or rule-directed. Students are presented with a narrative case and warned of the presence of an ethical dilemma. In practice, however, decision makers do not usually enjoy the benefit of being forewarned. They identify problems, formulate solutions, and implement plans, while performing under varying degrees of stress, engaging in parallel sets of concomitant activities, and interacting with a number of people over a range of decisions, all within a time constraint. Normative models of ethical decision making seek absolute truths about a decision. Positive models seek to explain actual decision-making behavior. Unfortunately, these models prove inadequate in their capacity to explain virtuous behavior.

The contemporary discussion on the ethics and moral obligation of decision making has been dominated by two major theories of principle: The deontological approach studies the decision-making behavior in terms of binding obligations, as in duty. The utilitarian approach studies decision-making behavior in terms of the importance of utility over beauty or other considerations. Unfortunately, these theories fall short of explaining behavior emanating from considerations of virtue. Other theories, such as Hobbes' non-utilitarian theory of egoism, too, are of little help in understanding virtuous decision making. However, Dhir has recently offered a new paradigm for the analysis of the virtuous decision making behavior.^[15] According to him, the process of virtuous decision making is evident when an agent attempts to ameliorate a

special class of ethical dilemma. The elements present in such decision-making process are as follows:

- Action is demanded of an agent.
- The situation is confronted unexpectedly, with little foreknowledge of the emergent situation.
- The situation is not of the agent's making.
- The situation consists of a conflict between alternative courses of action.
- At least one of the courses of action benefits the agent's self interests, satisfying the agent's individual obligations, often with relatively low associated costs or risks borne by the agent.
- Other courses of action are in interest of others, raising the issue of the individual's social responsibility, often with relatively high associated costs or risks borne by the agent primarily for the benefit of others.
- The timing may be awkward or inconvenient.
- The choice made may have significant consequences for the agent and for others.
- To be virtuous, the agent must first recognize the dilemma, that is, the agent must recognize that 1) there are conflicting obligations to be met; and 2) there is no solution that would satisfy all demands of the situation.
- Knowledge is a prerequisite for virtuous decision making. Through knowledge, the agent must seek to wisdom, or understanding, or deep insight.
- The individual must make an informed choice, with awareness of the consequences posed by the alternative actions available to the agent for all parties affected.

Consider the actions of whistleblower Jeffrey Wigand, who exposed corporate wrongdoing in spite of threats to his life and career. Dr. Wigand was Vice President of Research and Development at Brown & Williamson from 1989 to 1993. After 10 months with the company, he realized that he had a dilemma. His employers were denying that their tobacco-based products had harmful effects. He had an attractive income. He had a wife, and two daughters, of whom one required extensive medical coverage. So he looked the other way. He was not ready to face the consequences of taking an adversarial role against Brown & Williamson. Then, laboratory testing showed a controversial pipe tobacco additive, called Coumarin, to be a lung-specific carcinogen in mice and rats. In 1993, he took issue with the Brown & Williamson's continued use of Coumarin in pipe tobacco. They terminated him.

In 1994, the CEOs of seven major tobacco companies swore at congressional hearings on the effects of tobacco that nicotine was not addictive. Wigand had a confidentiality agreement with Brown & Williamson.

He decided to expose the perjury. He broke the confidentiality agreement with his former employers and appeared on *60 Minutes* with Mike Wallace in 1996. He talked about the smoking issues. He also described how his family was being harassed with death threats. Initially, CBS shelved the interview, fearing a lawsuit from his former employer. Personal fallout from the stress included a divorce from his wife that same year and financial insecurity. The interview was subsequently aired. Brown & Williamson sued Wigand for breach of confidentiality. However, with the settlement between the tobacco industry and the states, this suit was dropped in 1997.

In relation to the paradigm presented here, Wigand recognized that he had a dilemma in terms of the conflicting demands of 1) his personal well being, security of his family, and continuity of his career and 2) obligations to others, including: 1) his confidentiality agreement with Brown & Williamson to protect their secrets and 2) obligation to the society in terms of saving human lives and protecting human health. He had to make act, unexpectedly, within months of joining Brown & Williamson. The timing was awkward in that his daughter needed medical attention. Costs, both financial and otherwise, were high. Wigand made an informed choice after considerable contemplation.

CONCLUSION

In this article we have described a set of paradigms to aid the understanding of three different paradigms of decision-making behavior found in the health care industry. After differentiating between a paradigm and a model, we described paradigms of ethical, courageous, and virtuous decision-making behavior. Innovations and discoveries in health care have created ethical dilemmas never before encountered. The regulatory ethics paradigm, which has effectively guided ethical decision-making behavior of health care professionals, is now being questioned. New advances, as in the case of surgical techniques, show inadequacies of this paradigm. Greater degree of freedom for decision making by the physician is being demanded.

Courageous decision-making behavior of health care professionals has not been adequately studied. However, it has been found that becoming courageous is a lifelong process. Courageous behavior among patients with long-term health develops over time after full acceptance of reality by the patient. The patients are then able to manage their problems based on this discernment of reality, pushing themselves beyond coping to be productive and responsible.

Opportunities to manifest virtuous decision-making behavior arise when a decision maker must unexpectedly choose from alternative courses of action, of

which 1) one benefits the decision maker's self-interest, satisfying his or her individual obligations with relatively low associated cost or risks borne by the decision maker and 2) the other is in the interest of others, raising the issue of the decision maker's social responsibility, often with relatively high associated costs or risks borne by the decision maker. The decision maker must recognize the opportunity and then make an informed choice.

REFERENCES

1. Beauchamp, T.L.; Childress, J.F. *Principles of Biomedical Ethics*, 2nd Ed.; Oxford University Press: New York, 1983.
2. Smith, W.J. *The Culture of Death: The Assault on Medical Ethics in America*; Encounter Books: San Francisco, 2001.
3. Baker, R.B.; Caplan, A.L.; Emanuel, L.L.; Latham, S.R. *The American Medical Ethics Revolution: How the AMA's Code of Ethics Has Transformed Physicians' Relationships to Patients, Professionals, and Society*; Johns Hopkins University Press: Baltimore, MD, 2001.
4. Epstein, E.M. Business ethics and corporate social policy. *Bus. Soc.* **1998**, *37* (1), 7–40.
5. Bates, T. Ethics of consent to surgical treatment. *Br. J. Surg.* **2001**, *88* (10), 1283–1284.
6. Weijer, C. I need a placebo like I need a hole in the head. *J. Law Med. Ethics* **2002**, *30*, 69–72.
7. Agich, G.J. Ethics and innovation in medicine. *J. Med. Ethics* **2001**, *27* (5), 295–296.
8. Finfgeld, D.L. Courage as a process of pushing beyond the struggle. *Qual. Health Res.* **1999**, *9* (6), 803–815.
9. Rawnsley, M.M. Recurrence of cancer: a crisis of courage. *Cancer Nurs.* **1994**, *17*, 342–347.
10. Haitch, R. How Tillich and Kohut both find courage in faith. *Pastor. Psychol.* **1995**, *44* (2), 83–97.
11. Deutsch, M. Courage as a concept in social psychology. *J. Soc. Psychol.* **1961**, *55*, 49–58.
12. Rorty, A.O. The two faces of courage. *Philosophy* **1986**, *61*, 151–171.
13. Walton, D.N. *Courage: A Philosophical Investigation*; University of California Press: Berkeley, CA, 1986.
14. Cuff, W.T. The Experience of Courage and the Characteristics of Courageous People. Dissertation Abstracts International, University of Minnesota, 1993; Vol. 53, 5408B. Doctoral dissertation.
15. Dhir, K.S. A paradigm for the study of virtuous decision making behavior. Proceedings of the 2002 Western Decision Sciences Institute Conference, Las Vegas, NV, April 2–5, 2002; Dhir, K.S., Ed.; Decision Sciences Institute, Georgia State University: Atlanta, GA, 2002; 94–99.
16. Pellegrino, E.D. Toward a virtue-based normative ethics for the health professions. *Kentucky Inst. Ethics J.* **1995**, *5*, 253–277.

Health Care Finance in the Twenty-First Century

Judith J. Kirchhoff

MPA Program–SBPAIS, Long Island University, University Plaza, Brooklyn, New York, U.S.A.

INTRODUCTION

In the early years of the twenty-first century, it is fair to say that changes in the financing and delivery of health care services have led to a health care delivery system far different from the fragmented delivery by independent, unconnected service units characterizing most of the twentieth century. A health delivery system or integrated network offering a continuum of prepaid managed medical care is now the norm, covering 75% of the U.S. insured population.^[1] The term “managed care” has come to refer to organized networks of service delivery providing or arranging provision of a continuum of services to a defined population.^[2] In organizational terms, a managed care system “seeks vertical integration,”^[1] a state of control by administrative decision.

The shift to widespread adoption of managed health care in the United States can be traced to 1973, when the U.S. federal government promoted competition as a method to slow health care cost inflation deriving from the launch of Medicare and Medicaid in 1967.^[3] Under the 1973 law, widely referred to as the HMO Act,^[4] federally qualified health maintenance organizations (HMOs) could require employers with more than 25 employees to offer an HMO option. This gave employers, and their employees, choices of health plans beyond the traditional indemnity insurance that had prevailed until then.

Growth of managed health plans initially was slow but gained momentum as the 1970s gave way to the 1980s and 1990s.^[5] More health plans drew members away from indemnity insurance plans using lower premiums and broader coverage as enticements. Managed health plans were able to lower member premiums by shifting from provider-set fees to negotiated contracts involving volume discounts.

At the same time, continuing escalation of health care costs led to higher numbers of uninsured or underinsured citizens, 40 million citizens on average in the mid-1990s.^[6] This number is much smaller than the 54 million people who lacked insurance coverage for at least one month (1993) and the 67 million who lacked coverage at some point during the two-year period 1992–1993.^[5]

Because the average number of citizens without insurance increased to an estimated 44 million by

1999,^[7] it is likely that the dynamic numbers have increased as well. It should be noted that underinsurance involves, to the extent permitted by law, exclusions for preexisting conditions, services not covered, absence of catastrophic coverage, insurance deductibles and copayments, gaps in medicare coverage, and lack of long-term care coverage.^[5] One particularly problematic form of underinsurance occurs for individuals who reach benefit limits while still needing care;^[8] such individuals often become dependent on government health care and income support, if the condition is debilitating.

In the early 1990s, the Clinton Administration proposed universal health care coverage through a complex system of “managed competition” to control costs.^[9,10] Though the Clinton proposal failed to become legislation, it energized the private sector into aggressive pursuit of alternatives, resulting in a pronounced shift to managed health plans.^[11]

Health care finance at the dawn of the twenty-first century can be understood in terms of decision choices of participants in the health care system and payment mechanism incentives.

PARTICIPATION DECISIONS AND FINANCING

Health plans are entities providing health care services at a fixed monthly premium for a specific period of time. Employees join health plans through their employers, who pay a fixed premium per employee for a specified benefits package for a specific period of time, usually one year. The health plan selects providers and negotiates price, volume, and payment method, which may vary depending on the type of provider. The health plan “manages” members and delivery of care using administrative tools and analytical processes. Providers include primary care physicians and specialists, inpatient and outpatient services, labs and imaging centers, pharmacies, therapists, and home health organizations.^[12] Patients are health plan members who access care through in-plan providers or, at extra personal cost, out-of-plan providers.

Employers are the dominant purchasers of health care coverage for their employees.^[6] Taxpayers, through government purchasers, finance Medicare for

the elderly and other eligible patients, Medicaid for the eligible poor, care for the Armed Forces and their dependents, eligible veterans, American Indians, children and the working poor, and other eligible populations.^[13] More recently, the 1997 federal State Children's Health Insurance Program (SCHIP) provided strong incentives to states to expand health care coverage for children in families with incomes of less than 200% of the U.S. poverty level but higher than Medicaid eligibility levels.^[14] By the end of 2001, the program provided coverage for three million children.

In transactions with health plans, employers are motivated to keep their expenditures for health care coverage low, choosing health plans offering benefits that attract and keep good employees while minimizing employer health care expenses. Some employers drop coverage for their employees if premiums increase too much.^[15] Employees participate in health care coverage costs to the degree of their contributions to premium payments and the amount of co-pays to health providers.

For example, a relatively young and healthy single employee may perceive little need for health care coverage and participate only to the degree required, effectively gambling on remaining well and accident-free. Other employees may have conditions requiring continuous or expensive medical care, or family members with such conditions. These employees know that they will use more benefits than they pay for, and are grateful that they have coverage at all.

Government purchasers, because they are purchasing for large numbers, have great leverage in negotiating prices. Indeed, the federal government prospectively sets prices as one strategy for estimating cost outlays. The ability to estimate is limited, because Medicare and Medicaid are entitlement programs for their eligible populations, meaning all who are eligible may enroll, and all who enroll are entitled to all medically indicated approved services.^[16] Thus, demand for approved services ultimately determines government expenditures for these programs in any given year.

Health plans seek profitability, that is, revenues greater than expenditures. Health plan "products" are benefit packages sold for the "monthly premiums" that employers pay and health plans use to pay providers and other operating costs. More generous benefit packages carry higher potential costs for the health plan and require higher premiums. Federal and state mandates require some types of coverage, and others are left to the health plan's discretion. Health plans calculate annual premiums using actuarial data, i.e., members' demographic characteristics and historical health care utilization data. Plan profitability depends in part on skilled forecasting of members' service utilization and in part on setting premium rates high enough to be profitable without discouraging membership in the plans.

In order to meet these goals, health plans seek control of plan members' health care utilization decisions through financial incentives, for example, co-payments and charges for out-of-network care. If the plan member seeks in-plan provider network care, one co-payment applies. Out-of-network providers carry a higher co-payment. Staff model health plans restrict access to staff providers and others only as authorized by the staff providers.^[1]

Changes in health plan membership take place annually during an open enrollment period. Employers may elect to discontinue coverage for all employees if premium costs increase too much, a choice many have exercised in recent years. Or, a large employer may move to self-funding health care costs, that is, become self-insured. Governments cannot eliminate health benefits to eligible recipients without legislative authorization. Plan members can change health plans but may not withdraw from at least minimum coverage by their employer. Health plans may elect not to offer coverage to an employer or membership categories, for example, the elderly, as some health plans have opted to do. In addition to federal regulation, many states regulate health plan choices about offering and withdrawing coverage, among other practices.^[4]

Employers and employees, governments, and health plans also can seek to limit their risks by limiting benefits offered (services covered), shifting more costs to others, or, in the case of governments, by limiting eligibility for the entitlement. Health plans also have the option, with constraints already mentioned, of limiting benefits or raising premium rates, both at the risk of encouraging employers to respond by self-insuring, searching for less costly health plans, or eliminating coverage. Employee options are limited by employer decisions about coverage. The employee may seek another employer, buy additional coverage through an outside source, or, if a spouse also has employer-sponsored coverage, elect to be covered under the spouse's plan.

MANAGED CARE FINANCIAL INCENTIVES

The financial incentives used by health plans profoundly affect the way health care services are delivered. The goal of the health plan in these transactions, as already noted, is profitability, which depends in part on the ability of the health plan to control expenditures to providers and internal operating costs. Controlling expenditures to providers requires managing utilization of care by plan members and delivery of care by providers.^[17] In their efforts to manage providers and patients, health plans employ incentives designed to reduce the amount, frequency, and intensity of services delivered by providers and used

by patients^[18] and to shift risk to members and providers.

Financial incentives vary by provider type. Capitated rates are most frequently used with primary care physicians (PCPs).^[19] PCPs are “gatekeeper” physicians who control access to the rest of the medical care system. Capitation pays per member per month fees for each enrollee registered with a provider. The capitated payment is designed to cover medical care within the practitioner’s scope of practice and often includes routine lab tests and specialist referrals. Capitated payments make provider forecasts of plan member utilization extremely important.

For care outside the PCPs scope of practice, PCPs refer to specialists under another financial mechanism that often links additional PCP revenue to number of referrals. A percentage of PCP capitation revenue may be “withheld” to create a pool of funds used to pay specialists.^[19] Alternatively, the cost of specialists may come directly out of the PCP’s capitation income.^[19]

In the first case, whatever is “saved” by minimizing the use of specialists is shared by the plan’s PCPs at the end of the contract year. Those with fewer referrals often earn the largest share of the pool. Some states regulate such revenue-sharing practices. In the second case, the PCP may make significant “profit” if many enrollees seek no care during the year, but the PCP is at risk of substantial losses if some patients have high specialty medicine costs. PCPs must carry stop-loss insurance to cover for such situations.^[20]

Some services are “carved out” of the capitation payment system, as are certain kinds of care, for example, mental health services^[19] and expensive but relatively infrequent procedures like neurosurgery and cardiac surgery.

Capitation rates base health plan expenditures and PCP revenues on number of enrollees. Health plans can forecast expenditures, and physicians can forecast their annual revenues from the plan based on the number of plan members they serve.^[20] Total physician revenues will be calculated from the number of plans accepted and the number of enrollees from each plan. Total health plan expenditures to each physician also are settled for the contract year based on members enrolled in the PCP’s practice. Members can change PCPs freely within a health plan’s provider list, and health plans manage such changes carefully, because plans do not want to pay two PCPs for the member who changes from one PCP to another. The capitated payment follows the member from the “losing” physician to the “gaining” physician.

Discounted fees and prospectively set case rates are contract payment mechanisms commonly used for hospitals and other providers. Discounted fees simply pay a set amount for the service. For example, a physical

therapist may charge \$90 for a service unit, and the health plan may pay \$60. The economic incentive for the physical therapist is to shorten the service unit time of \$60 patients. A case rate bundles services into an identifiable episode of care reimbursed at a flat fee. Medicare’s diagnostic-related groups are perhaps the most familiar type of case rate^[21] for inpatient care.

For outpatient care, Resource Based Relative Value Units are a form of case rate.^[19] Discount and case rate contracts may involve negotiations between payers and providers over the amount of the discount from charges (providers’ regular price) and assumptions used in setting case rates or global fees for specified categories of members or care episodes.

Revenues are more difficult to forecast with discounted fees and case rates, because they are based on service units rather than a fixed number of patients or episodes of care. Case rates incorporate assumptions about episode duration and procedures.^[19]

Health plan internal operations are controlled by executive decisions and management of employees. Internal operations of health plans constitute administrative costs. Health plans have marketing and sales functions, processing and paying enrollee medical care claims functions, and provider and patient relations functions. Marketing and sales functions are related to selling benefits packages to employers, medical care claims are a payment function, and provider and patient relations are new health plan management functions associated with controlling the amount of care provided.

Health plan operations management includes developing provider networks to deliver medical services to enrollees; negotiating prices with provider representatives or providers; precertifying procedures; coordinating benefits for enrollees with separate coverage; managing individual episodes of care; reviewing patient utilization, physician diagnosis and treatment patterns, and hospital lengths of stay; encouraging enrollees to undergo preventive tests; ensuring good relationships with providers and enrollees; and monitoring provider and patient satisfaction.^[22]

These activities are replicated in provider organizations to fulfill health plan reporting requirements and to generate management information for complex contract decisions.^[23] This adds administrative costs to providers’ operations. At the same time, federal and state governments add regulatory complexity, requiring physicians and hospitals to provide more and better documentation to prevent fraud and ensure quality care at reasonable cost.^[24] Additionally, increasing amounts of quality information about health plans is available to assist employers in making coverage decisions, but studies show that it is rarely used.^[25] All of these functions add administrative costs to health care service delivery in the United States, costs incorporated

into the sales, premium, expenditure cycle described herein.

In summary, the health care system at the turn of the twenty-first century is far different from that which dominated for much of the twentieth century. The health care system is rapidly becoming organized and managed by profit-making health care plans using financial incentives that spread costs and risks among the participants. At this point in time, however, it cannot be said that managed health care is less costly or more efficient. After a short period of declining inflation in health care costs, costs to all participants are on the rise again.

REFERENCES

- Folland, S.; Goodman, A.; Stano, M. Managed care. In *The Economics of Health and Health Care*; Prentice-Hall: Upper Saddle River, NJ, 2001; 248–280.
- Kovner, A.R. Health maintenance organizations and managed care. In *Jonas and Kovner's Health Delivery in the United States*, 6th Ed.; Kovner, A.R., Jonas, S., Eds.; Springer: New York, 2001; 279–306.
- Patel, K.; Rushefsky, M.E. Health care policy in the United States. In *Health Care Politics and Policy in the United States*, 2nd Ed.; M. E. Sharpe: Armonk, NY, 1999; 25–54.
- Pozgar, G.D. Managed care and organizational restructuring. In *Legal Aspects of Health Care Administration*; Aspen: Gaithersburg, MD, 1996; 615–653.
- Bodenheimer, T.S.; Grumbach, K. Access to health care. In *Understanding Health Care Policy, A Clinical Approach*; McGraw Hill: New York, 2002; 15–32.
- Folland, S.; Goodman, A.; Stano, M. The organization of health insurance markets. In *The Economics of Health and Health Care*; Prentice-Hall: Upper Saddle River, NJ, 2001; 227–247.
- Fershein, J.; Sandy, L.G. The changing approach to managed care. In *To Improve Health and Health Care 2001*; Isaacs, S.L., Knickman, J.R., Eds.; Robert Wood Johnson Foundation: Princeton, NJ, 2001; 77–100.
- Billings, J. Access to health care services. In *Jonas and Kovner's Health Delivery in the United States*, 6th Ed.; Kovner, A.R., Jonas, S., Eds.; Springer: New York, 2001; 401–438.
- Maxwell, J.; Temin, P. Managed competition versus industrial purchasing of health care among the Fortune 500. *J. Health Polit. Policy Law* **2002**, 27 (1), 5–30.
- Starr, P. *The Logic of Health Care Reform*; Penguin Group Publications: New York, 1994.
- Loomis, C.J. The real action in health care. *Fortune* **1994**, July 11, 149–154.
- Kovner, C.; Salsberg, E.S. The health care work force. In *Jonas and Kovner's Health Delivery in the United States*, 6th Ed.; Kovner, A.R., Jonas, S., Eds.; Springer: New York, 2001; 64–115.
- Patel, K.; Rushefsky, M.E. Health care policy in the United States. In *Health Care Politics and Policy in America*; M. E. Sharpe: Armonk, NY, 1999; 25–54.
- Bodenheimer, T.S.; Grumbach, K. Paying for health care. In *Understanding Health Care Policy, A Clinical Approach*; McGraw Hill: New York, 2002; 5–14.
- Thorpe, K.E.; Knickman, J.R. Financing for health care. In *Jonas and Kovner's Health Care Delivery in the United States*, 6th Ed.; Springer: New York, 2001; 32–63.
- Patel, K.; Rushefsky, M.E. Medicaid health care for the poor. In *Health Care Politics and Policy in America*; M. E. Sharpe: Armonk, NY, 1999; 55–94.
- Shimshak, D.G.; DeFuria, M.C.; DiGiorgio, J.J.; Getson, J. Controlling disenrollment in health maintenance organizations. In *Managed Care, Strategies, Networks and Management*; Brown, M., Ed.; Aspen: Gaithersburg, MD, 1994; 231–238.
- Bodenheimer, T.S.; Grumbach, K. Mechanisms for controlling costs. In *Understanding Health Care Policy, A Clinical Approach*; McGraw Hill: New York, 2002; 89–103.
- Zelman, W.N.; McCue, M.J.; Millikan, A.R. Alternative payment systems and provider cost. In *Financial Management of Health Care Organizations*; Blackwell: Malden, MA, 1998; 342–370.
- Bodenheimer, T.S.; Grumbach, K. Capitation payment in managed care. In *Understanding Health Care Policy, A Clinical Approach*; McGraw Hill: New York, 2002; 44–52.
- Patel, K.; Rushefsky, M.E. Health care cost containment. In *Health Care Politics and Policy in the America*; M. E. Sharpe: Armonk, NY, 1999; 162–195.
- Duncan, W.J.; Ginter, P.M.; Swayne, L.E. *Strategic Management of Health Care Organizations*, 2nd Ed.; Blackwell: Cambridge, MA, 1995.
- Bodenheimer, T.S.; Grumbach, K. Reimbursing health care providers. In *Understanding Health Care Policy, A Clinical Approach*; McGraw Hill: New York, 2002; 33–43.
- Sparrow, M. *License to Steal: Why Fraud Plagues America's Health Care System*; Westview Press: Boulder, CO, 1996.
- Patel, K.; Rushefsky, M.E. The triumph of incrementalism. In *Health Care Politics and Policy in America*; M. E. Sharpe: Armonk, NY, 1999; 292–339.

Health Care Organizations, Managing

Grant T. Savage

*Department of Management and Marketing, University of Alabama,
Tuscaloosa, Alabama, U.S.A.*

Julie W. Robinson

*Veteran Affairs Medical Center (151C), Tuscaloosa Research and Education Advancement
Corporation, Tuscaloosa, Alabama, U.S.A.*

INTRODUCTION

The United States has experienced major changes in the way health care is delivered and coordinated, making the task of managing these organizations both complex and uncertain. Therefore, top managers need not only the traditional skills of communication, leadership, planning, and organizing, but also conflict, stakeholder, and strategic management skills. This article provides a current snapshot of the health care industry, trace its history, and note current trends in health care management. It concludes by noting the need to expand managerial competencies, as well as the vision of what health care organizations can accomplish.

A CURRENT SNAPSHOT OF THE HEALTH CARE INDUSTRY

Hospitals are no longer the dominant centers for health care delivery. Ambulatory, home health, and long-term care facilities represent significant rivals. Moreover, managed care organizations, for-profit corporations, and the government exert significant influence on the industry. Ambulatory care services have grown and expanded with advances in medical technology. These advances also are making it possible for home health agencies to provide care to more patients. At the same time, managed care and government programs have significantly reduced reimbursements, and hospitals have integrated with other care providers, creating organized delivery systems and networks. Long-term care has also been transformed, with assisted living centers having been built at a rapid pace. Fig. 1 shows the major players that influence the health service delivery sector.

Buyers

Employers, state and federal governments, and insurers are buyers of health services. Because of cost

containment pressures from buyers, health care providers have reduced staff and sought economies of scale through mergers and acquisitions.^[1] Addressing patient safety and quality expectations of buyers is a major challenge for health care providers.

Suppliers

Suppliers include the pharmaceutical, medical, and durable goods and devices that enable health care organizations to provide services. Suppliers are declining in number. Competition, rising costs, and consolidation have encouraged this trend.^[1]

Substitutes

Alternative and complementary medicine is gaining greater acceptance and use by health care consumers. Complementary and alternative medicine is defined as “diagnosis, treatment and/or prevention which complements mainstream medicine by contributing to a common whole, by satisfying a demand not met by orthodoxy or by diversifying the conceptual frameworks of medicine.”^[2] Many people have turned to alternative or complementary treatments such as acupuncture, chiropractic, and herbal remedies out of frustration with traditional medicine.

New Entrants

Long-term care facilities—in particular, assisted living centers and web-enabled home health—are just two of the emerging entrants in the health care industry. Assisted living centers provide a combination of housing, support services, and limited health care. As of 2000, more than 1 million Americans were living in assisted living residences.^[3] Assisted living, a consumer-driven sector of the health care industry, appeals to the healthiest segment of those in long-term care, with the sickest and least capable patients concentrated in nursing homes. Web-enabled home

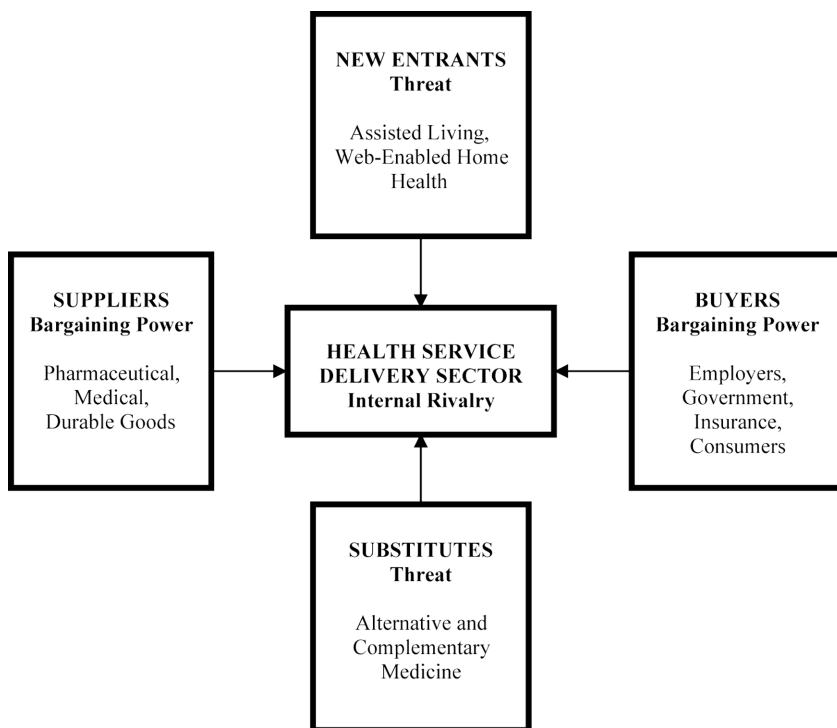


Fig. 1 Market forces model.

health is further accentuating this trend, as telemedicine allows chronically ill patients to be monitored at a distance.

BRIEF HISTORY OF HEALTH CARE ORGANIZATION AND MANAGEMENT

Before the advent of hospitals, health care was a private matter, delivered in the patient's home. The services of private physicians, midwives, and nurses were obtained when needed and financially feasible. Chemists sold medications or "tonics." Health care has since centralized, decentralized, and is now integrated (Table 1).

Centralization of Care

Hospitals in the United States began with the almshouses and pesthouses of the 1700s.^[4] Local governments established these facilities to feed and shelter the orphaned, homeless, elderly, disabled, and chronically or mentally ill, and they provided health care as a secondary function. Hospitalization carried with it the stigma of disgrace.

During the Industrial Revolution in the United States, medical technology and training and the desire to standardize medical education all aided in the demand for and the subsequent development of hospitals. As medical technology, such as anesthesia

and the X-ray, grew so did the demand by physicians for facilities with the capabilities to practice these advanced techniques.^[4] Neither the physician's home nor the home of a patient was now equipped to diagnose and practice medicine.

Along with medical advancements came the need to standardize medical education and training. In 1910, Abraham Flexner led a study of medical education. The Flexner Report sparked systematic efforts to standardize medical education.^[4] Apprenticeships with practitioners gave way to hands-on experience in a facility with the capabilities to train students in the most advanced medical technology. Hospitals became not only places of care, but also teaching, research, and training facilities. During the early twentieth century, most administrators were physicians or clergymen who often followed the preferences of local physicians. Most hospitals centralized power at the executive level, much like a university. The competencies necessary to manage included internally focused communication, leadership, organizing, and planning skills.

World War II contributed to the development of hospitals as not-for-profit institutions. Americans began organizing care for mentally and physically wounded veterans. The number of skilled and technical workers increased in hospitals, and management's role became one of creating an environment where health care professionals could do their job, as opposed to controlling the workplace.^[5] Moreover, employers began paying for health insurance. The

Table 1 Organizational phases

Organizational phase	Period		
Centralization of care	1850 to 1899	1900 to 1964	1965 to 1970
	• Development of the Pennsylvania Hospital • Growth of medical and caring professions • Introduction of public health	• Beginning of advances in medical technology, training, knowledge and standardization of medical education • Formation of insurance plans • Hill-Burton Act	• Creation and expansion of Medicare and Medicaid • Growth in health services and medical specialization
Decentralization of care	1971 to 1980	1983 to 1985	1986 to 1990
	• Shift from inpatient to outpatient care begins • Beginning of cost containment efforts by insurers and government • For-profit hospital chains emerge	• Medicare prospective payment system enacted and implemented • Managed care gains market share • Growth in joint ventures between hospitals and physicians	• Freestanding ambulatory care facilities double • Many rural and secondary hospitals close • Hospitals diversify to counter market competition
Integration of care	1990s	1996	1999
	• Integrated Delivery Systems emerge as a result of consolidation, mergers, and acquisitions	• Over 55% of U.S. hospitals report being in an alliance	• 72% of hospitals reported being in an alliance

federal government had placed wage and price controls on war industries, so employers used health coverage for recruitment and retention.^[6]

Direct federal involvement in hospitals began in 1947. The Hill-Burton Act was intended to fund the construction of hospitals in rural areas, but amendments extended it to provide grants that matched the funds generated by a community.^[4] The federal government's involvement continued to grow in the mid-1960s with the creation of Medicare and Medicaid. These programs, along with employer-sponsored health insurance, increased the demand for hospital-based health services. Top managers of hospitals were referred to as administrators, and often had either clinical or accounting backgrounds.^[7] By the 1970s, hospitals were the center of medical activity, and the administrator's duties had become increasingly complex. Managerial competencies included not only communication, leadership, organizing, and planning skills, but also increasingly skills in conflict management and external relations.

Decentralization of Care

The 1980s and 1990s gave rise to the professionalization of health care management. Hospital administrators

were now referred to as chief executive officers (CEOs). The majority of CEOs were educated—as they are today—at the master's degree level in health services administration.^[8] A shift from inpatient care to outpatient care also occurred during the 1980s. Reimbursement incentives, continued medical advances, and the introduction of managed care were three factors that prompted the shift.

The enactment of the Medicare prospective payment system (PPS) in 1983 resulted in hospitals being paid a “predetermined amount for Medicare inpatients based on diagnosis-related groups (DRGs).”^[9] The PPS encouraged hospitals, physicians, and health care entrepreneurs to enter the ambulatory care arena. Ambulatory care takes place in various locales including urgent care centers, physician offices, hospital emergency rooms, and patients' homes. Increasingly, ambulatory care services are often the patient's first point of contact in the health care delivery system.^[10] This new form of medical practice required a new style of management practice. The traditional administrator who managed internal affairs now had to focus outward and compete for patients based on quality *and* on cost. Business skills gained currency, as assuming financial risk for the care provided, contracting with physicians, and marketing services were added to the already complex duties of the health care manager.^[11]

Moreover, medical technology permitted sophisticated procedures to be delivered in ambulatory rather than inpatient facilities.^[12] Between 1980 and 1990, the American Hospital Association reported that the number of freestanding ambulatory care facilities doubled.^[10] Many diagnostic and therapeutic procedures, as well as treatment, do not necessitate overnight stays, which plays to patients' preferences for accessible, convenient, and prompt care.

Finally, the growth of managed care expanded the ambulatory care sector. The financial incentive to provide care in an outpatient setting was a natural fit with the managed care philosophy of using gatekeepers to reduce costly hospitalization. As health care consumers continue to be vocal in their demands and medical technology continues to advance, the market for ambulatory care will increase.

Integration of Care

The 1990s were a decade of vertical and horizontal integration in the health care industry. To integrate in the health care industry means to bring together practitioners, hospitals, health plans, employers, and patients.^[13] Under this strategy, repetitive services are eliminated and the expense of technology is consolidated in an effort to streamline health care delivery. The increasing cost of delivering health care and patient's demands for convenient "one-stop shopping" were two of the drivers for integrated delivery systems (IDSs). In 1996, over 55% of America's hospitals were reported to be in an alliance;^[14] in 1999, that number rose to 72%.^[15]

Hospitals first ventured into horizontal integration. *Horizontal* integration consists of alliances between separate hospitals in an effort to become more efficient by creating economies of scale. *Vertical* integration is the merger and/or acquisition of health care providers at different positions along the continuum of care to improve efficiency and coordinate care across the entire organization.^[14] An IDS combines vertical and horizontal integration so managers have accountability at every level.^[5] Chief executive officers, departmental managers, and entry-level managers all must manage the expectations of stakeholders. Significantly, management competencies expanded from managing an organization to managing a network of services and alliances.^[16]

CURRENT TRENDS IN HEALTH CARE

The delivery of health services will continue to be buffeted by change (Fig. 2). Forces for change include advances in information and medical technology,

increased governmental regulation, growing consumerism, and aging demographics. Health care executives will have to manage conflicting values as they face pressures to contain costs while maintaining access and improving quality. Management trends addressing cost and quality concerns include the use of continuous quality improvement (CQI), population health management, and information technology.

CQI

Less than stellar results with static, quality assurance programs led to the health care industry's experimentation with total quality management (TQM), typically called CQI. Continuous quality improvement builds on traditional assurance methods, emphasizing all three of Donabedian's criteria: structure, process, and outcomes. CQI looks for improvement *opportunities*. Three practices distinguish CQI from previous quality assurance models: consumer-focus, systems thinking, and statistical quality control methods.

Continuous quality improvement recognizes the rise in health care consumerism.^[17] As consumers are becoming more knowledgeable about their health and health care delivery, they are demanding better care, better information, and better coordination of their care. The definition of quality of care has broadened in response to these social changes and now includes patient-provider interaction, the setting in which care is given, and quality of life.^[18]

"Customers," in CQI terms, are those people affected by a product or process, including both internal and external customers.^[19] External customers are patients, benefactors, payers, or purchasers of health care services. Health care providers, their colleagues, or other departments and services within the organization are internal customers. Continuous quality improvement encourages the development and use of methods to discover consumer needs and expectations.

"Systems thinking" recognizes that processes are interrelated.^[19] Continuous quality improvement avoids personal blame and focuses on the managerial and professional processes associated with an outcome.^[17] Continuous quality improvement involves each person associated with the process in the problem identification or quality opportunity phase and the actual redesign process. It encourages each member of the organization to see the process through to the end and consider the "big picture."

Quality circles and improvement teams that were popular in the 1970s and 1980s encouraged the use of statistical tools, but CQI requires these fact-based decision-making tools.^[17] Employees as well as managers, study possible causes and brainstorm

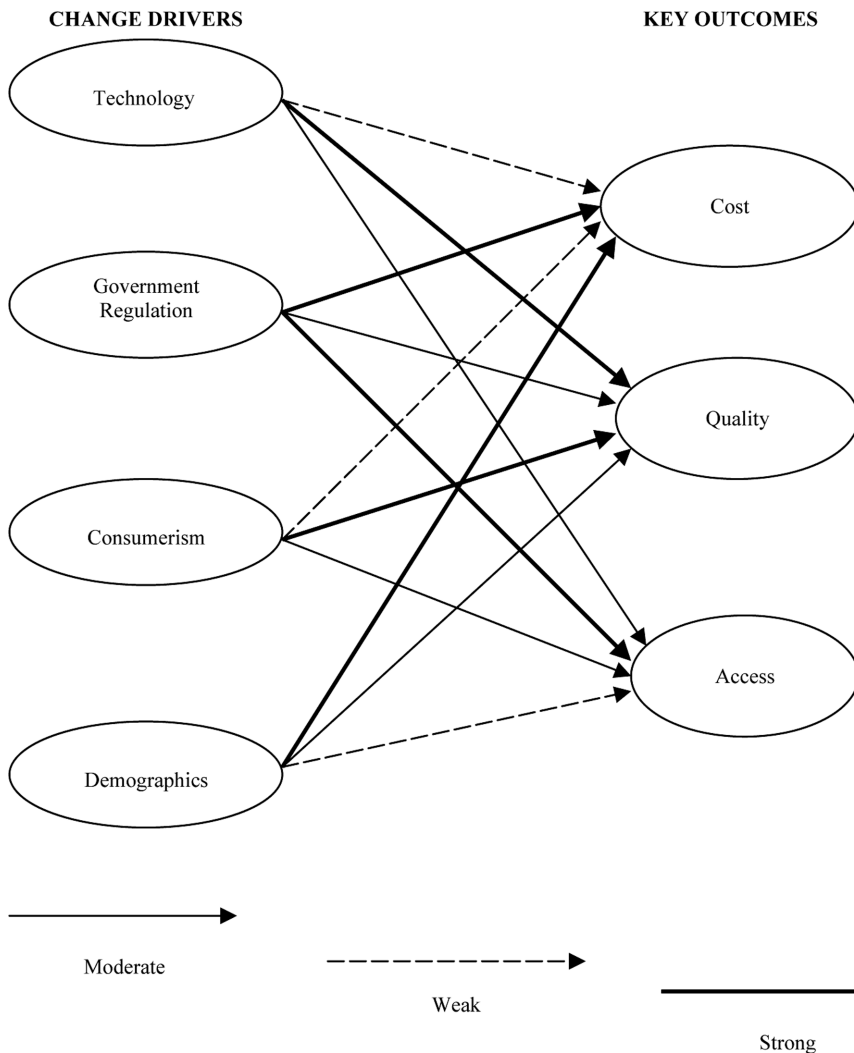


Fig. 2 Factors for change. (From Ref.^[1].)

systemwide solutions, requiring personnel educated in statistical analysis and control processes. “The use of data-based management and scientific principles to the clinical and administrative processes that produce patient care is what CQI is all about.”^[17]

Population Health Management

Population health management refers to the change in focus from caring for individual illnesses to encouraging the wellness and health of a population. An ideal system would deliver comprehensive and personalized preventive through tertiary care, while guiding and tracking patients through a continuum of services.^[14] Some health systems are now providing primary through tertiary care, as well as home and long-term care. More specifically, population health management relies on the evidence-based practices of catastrophic care, demand, disease, disability, and lifestyle.^[21]

These five forms of population health management have varying foci and benefit different stakeholders. Lifestyle management emphasizes reducing health risks and preventing illnesses by targeting a relatively healthy population, such as young employees. Demand management concentrates on consumer requests and inquiries for medical care services, supplying advice, counseling, and referrals to appropriate providers, thus lowering the costs and improving the efficiency of care seeking. Disease management focuses on individuals with chronic conditions, such as diabetes, congestive heart failure, or asthma, coordinating the services needed for these specific diseases. Catastrophic care management centers on providing health services for individuals with catastrophic illnesses or injuries, such as acquired immune deficiency syndrome or stroke patients. Disability management focuses on reducing injuries, preventing job-related illnesses, and managing employee disabilities. All health population management approaches add value by reducing costs and improving the quality of care.^[21]

Information Technology

Information technology (IT) is a valuable resource for CQI and population health management efforts. Information technology provides the foundation for the flow, storage, and analysis of data, converting it into usable information for making operational and strategic decisions.

More important, IT has the potential to radically transform the health care industry, improving quality and access while minimizing costs. Telemedicine demonstrates IT's benefits. Radiology, for example, is a common application within telemedicine that Medicare now covers. The managed care industry quickly saw the opportunity to contain costs through IT. By deploying IT, health service organizations can improve clinical outcomes, ensure patient safety, and streamline administrative tasks. However, without industrywide standards, the full benefits of IT cannot be realized.^[20]

Indeed, both population health management and CQI require a robust information infrastructure.^[22,23] Without an information infrastructure that connects all providers, insurers, third-party administrators, and employer/payers, effective disease and other population health management measures are much less efficient. At the same time, the coordination of care that results from establishing health networks is also more effective and efficient when an information infrastructure connects primary, secondary, and tertiary providers within and across communities.

CONCLUSION

Health care managers face the challenge of developing and managing delivery organizations that are internally complex, and involve external networks of other providers. Health care executives must manage conflicting values among internal stakeholders, while also managing complex strategic alliance relationships. Relational and strategic competencies, as well as abilities in managing technology, quality improvement, and population health are now prerequisites for health system CEOs. As health system managers develop information infrastructures, manage population health, and improve internal quality processes and outcomes, they begin to achieve a community health care management focus. The challenge for health care executives is both to articulate a vision, and to master the skills needed for managing these complex and emergent organizational forms.

REFERENCES

1. Savage, G.T.; Campbell, K.S.; Patman, T.; Nunnelley, L.L. Beyond managed costs. *Health Care Manage. Rev.* **2000**, *25* (1), 93–108.
2. Ernst, E. The role of complementary and alternative medicine. *Br. Med. J.* **2000**, *321*, 4.
3. Assisted Living Federation of America. **2002**. What is Assisted Living? [On-Line]. Available: <http://www.alfa.org/public/articles/details.cfm?id=96>.
4. Haglund, C.L.; Dowling, W.L. The hospital. In *Introduction to Health Services*, 4th Ed.; Williams, S.J., Torrens, P.R., Eds.; Delmar Publishers, Inc.: Albany, NY, 1993; 134–159.
5. Kovner, A.R.; Neuhauser, D. *Health Services Management*; Health Administration Press: Chicago, 2001.
6. Knight, W. *Managed Care*; Aspen Publishers: Gaithersburg, MD, 1998.
7. Wolper, L.F.; Pena, J.J. History of hospitals. In *Health Care Administration*, 3rd Ed.; Wolper, L.F., Ed.; Aspen Publishers: Gaithersburg, MD, 1999; 391–400.
8. McConnell, C. The changing face of health care management. *Health Care Manage.* **2000**, *18* (3), 1–17.
9. Hosking, J.E. Planning health care facilities and managing the development process. In *Health Care Administration*, 3rd Ed.; Wolper, L.F., Ed.; Aspen Publishers, Inc.: Gaithersburg, MD, 1999; 237–256.
10. McGuire, J.P. Ambulatory care services continue to grow. *Healthc. Financ. Manage.* **1994**, *48* (11), 10–12.
11. Nelson, R.A. The role of the medical group manager. In *Ambulatory Care Management and Practice*; Barnett, A.E., Mayer, G.G., Eds.; Aspen Publishers: Gaithersburg, MD, 1992.
12. Barr, K.W.; Breindel, C.L. Ambulatory care. In *Health Care Administration*, 3rd Ed.; Wolper, L.F., Ed.; Aspen Publishers: Gaithersburg, MD, 1999; 431–466.
13. McBride, S.H. Integration: Key to surviving managed care? *Dermatol. Times* **1995**, *6* (4), 43–48.
14. Barber, J.B. Evolution of an integrated health system: A life cycle framework. *J. Healthc. Manage.* **1998**, *43* (4), 359–377.
15. Friedman, L. Why integrated health networks have failed. *Frontiers of Health Services Management* **2001**, *17* (4), 3–54.
16. Savage, G.T.; Roboski, A.M. Integration as networks and systems: A strategic stakeholder analysis. In *Advances in Health Care Management*; Fottler, M.D., Savage, G.T., Blair, J.D., Eds.; JAI Press: New York, 2001; Vol. 2, 37–62.
17. McLaughlin, C.P.; Kaluzny, A.D. *Continuous Quality Improvement in Health Care*; Aspen Publications: Gaithersburg, MD, 1999.
18. Donabedian, A. Commentary on some studies of the quality of care. *Health Care Financ. Rev.* **1987**, December; Spec No: 75–85.
19. Kongstvedt, P.R. *Essentials of Managed Health Care*, 4th Ed.; Aspen Publishers: Gaithersburg, MD, 2001.
20. van der Reis, L. The idea behind information systems. *San Francisco Medicine* **1996**, *69*, 7.
21. McAlearney, A.S. Population health management in theory and practice. In *Advances in Health Care Management*; Savage, G.T., Blair, J.D., Fottler, M.D., Eds.; JAI Press: New York, 2002; Vol. 3, 117–158.
22. Feifer, C.; Ornstein, S.M.; Nietert, P.J.; Jenkins, R.G. System supports for chronic illness care and their relationship to clinical outcomes. *Top. Health Inf. Manage.* **2001**, *22* (2), 65–72.
23. Payton, F.C. How a community health information network is really used. *Commun. ACM* **1999**, *42* (12), 85–90.

Health Care Policy

Walter J. Jones

Medical University of South Carolina, Charleston, South Carolina, U.S.A.

INTRODUCTION

“Health policy” refers to the area of study that focuses on the outputs (products) and outcomes (final results) generated by a society that relate to the nature of the health services provided to the society’s population. The term also refers to the processes by which societies create these outputs and outcomes. This entry examines and analyzes the various issues related to health policy making and policies.

THE NATURE OF HEALTH POLICIES AND POLICY MAKING

All societies produce health policies. During the twentieth century, as medical science developed, societies began to refine their health care systems through policies created by governmental and private bodies. In most of Europe, this process was soon dominated by national governments. By the end of the twentieth century, most economically advanced nations had enacted health policies that provided for universal health insurance coverage for all citizens. In contrast, the United States enacted policies that resulted in a very well-funded, but relatively unorganized, health care system, with health insurance provided not by citizenship, but by employment status and, more recently, by age category (Medicare) and “qualifying poor” status (Medicaid).^[1]

THE HEALTH POLICY-MAKING PROCESS

Like other policy-making processes, health policy making is *cyclical*, not linear. It has no defined beginning or end. This also means that policy issues are never finally and completely resolved. Previous policy outputs and outcomes lead to further system inputs, which keep the process operating. In policy making, there are no final successes or failures.^[2]

The health policy-making process comprises four phases.

1. *Entry of Policy Inputs*: In this phase, policy-making institutions and individuals are exposed to stimuli for action coming from individuals,

groups, and institutions that comprise their relevant society. These inputs may include *demands* for policy making (or for opposition to proposed policy making), *supports* for previous policy making (including reelection to public office, or reappointment to an administrative or judicial position), *resources* for additional policy making (such as tax revenues), and *costs* incurred through previous policy making (such as budgetary debts, or the loss of political support as a result of previous policy decisions).

As a result of the relative impact of each input and their interactions, the *policy agenda* for the current cycle of policy making develops. The agenda includes the issues that policy makers will focus on in their current activities. Agenda setting is an important political skill for policy participants. It involves attracting public and policy maker attention through media publicity, published research, and other mechanisms. Some successful agenda settings are the result of such deliberative efforts (e.g., the efforts in the 1990s of groups such as ACT-UP to bring attention to the plight of people with HIV infection or AIDS). On the other hand, policy agendas are often set by the cumulative social results of previous policy making (for instance, the health care cost inflation in the United States mentioned below).^[3]

2. *Policy Formation*: Once policy agendas are set, policy makers work through existing institutions to enact preferred policy responses. In democracies such as the United States and the UK, this usually involves the interaction of executive, legislative, judicial, regulatory, and private sector institutions.

Some democracies, such as the United States, have policy institutions that were designed to prevent hasty or dictatorial policy making. They feature the separation of policy-making powers between governmental branches, the further division of policy-making powers between the national and state governments, and the potential review of much executive and legislative policy making by the judicial branch.^[4] Effective

policy making in some areas (such as tobacco industry liability) ends up in the courts, with lawyers representing the various sides dueling over one or more lawsuits.

Because of divided power, U.S. health policy making is often characterized by *incrementalism*, with prevailing policy making consisting of only modest and gradual adjustments of current policies. Reforms of large, complex, and popular programs such as Medicare and Medicaid usually do not involve radical change. On the other hand, many other democracies, such as the UK, have more unitary systems of government. With the British Parliament in fact controlled by a government with a strong Prime Minister, and only modest judicial review, it is easier for British policy makers to enact new laws.^[5] A countervailing advantage of the U.S. system may be flexibility and ease of experimentation. With some national programs such as Medicaid administered by the states with significant autonomy, it is possible for the United States to experiment with differing program coverages and methods of administration.

One other important feature of U.S. policy formation is the involvement of and, at times, control by private sector organizations. In the United States, slightly over half of health care spending comes from private sources, and most service deliveries are through private sector organizations. Consequently, private sector organizations have a significant amount of input into all health policy making. In some areas such as clinical certification and facilities accreditation, primarily private associations such as the Joint Commission for Accreditation of Health Organizations (JCAHO) have de facto control over far-reaching public policy functions.

3. *Policy Outputs:* Policy formation by the various institutions described above leads to policy outputs. Some outputs are very tangible. *Laws* are passed; P.L. 89-97, the Social Security Amendments, established Medicare for U.S. senior citizens in 1965, whereas the Balanced Budget of 1997 substantially lowered Medicare hospital reimbursement to curb rapidly increasing Federal Government expenditures. Laws can also come in the form of *subsidies* to relevant individuals or institutions; in 1990, P.L. 101-381, the Ryan White Comprehensive AIDS Resources Emergency Act, provided AIDS treatment funding for states and metropolitan areas heavily affected by the HIV-AIDS epidemic. Laws result in money and/or legal authority being allocated to other government bodies (federal,

state, or local), organized interest groups, families, and/or individuals.

In contrast, some policy outputs result from subsequent *regulatory decisions* rather than executive/legislative branch deliberations by elected policy makers. The Food and Drug Administration (FDA) effectively makes policy when it decides to withhold or withdraw medical products or devices from the market (as it did when it banned the use of the drug thalidomide in the 1960s). However, regulatory agencies derive their authority from executive and legislative branch mandates. At times, if their decisions create a sufficient public outcry, the other branches of government may act to rescind them. For example, because of public opposition, Congress overrode the FDA ban on saccharine in the 1970s.^[6]

In a similar fashion, policy making may flow directly from *judicial decisions*. A gridlock between the legislative and executive branches over tobacco policy was partially broken in the mid-1990s when public health lawyers and state attorney generals began to successfully sue tobacco companies in state courts over liability for Medicaid spending on tobacco-related illnesses. The hundreds of billions of dollars involved in the various suits prompted both sides to begin developing a nationwide settlement that would preempt further state legal action.

More nebulous, but very real, are variations in policy making that have been termed *symbolic politics* and *non-decision making*. Most policies have at least some symbolical intent, as the titles of some legislations suggest. For example, the 1996 law that ended the linkage between Medicaid and the primary public assistance program for families with children (Aid to Families with Dependent Children, or AFDC) was named the Personal Responsibility and Work Opportunity Act. This made the public point that there would be no more “free ride” for welfare recipients. The symbolical component of policies provides emotional satisfaction for political supporters of the policy makers and sends a more or less open warning to potential opponents.^[7]

A non-decision is the output of a policy-making body when it defeats a proposed policy initiative through inaction, rather than any overt decision.^[8] The most recent and important example of a non-decision was the 1994 defeat in Congress of President Clinton’s proposed Health Security Act (HSA), which would have established a universal health insurance program for Americans. After President Clinton sent the bill to Congress, its opponents mounted an

effective negative mass media campaign. Support for the bill dropped, and its opponents in Congress were able to prevent any vote on the legislation, either on the floor of Congress, or in any relevant committee. This non-decision preserved the decentralized status quo in U.S. health care organization and financing.

4. *Outcomes:* Health policies are intended to have some specific impacts, or to create a changed state of affairs, at some level of society. Examples of outcomes include situations in which enacted policies lead to defined groups of individuals receiving some enhanced access to health services, institutions receiving additional funding and/or legal authority, elected policy makers increasing their public popularity and ability to get campaign contributions, thus increasing their likelihood of reelection, and so forth. Most commonly, health policies aim to improve health care access, cost control, and/or quality.

However, policies often do not have their expected impact because their *implementation* varies in terms of efficiency and effectiveness. Imperfections in implementation will inevitably result in outcomes that are less significant than expected. For example, one of the health policy outputs of the 1997 Balanced Budget Act was funding for the establishment of State Children's Health Insurance Programs (SCHIPs). The intention was to increase health insurance coverage for children. However, current research shows that state implementation of these programs has been quite uneven.^[9]

Because health policy making usually involves considerable complexity and uncertainty, there is often a significant gap between intended and actual outcomes.^[10] Sometimes important policy outcomes are unintended and unexpected. When enacted in 1965, the intended outcomes of Medicare and Medicaid included increased access to health services for senior citizens and some categories of the poor. However, the increased health spending and increased utilization were also primary factors in the U.S. health care cost inflation that has persisted to this day.

U.S. HEALTH POLICY OUTPUTS AND OUTCOMES: 1965–PRESENT

Since 1965, the United States has far outspent every other nation in the world, both in the aggregate and per capita, in health services. With the Federal Government taking the lead, public policies that dramatically expand the governmental role in organizing, financing, and delivering health services have been enacted.

The list of health policy outputs since 1965 is extensive. Programs such as Medicare and Medicaid have significantly extended health insurance to defined population categories. The Federal Government has also spearheaded a movement to improve health services through more effective research into health outcomes, subsequently linked to clinical and organizational improvements. In 1989, the Agency for Healthcare Research and Quality (AHRQ) was established. It works with academical and industry researchers to advance health outcomes measurement, analysis, and applications.^[11]

In reaction to the rapidly growing portion of the gross domestic product devoted to health care, the Federal Government has steadily put cost containment measures into place, including a Prospective Payment System (PPS) for hospital Medicare services in 1983 (P.L. 89-97) and a Resource-Based Relative Value Scale (RBRVS) for physician Medicare reimbursement in 1989 (P.L. 101-239). Most recently, the 1997 Balanced Budget Act reduced Medicare and Medicaid provider reimbursements by an estimated US \$128 billion in 5 years (but recently, the provisions of the Act have been significantly relaxed).

State governments have also been active health policy makers.^[12] Many states have experimented with new delivery systems in their Medicaid programs, and the general range and level of benefits have gone up steadily in the last 20 years. Some states have instituted nearly universal employer-based health insurance systems (Hawaii), or have attempted to greatly expand health insurance coverage through state mechanisms (Tennessee).

With all of the above-mentioned public policy *outputs*, how have health *outcomes* been affected? It appears that public policies have, at best, a limited impact on population health. Some public health measures, such as disease immunization and water purification, have dramatically affected life expectancy in the last century. However, beyond that, it appears that most health outcomes are the product of overall environmental conditions and individual lifestyle behaviors, not health policy measures.^[13] Because of less effective prenatal care, along with less healthy population lifestyles, Americans on average have a shorter life expectancy than most citizens of advanced Western European and Asian nations. To improve these results, American policy makers must develop more effective methods of health education.

TWENTY-FIRST CENTURY HEALTH CARE POLICY CHALLENGES

In addition to the ongoing task of cost containment balanced with reasonable access to the ever-expanding

array of medical technologies, the twenty-first century will see the development of additional health policy challenges.

Health Disparities and the Market

As medical technologies develop, the gap between possible treatments and available resources to pay for these treatments grows. All societies, including the United States, will have to find some acceptable ways in which to ration the most expensive medical technologies. How will this be done? From a pure efficiency standpoint, the best way to ration resources is through the use of market mechanisms, such as price. The United States has already gone farther than most countries in turning health care over to markets as an economic commodity.

However, markets that are somewhat “free” imply that those who cannot pay will go without. To what extent are Americans willing to accept this? The United States already has significant disparities in medical access. If the United States wants to keep some semblance of rough equality in the provision of essential health services, it will have to increase its subsidies to the medically indigent to keep pace with the growing cost of care. If Americans are unwilling to do this, the nation could wind up with enormous and growing gaps in medical access between social groups—gaps large enough to create future political unrest.

Genetic Technologies

With the ongoing mapping of DNA coming from the Human Genome Project, twenty-first-century health providers will increasingly have the ability to directly intervene in human genetic structures and functioning. This potentially means the ability to treat, or even prevent, the development of genetically based disorders, including cancer and cardiovascular diseases. Success here could materially increase average human life span, from the current 70–80 to 130 years or more.^[14] This increase, if attained, would be a mixed blessing, bringing on funding challenges in geriatric health care that dwarf our current problems.

Growing abilities to manipulate human genetics would also give us the ability to eliminate almost *any* genetic feature, not simply those generally considered undesirable. Thus, there could be sharp policy debates over the social and ethical implications of using genetic medicine to select (or eliminate) actual or potential offspring by gender, physical appearance, or expected intelligence.

CONCLUSION

Based on the above suggestions as to twenty-first-century health policy challenges, one of the insights provided earlier should come to mind. Health policy making does not “solve” problems in a once-and-for-all fashion. Often, current problems that policy makers face result from prior policy initiatives that were and still are considered successes. It should not be hard to imagine future policy scholars drawing the same conclusion about such twenty-first-century advances as genetic medicine. The more things change, the more they remain the same.

REFERENCES

1. Anderson, O.W. *The Health Services Continuum in Democratic States*; Health Administration Press: Ann Arbor, MI, 1989.
2. Longest, B.B., Jr. *Health Policymaking in the United States*; Health Administration Press: Chicago, IL, 1998; Chapter 3.
3. Kingdon, J.W. *Agendas, Alternatives, and Public Policies*; HarperCollins College Publishers: New York, NY, 1995.
4. Ladd, E.C. *The American Polity*; W.W. Norton and Company: New York, NY, 1989; Chapter 5.
5. White, J. *Competing Solutions*; The Brookings Institution: Washington, DC, 1995; Chapter 10.
6. Foreman, C.H., Jr. *Plagues, Products, and Politics*; The Brookings Institution: Washington, DC, 1994; Chapter 6.
7. Edelman, M. *The Symbolic Uses of Politics*; University of Illinois Press: Urbana, IL, 1974.
8. Crenson, M.A. *The Unpolitics of Air Pollution*; Johns Hopkins University Press: Baltimore, MD, 1974.
9. Bilheimer, L.T.; Colby, D.C. Expanding coverage: reflections on recent efforts. *Health Aff.* **2001**, *20* (1), 83–95.
10. Thompson, F.J. The evolving challenge of health policy implementation. In *Health Politics and Policy*, 3rd Ed.; Litman, T.J., Robins, L.S., Eds.; Delmar Publishers: New York, 1997; 155–175.
11. Gaus, C.; Fraser, I. Shifting paradigms and the role of research. *Health Aff.* **1996**, *15* (2), 235–241.
12. Holahan, J.; Nichols, L. Health policy in the states: overview. In *Health Policy, Federalism and the American States*; Rich, R.F., White, W.D., Eds.; Urban Institute Press: Washington, DC, 1996; 39–70.
13. Feldstein, P.J. *Health Care Economics*; Delmar Publishers: New York, 1998; Chapter 2.
14. Schwartz, W.B. *Life Without Disease*; University of California Press: Berkeley, CA, 1998.

Health Care Settings, Change in

R. Wayne Boss

Leeds School of Business, University of Colorado at Boulder, Boulder, Colorado, U.S.A.

Mark L. McConkie

Graduate School of Public Affairs, University of Colorado at Colorado Springs, Colorado Springs, Colorado, U.S.A.

INTRODUCTION

Health care organizations are unique. Hospitals, for example, are different from other organizations in both the private and public sectors, and their differences make the management of change much more difficult. In contrast to industrial firms, for example, hospitals have the following characteristics:

- They tend to be more vitally important to society.
- They have more highly differentiated, specialized, and complex organizations.
- They contain four major power groups (physicians, nurses, administrators, and boards of trustees) with different interests, poor communication, and weak linkages.
- They are dominated by one power group (physicians) who are socialized as professionals, assume cosmopolitan roles, and function autonomously and independently from the rest of the organization.
- They produce personal, unique services with an unusually strong emphasis on quality.
- They have vague goals, produce results that are difficult to measure, and show little standardization among organizations.
- They tend toward random delivery of services with little organizational control.
- They are less adaptable to planned change, because so much of the energy is devoted to adapting to the external environment.
- They provide fewer incentives for effective management, primarily because effective management is not highly valued by the physicians.
- They experience greater and more complex levels of intraorganizational interdependence among professionals and non-professionals.
- They need many highly trained and technically competent professionals with related but often mutually exclusive skills and expertise.
- They are staffed by personnel who are often unwilling to take risks and try new modes of delivery.
- They are paid for good intentions and programs, rather than for their products and results.
- They are subject to government regulation that has an immediate and far-reaching impact on the industry, particularly in terms of reimbursement patterns.
- They are often controlled by insurance companies that make reimbursement decisions (in terms of the amount of reimbursement and the services for which the hospital will be reimbursed), and determine from what institutions patients can receive treatment. This takes the decision for health care services out of the hands of the physicians and puts them into the hands of bureaucrats and administrators.
- Their personnel populations have reached almost a crisis point because of increased demands for services (due to the aging population) and the decreased numbers of people entering the health care field.
- They are far more interdependent, to the point that it is impossible for some departments to deliver services to patients without the cooperation of many other departments. Indeed, if one department, such as the pharmacy, fails to do its job, it can practically close down the entire organization (Adapted from Ref.^[1], pp. 15–16). Given these differences, many consultants and health care administrators find the process of change difficult to manage. However, the research indicates that significant changes can take place when those involved in the change process pay attention to the unique characteristics of health care organizations.

ORGANIZATION DEVELOPMENT

The typology used here comes from an Organization Development (OD) perspective. Although the definitions of OD vary significantly, the most widely accepted is that given by Beckhard: “Organization development is an effort that is planned, organization wide, and managed from the top to increase organization effectiveness and health through planned interventions in the organization’s processes, using behavioral science

knowledge.”^[2] OD has a long-range focus. It is designed to improve an organization’s health, efficiency, and effectiveness in effecting change, reaching decisions, and taking advantage of the full potential of the organization’s most valuable resource: its people. In short, OD applies behavioral science knowledge to facilitate change, to build skills in people, and to resolve problems, whether personal, interpersonal, intragroup, intergroup, or organizational (see Ref.^[1], p. 18).

THE ACTION RESEARCH MODEL

The basic approach of most OD practitioners is to use the action research model, a data-based, problem-solving model designed to replicate the scientific method of inquiry (for a general discussion of action research, see Ref.^[3]). The specific steps in the action research model include data collection from individuals and groups within the organization, feedback of the data to the organization personnel, problem diagnosis, and action planning and interventions within the system. Once the interventions have been completed, additional data are collected and fed back to the client system, problems are diagnosed, and additional action planning of subsequent interventions takes place. Thus, implementation of the action research model involves a cycle of interactions designed to provide immediate feedback to the client on the status of problems within the system. Action research, therefore, implies a meaningful link between research and action and connotes action based on valid information. It is important to realize that each phase of the action research model constitutes a clear intervention into the client system (Ref.^[4], p. 260).

OD INTERVENTIONS

The term “OD interventions” refers to a range of structured activities designed to resolve problems and enable personnel to develop and maintain healthy organizations. These methods and techniques can be implemented by the client and the consultant and constitute an evolving array of methods and techniques, based on the diagnosis of the problems and the needs of the client system.

Although few OD projects include all possible intervention activities, the following constitutes a generally agreed upon set of definitions for the major types of OD interventions.

- *Team building*: Activities designed to enhance the effective operation of system teams.
- *Intergroup activities*: Activities designed to improve effectiveness of interdependent groups. They focus on joint activities and the output of groups considered as a single system rather than as two subsystems.
- *Survey-feedback activities*: Although related directly to the diagnostic process described above, these activities tend to become an ongoing process in the organization. Data collection activities, in the form of questionnaire surveys of the target population and interviews with key personnel, take place on a regular basis. The data are analyzed, problem areas are identified by the organization’s leadership, and action plans are developed to deal with problems related to the total organization. The data are then shared with the organization’s personnel in department or division meetings, and specific action plans are formulated for problems that relate to the specific units.
- *Education and training activities*: Activities designed to improve skills, abilities, and knowledge of individuals. The activities may be directed toward technical skills required for effective task performance or may be directed toward improving interpersonal competence. The activities may be directed toward leadership issues, responsibilities and functions of group members, decision making, problem solving, goal setting, planning, and so on.
- *Technostructural or structural activities*: Activities designed to improve the effectiveness of the technical or structural inputs and constraints affecting individuals or groups. The activities may take the form of experimenting with new organization structures and evaluating their effectiveness in terms of specific goals or devising new ways to bring technical resources to bear on problems. Structural interventions are defined as the broad class of interventions or change efforts aimed at improving organization effectiveness through changes in the task, structural, and technological subsystems. Included in these activities are certain forms of job enrichment, management by objectives, collateral organizations, and physical settings interventions.
- *Sociotechnical systems*: Activities designed to focus simultaneously on the technical systems and the social systems of the organization. Such activities often involve the restructuring of work methods, rearrangements of technology, or the redesign of organizational social structures. The objective is to maximize the “fit” between the technical system and the human system. Often, the implementation of sociotechnical systems involves the creation of work teams or autonomous work groups, which are responsible for a specific product or output. The concept, originally developed in London at the Tavistock Institute of Human Relations, has had broad applications in mines and factories in

England and the Scandinavian countries. Successful sociotechnical interventions have also been reported in joint United Auto Workers-General Motors plants in the United States.

- *Process consultation*: Activities on the part of the consultant to assist the client in perceiving, understanding, and acting upon process events that occur in the client's environment. Specific emphasis is placed on the human processes that occur at the individual, interpersonal, and intergroup levels. These processes include communications, group problem solving and decision making, the functional roles of group members, group norms and group growth, leadership and authority, and intergroup processes. Because organizations are made up of networks of people, with these processes continually occurring between them, the objective of process consultant activities is to help the client gain the skills to understand, diagnose, and manage such processes effectively.
- *Third-party consultation*: Activities conducted by a skilled third party, which are designed to resolve problems regarding interpersonal or substantive issues and help those parties deal effectively with their conflict. Such interventions involve conflict resolution in superior-subordinate and peer relationships. Among the substantive issues are disagreements over policies and procedures, different conceptions of roles, and competition for scarce resources. The interpersonal issues often involve feelings of anger, distrust, fear, scorn, resentment, or rejection.
- *Coaching and counseling activities*: Activities that involve the consultant or other organization members working with individuals to help them define learning goals; learn how others see their behavior; and learn new modes of behavior to see if these help them to better achieve their goals. A central feature of this activity is the non-evaluative feedback given by others to an individual. A second feature is the joint exploration of alternative behaviors.
- *Life and career planning activities*: Activities that enable individuals to focus on their life and career objectives and how they might go about achieving them. Structured activities lead to production of life and career inventories, discussions of goals and objectives, assessment of capabilities, needed additional training, and areas of strength and deficiency.
- *Planning and goal-setting activities*: Activities that include theory and experience in planning and goal setting, utilizing problem-solving models, and the like. The goal of all such activities is to improve these skills at individual, group, and organization levels (see Ref.^[4], pp. 262–265).

A MODEL FOR CHANGE

Golembiewski provided a model of organizational change that has proven effective in a number of different settings. Specifically, to introduce change into simple systems, he suggested include the following:

1. Move slowly.
2. Implement one innovation at a time.
3. Precede each step with ample warning.
4. Accompany each step by a statement of reasons and benefits.
5. Offer maximum opportunities to participate.
6. Provide outlets for inevitable hostilities.

Occasionally, harsh or dramatic changes are desirable; sometimes it is the only way to facilitate change. When the manager feels it practically necessary or morally defensible to so induce change, that manager simply will neglect the above list. Even then, however, the manager can expect a low probability of success and should anticipate much time and effort sweeping up afterwards.

A comparable list of steps for changing complex more systems is less straightforward or complete. At least fragmentarily, then, the managerial change agent intent on affecting a complex system should keep such general guidelines in mind:

1. Changes in any subsystem often require changes in other systems.
2. Changes at any one level of organization often require changes at the level above and below.
3. Change is usefully begun where stresses exist.
4. Planning any change should involve formal and informal organizations.
5. The higher the level at which change begins, the greater the chances of success.^[5]

Given the complex nature of health care organizations, change agents in hospitals do well to adhere to the guidelines suggested for complex systems, particularly given the highly interdependent nature of health care organizations. For example, the leaders of an outpatient surgery department in a major medical center decided to move the nurses' station from one side of a room to the other. This change directly impacted 29 other departments.

THREE CASE STUDIES

Given the challenges currently facing health care, one would expect it to prove a fertile field for applying OD technology. Consider, for example, the following applications of OD technology to a variety of different

health care settings (materials in this section is adapted from Ref.^[1], pp. 1–2):

Lincoln Hospital's Operating Room

Lincoln is a 400-bed for-profit hospital and medical center located in the southwestern United States. Lincoln had long had trouble with egotistical and dictatorial physicians who constantly threatened to take their patients elsewhere if their demands were not met. Historically, the hospital administration had catered to these doctors, even firing nurses who had alienated particular physicians.

Mary, the OR director for 13 years, had a reputation as being a conservative, strong-willed, strict constructionist who would brook no deviations from hospital rules. Don, the incoming chief of surgery, was an outspoken critic of the hospital. Nurses and administrators feared his sharp tongue, quick temper, and critical disposition, but Mary was not the least bit intimidated by his behavior, and the two had openly clashed on numerous occasions. Faced with the new broom's need to make a difference, Don openly declared that his major goal for his first year as chief of surgery was to have Mary fired.

One year later, following multiple OD interventions, Mary had successfully completed her 14th year as OR director. Not only did she and Don trust each other, but the two had engineered some major changes in the surgical services area, including a decrease in the turnover rate among surgical nurses from 40% to zero, a resolution of the difficult problem of scheduling operations for a group of *prima donna* surgeons, and a significant decrease in physicians' verbal abuse of nurses.

Manchester Nursing Services Division

Manchester hospital and medical center is a 300-bed for-profit facility located in the northwestern United States. Although all of Manchester's 900 employees generally felt frustrated with the problems facing the organization, low morale was most pronounced in the 500-employee nursing services division. Several factors contributed to the nurses' dissatisfaction: the decentralization of nursing management, enormous technical and managerial problems for the new managers; an almost 100% turnover among head nurses; mistakes in filling key management positions; distrust for the administration; and physician abuse. The nurses felt unappreciated, undervalued, and powerless in dealing with the administration and with the physicians, and the turnover rate among nurses had fluctuated between 30 and 40% for the previous four years.

Three years later, the Manchester nursing services division achieved significant improvements in organization climate and group effectiveness. The turnover rate among nurses dropped from over 32% the year before the OD project started to 5.8% the third year. This decrease in turnover saved Manchester over \$1.5 million during the three-year period.

City Hospital and Medical Center

City Hospital is a 250-bed denominational hospital and medical center with 850 employees, located in a fast-growing southeastern city. After holding his position for 10 years, the hospital's administrator had run into trouble with the medical staff, the employees, and the corporate board, and had left for a hospital in a different region.

The new administrator inherited several serious problems: a lack of effective leadership, ineffective communication, poor decision making, an absence of planning, a hostile medical staff, a poor public image, dissatisfied nurses, an employee population that felt overworked, underrewarded, and burned out, and serious financial problems (to the point that the hospital had recorded its first financial loss in over 50 years).

Three years later, the number of City Hospital employees in the advanced stages of burnout had decreased by 31%. Net revenue as a percent of gross revenue (profits) showed double-digit increases, and market share had increased by 6.3%. Days of revenue in gross accounts receivable were down by 18.9 days, for an increase in revenue of approximately \$2,835,000 during the three years. In addition, employee turnover dropped from over 25% to 4%. By the most conservative estimates, the hospital had saved over \$2 million in recruitment costs for registered nurses alone during the three-year period.

ESSENTIALS FOR SUCCESS

The purpose of OD is to improve an organization's effectiveness in making decisions, accomplishing its missions, and in reaching the full potential of its people. Successful OD efforts will simultaneously create a regenerating system characterized by trust, openness, a flow of accurate information, freedom for individuals to grow and develop, and a minimal interpersonal risk or perceived threat. Under such conditions, minimal resistance to change is generated, and a high level of collaboration is ensured (material in this section is adapted from Ref.^[1], pp. 102–105).

Of course, not all OD efforts attain these objectives. Often, the project is doomed to failure simply because many essential elements of effective change are

missing. A successful OD project will include the following variables:

1. Commitment of the chief executive officer.
2. Cultural preparedness in the organization.
3. Realistic expectations about the success of the project.
4. Internal tension and pain.
5. Involvement of other people in positions of power.
6. Development of internal resources.
7. Methods for holding people accountable.
8. A competent consultant.
9. A clear psychological contract between the CEO and the consultant.
10. Technically competent supervisors.

Commitment of the Chief Executive Officer

The CEO is the single most important person in the change effort, largely because the CEO controls the reward structure, which in turn, controls individual behaviors. Two further examples illustrate this point.

First, 37 top-level administrators in seven criminal justice agencies participated in a 6-day confrontation team building meeting. Thirty-three were members of six natural teams or organizational family groups. The remaining four comprised a cousin group from a seventh organization. Cousin group members represented four different departments in a state agency. No two members reported to the same supervisor. The major variable among the six natural teams was that only one team met with its CEO. Although the natural team with the CEO showed significant improvement between the before and after scores on standard organization climate measures, the five leaderless teams, the cousin group, and a comparison group showed no statistically significant change.

Second, a seasoned hospital administrator retained a consultant and launched a major OD effort. During the next two years, the consultant worked closely with the administrator, his administrative council, and department managers to resolve problems, open channels of communication, enhance cooperation, and increase organizational effectiveness. Hospital personnel, however, perceived the administrator's behavior to be incongruent with the values of openness, honesty, and candor that he had professed to espouse. They saw him as deceitful and manipulative and felt that OD was a gimmick. Two years after the project began, the consultant quit; six months later, the administrator was fired.

Within two months, a new administrator was appointed. He immediately hired the same consultant and began another 2-year OD effort, with the same

people and the same objectives. Both projects were characterized by the following:

- The same consultant was employed.
- The consultant used the same approach.
- The consultant spent approximately the same number of days in the organization each year.
- The organization experienced little turnover, and there was none among the administrative council level.
- The general approach to hospital work was the same.

The only major difference was the person who served as administrator. Data collected at the start of the first OD effort, 2 weeks before the administrator was fired, and 2 years after the new chief was hired, showed that the organization retrogressed between the first and second data collections but improved significantly in organizational climate and leader effectiveness by the time of the third collection.^[6]

The results of these and other studies suggest that the CEO is the single most important factor in determining the success or failure of an OD effort. Without the CEO's total support, the potential for a successful OD project is questionable at best; if that support is marginal, the OD project should not be started at all.

Cultural Preparedness

To achieve lasting change, an organization must be culturally prepared; that is, its employees' psychological energies must be directed toward achieving, not sabotaging, the organization's goals. Below a certain threshold of organizational health, successful OD efforts are impossible. Important elements of cultural preparedness include interpersonal trust among employees, their willingness to resolve problems, their willingness to be held accountable, and their psychological health.

Interpersonal Trust

Trust is essential to organizational effectiveness. People who trust each other are significantly more effective in group problem solving than people who do not. The degree to which people trust one another determines the degree to which they will communicate honestly, share relevant information, try to control others, and permit others to influence their own behavior and decisions. The level of interpersonal trust at the outset of an OD project is typically an accurate indicator of the potential success of the effort. If trust has degenerated below a minimum level, the consultant or manager must scrap the project or replace the

offensive party. Without minimal trust, the chances of successful changes are practically zero.

A Willingness to Resolve Problems

Organization development, by definition, involves problem solving. Those involved must be willing to resolve their problems or change cannot take place. For example, in one client organization, the line personnel had major conflicts with their supervisors. During the diagnostic phase of the project, the consultant interviewed all the parties, but 60% of the line personnel refused even to sit down in the same room with their supervisors and discuss the problems. Under such circumstances, the objectives of OD could not be achieved; only power-oriented, win-lose behavior would have resolved problems.

A Willingness to be Held Accountable

Effective managers are not only able to hold employees accountable but also willing to be held accountable themselves. Competent people generally appreciate such responsibility. Incompetent people, on the other hand, resist it, because it exposes their weaknesses and inabilities.

Psychological Health

Although psychological health is relative, people must have a minimum level to resolve conflict effectively. Without it, supportive confrontation is difficult. If most of the members of a group need intense therapy, rational problem solving is impossible.

Realistic Expectations

Everyone involved in an OD project must understand at the outset that it is a long-range project requiring a substantial investment of time and energy. It took time to create the problems, and they should not expect those problems to be resolved quickly. Realistic expectations can prevent the anxiety, frustration, and tension associated with most change efforts.

Internal Tension and Pain

Any successful OD project results from recognizable discomfort, tension, and pain. Without a felt need for change, particularly among the administrators, no change can occur. Organizational change efforts have

a high probability of success only when the individuals involved have been experiencing internal stress.

Involvement of People in Power

Change efforts, where possible, should start with top-level administrators but must involve the informal (natural) leaders of the rank and file. Their support and cooperation can mean the difference between success and failure.

The Development of Internal Resources

Internal resources must be developed to ensure that the organization remains healthy after the consultant leaves. Without such resources, it remains dangerously dependent on the external consultant. This resource development can take two forms: intense managerial training and educational programs can build managerial competence throughout the leadership; or a few employees can be trained as internal change agents. For the latter, areas of emphasis should at least include process consultation, conflict management, team building, data gathering, problem diagnosis, and personnel training.

Methods for Holding People Accountable

Effective managers must be able to hold people accountable for the fulfillment of their responsibilities. One method used in the studies reported in this book is the personal management interview. Regular meetings between supervisor and subordinate provide opportunities for sharing information, training, building technical skills, dealing with interpersonal issues, and holding employees accountable. Notes of each meeting are kept, and action items are reviewed during subsequent meetings to ensure followup.

A Competent Consultant

Although many changes can be effected by resources within the organization, the objectives of OD generally require the help of someone outside it. Obviously, caution should be exercised in selecting that person. A client-consultant relationship, like a marriage, requires a great deal of interdependence and trust. The CEO should consider whether the consultant has the professional skills necessary to help diagnose and solve the problems and the interpersonal skills necessary for a trusting and open relationship with company personnel. The CEO should thoroughly interview

consultants and check with their previous clients, especially those in similar organizations.

A Clear Psychological Contract Between the CEO and the Consultant

It is absolutely essential that the consultant and the CEO reach a clear understanding about all aspects of their relationship before the consultant enters the system. They must spell out the client's expectations of the consultant and OD effort, the consultant's expectations of the client, the problems to be solved, the phases of the project, and the parameters within which the consultant and client will work. Such a psychological contract prevents misunderstandings and difficulties during the OD project.

Technically Competent Supervisors

Although not so central as the other essentials, the technical competence of the supervisors, especially the top administrators, is a major factor in successful change efforts. This factor is particularly important in occupations with a high degree of interdependence and personal risk. When subordinates put their destiny in the hands of someone else, the technical competence of the critical others is essential.

CONCLUSION

The factors listed above do not include all of the variables that may contribute to an effective change effort; and most consultants can cite cases where change efforts were successful in the absence of one or more of these conditions. Ignoring any of these essentials, however, will greatly hamper the change effort and substantially decrease the prospects of achieving the objectives of OD.

REFERENCES

1. Boss, R.W. *Organization Development In Health Care*; Addison-Wesley: Reading, MA, 1989; 15–16.
2. Beckhard, R. *Organization Development: Strategies and Models*; Addison-Wesley: Reading, MA, 1969; 9 pp.
3. French, W.L.; Bell, C.H. *Organization Development*, 5th Ed.; Prentice-Hall: Englewood Cliffs, NJ, 1995; 137–154.
4. Boss, R.W. Organization development: a vehicle for improving the quality of work life. In *Handbook of Organization Management*; William Eddy, W.B., Ed.; Marcel Dekker Inc.: New York, 1983; 260 pp.
5. Golembiewski, R.T. *Perspectives on Public Management*; F.E. Peacock Publishers: Itasca, IL, 1967; 35–36.
6. Boss, R.W.; Golembiewski, R.T. Do you have to start at the top? the chief executive officer's role in successful organization development efforts. *J. Appl. Behav. Sci.* 31 (3), 259–277.

Health Care, Assessment and Evaluation of

Leiyu Shi

Department of Health Policy and Management, Johns Hopkins University,
Baltimore, Maryland, U.S.A.

INTRODUCTION

This article provides an overview of assessment and evaluation of health care. It describes the purposes, characteristics, and applications of health care evaluation. It delineates the process to conduct evaluation including determining the scope of evaluation, getting acquainted with the program to evaluate, finalizing the methodology for evaluation, collecting data, analyzing data, and reporting findings. The article will benefit those intent on conducting evaluation studies or critiquing evaluation work.

BACKGROUND

Assessment and evaluation of health care, or health care evaluation for short, is the systematic application of scientific research methods for assessing the conceptualization, design, implementation, impact, and/or generalizability of health services programs. It may be used for program monitoring and refinement, or for policy application and expansion.^[1,2] The first purpose centers on the health service or program itself. It aims at finding out how a program actually operates, whether the program has fulfilled its objectives, and what improvements can be made toward program operations. In particular, evaluators are interested in addressing the following questions: How has the program been implemented compared to program plans? What are the characteristics and numbers of the beneficiaries? What are the strengths and weaknesses in its operations? To what extent has the program served its intended beneficiaries? What is the dropout rate? How cost-effective is the program? How cost-efficient is the program? What is the impact of the program? How can the program be improved for the future?

The second purpose goes beyond the program itself by analyzing the extent to which health policies achieve

particular effects and assessing whether the program can be expanded or applied to a different setting. Evaluators are concerned about the following: How generalizable are the findings regarding the program? What is the potential of the program being applied to other settings? What are the conditions for such an application? What will be the costs and intended benefits?

The difference between these two purposes is that the first purpose is limited to improving effectiveness of a specific program within that setting. There is no attempt to generalize findings beyond the setting in which the program is operating. The second purpose serves to enlighten policy makers or funders by providing relevant information for decision making regarding program modification, refinement, expansion, or promulgation. Such evaluation can inform action, enhance decision making, apply knowledge to solve health problems, help set priorities, and guide the allocation of scarce resources.

Health care evaluation has a number of characteristics.^[3] First, it is technical. Using established research methods, researchers design, implement, and assess a health-related project in a way that is reproducible by others. Rigorous application of scientific methods is necessary for the evaluation results to be valid and legitimate. Thus evaluators must be conversant with research methods commonly used for evaluation.

Second, health care evaluation is applied. In contrast to basic research, evaluation is undertaken to solve practical health problems and to have some real-world effect. It seeks to understand pressing organizational or societal problems and identify potential solutions. It studies the processes and outcomes of the attempted solutions. The researcher translates between the academic discipline and the world of action. The evaluator is often paid by the client who may be manager or administrator in a business, government, service, or private funding agency. The criteria for judging evaluation include its ability to provide useful knowledge, its utilization by decision makers, and its impact on health programs or conditions. Because the value of evaluation depends on its utilization by others, evaluators must be knowledgeable about the social dynamics of the setting in which they perform evaluation as well as the subject matter related to the evaluation.

Adapted from Shi, Leiyu, *Health Services Research Methods*. Albany, NY: Delmar Publishers, Inc., 1997, Chapter 9. Evaluation Research. pp. 182–214.

Third, evaluation should be objective. Program funders, administrators, and evaluators all have the duty of making the evaluation objective. On the part of program funders and administrators, the results of evaluation should not be tied to the current or future reward for the evaluator. If evaluators are personally or financially tied to the project they evaluate, they may be loath to report negative findings. Evaluations conducted under these circumstances may err in the direction of praise. On the part of evaluators, planning and conducting an evaluation often requires rapport, trust, and frequent interaction with program administrators. Evaluators should not let the process of gaining trust and rapport color their perspectives. Regardless of where or how they are employed, they need to find viable ways to maintain integrity, objectivity, and an appropriate sense of differentiation.

EXAMPLES OF HEALTH CARE EVALUATION

The field of health care evaluation is marked by diversity in disciplinary training, perspectives on appropriate method, and diversity in evaluation activities and arrangements. The major types include needs assessment, process evaluation, outcome evaluation, and policy analysis.^[2-4]

Needs Assessment

The purpose of needs assessment is to identify weaknesses or deficiency areas (i.e., needs) in the current situation which can be remedied or to project future conditions to which the program will need to adjust. The decisions following a needs assessment usually involve allocation of resources and efforts to meet high identified needs.

Needs assessment has several applications. It can be used to identify goals, services, problems, or conditions which should be addressed in future program planning. Important questions to be addressed include: What are the goals of the organization or community? Is there agreement on the goals from all groups within the organization or community? To what extent are these goals being met? What are the areas in which the organization is most seriously failing to achieve goals? What do clients perceive they need? What problems are they experiencing? How effective is the organization in addressing problems perceived by clients? What are the new challenges affecting the organization or community? How can those challenges be met? What additional resources are necessary in meeting the challenges?

Needs assessment is also used to determine the need for additional health services or providers in a

community. First, the need for the particular health service or provider in the relevant service area is examined. A trend study is usually necessary because the new service is likely to impact the service area population not immediately, but rather some time in the future. The second step is to look at the supply component in the given service area. Similarly, a trend analysis is necessary to project the future service level when the new competitor will be in the market. The third step is to assess if a gap exists between estimated requirements and estimated supplies. Finally, a determination is to be made regarding new entrance need based on the size of the gap and the impact the new service and provider will have on existing services and providers.

Process Evaluation

Process evaluation focuses on how a particular program operates. It is concerned with the activities, services, materials, staffing, and administrative arrangements of the program. Process evaluation is conducted not so much as to decide whether to continue or drop a program. Rather, it is conducted to monitor the implementation of the program, to find out how a requirement or procedure has been implemented by program administrators, where things are going as planned, where they may deviate from planned directions, and what factors are associated with such deviation. Monitoring also may include assessing whether program activities comply with legal and regulatory requirements—for example, whether affirmative action requirements have been met in the recruitment of staff and whether discrimination exists in the selection of clients. As a result of process evaluation, modifications may be made in the activities, services, staffing, organization, and other aspects of the program either to correct an unintended deviation from the plan or to address an unexpected problem encountered in the implementation stage.

There are many reasons for monitoring the process of implementing a program. First, program monitoring ascertains that program administrators conduct their day-to-day activities efficiently and, if not, identify ways to enhance efficiency. Second, program funders, sponsors, or stakeholders require evidence to indicate that the program is being implemented as planned and for the purpose that it was paid for. Third, process evaluation identifies unexpected problems that need to be corrected immediately rather than wait for the end of the normal duration of the program. Fourth, process evaluation is often a prelude to outcome evaluation. There is no point in assessing the impact or outcome of a program unless it has indeed taken place in the way intended. Finally, monitoring program costs

and resource expenditures provide essential information for estimating whether the benefits of a program justify the costs.

Important questions to be addressed include: What are the critical activities, services, staffing, and administrative arrangements, etc. in the program? How are resources including staff, money, and time allocated? What is a typical schedule of activities and of services? How are program activities supposed to lead to attainment of program objectives? How many participants and staff are taking part in the program? How are the important components of the program operating? To what extent has the program been implemented as designed? To what extent has the program reached its target population? What are the resource expenditures for providing a service? Is the program implemented efficiently? How can the program be more efficiently carried out? What adjustments in program management and support are needed? Are there any problems encountered in program operation? If yes, what is the nature of the problems? How can they be solved?

Many of the data elements required for process evaluation can be incorporated in a Management Information System that routinely collects information on a client-by-client basis about sociodemographic data, services provided, staff providing the services, diagnosis or reasons for program participation, treatments and their costs, impressions of outcome, and other comments. The information is essential for program monitoring as well as assessment at a later stage.

Outcome Evaluation

Outcome evaluation is concerned with the accomplishments and the impact of the program, and the effectiveness of the program in attaining the intended results. The purpose is to collect and present information needed for summary assessment about the program and its value. Program effectiveness is tied to its intended goals which are usually stated in terms of participants' outcomes. Examples can be health status (e.g., recovered from an illness), behaviors (e.g., primary care visits), performance (e.g., stop smoking), cognitive (e.g., knowledge or skill gained), or affective (e.g., satisfaction level). While evaluators primarily emphasize explicit program goals, they also notice unanticipated or unstated outcomes, both positive and negative. Outcome evaluation examines the extent to which a program's goals are or are not being achieved. The results of outcome evaluation may assist the sponsors or decision makers in deciding whether to continue or discontinue a program, or whether (and how) to expand or reduce it.

To judge whether the program is successful or not, evaluators often compare program's accomplishments with some comparable standards, such as the results of the status quo or some competing alternative programs with the same goals. Even if a program is not totally successful, evaluators may conduct an impact assessment that gauges the extent to which the program causes change in the desired direction. Because of resource constraints, in most cases, knowledge of program accomplishments alone is insufficient; the results produced by a program must be judged against its costs. The program's effects and costs are contrasted with those produced by an alternative program with the same goals. Programs that are cost-effective are preferred to those that are either ineffective or effective but costly.

Important questions to be addressed include: What are the goals of the program? How are they measured or assessed? What programs are available as alternatives to this program? How are the program's essential components (e.g., activities, services, staffing, and administrative arrangements) tied to achieving program goals? Why should these components contribute to achieving goals? Does the program move toward goal achievement? How successful is the program in accomplishing its intended results? What conclusion can be made about the effects of the program or its various components? Has progress been made toward meeting the program goals? Which activities or services best accomplish each of the program goals? What gaps exist in meeting the program goals? What adjustments in the program might lead to better attainment of the goals? Is the program or some aspects of it better suited to certain types of participants or certain settings? How effective is this program in comparison with alternative programs? How costly is this program in comparison with alternative programs? Are there unanticipated outcomes associated with the program? If so, what are they? Should the program be discontinued, continued, expanded, or modified?

Policy Analysis

Policy analysis lays out goals, uses logical processes to evaluate identified alternatives, and explores the best way to reach those goals. It rarely aspires to the standards of traditional academic research which are supposed to be able to withstand the scrutiny of time, and the continuing probes of fellow investigators. An analysis that can be finished until after the decision must be made is not useful to the decision maker. The purpose of policy analysis is to inform the decision maker, providing the best analysis possible given the limitations on time, information, and

resources. Analysis is presented not to overwhelm the decision maker with methodological details, rather in such a way that the essential points can be readily grasped. By relying on the analytic techniques developed in economics, mathematics, operations research, and systems analysis, policy analysis strives to improve our ability to predict the consequences of alternative policies, provide a framework of valuing those consequences, and make rational and better decisions.

The five-step framework^[4] for policy analysis includes: 1) establishing the context; 2) identifying the alternatives; 3) predicting the consequences; 4) valuing the outcomes; and 5) making a choice. Each of these steps contains important questions and issues to be addressed by the policy analysts. For example, to establish the context of analysis, policy researchers must find out the underlying problem that must be dealt with, and the goals to be pursued in confronting the problem.

To lay out the alternative courses of action, analysts need to be knowledgeable about the particular policy and program and know how to obtain further information for analysis. It is also an opportunity for creative thinking so that alternatives may be designed to take advantage of additional information as it becomes available and the decision process becomes flexible, enabling decision makers to change the course of action as they learn more about the real world.

In predicting the consequences or estimating the likelihood of the alternatives, researchers rely heavily on the analytic techniques of management sciences, in particular economics, and operational research (e.g., forecasting and simulation, benefit–cost analysis, discounting, decision analysis, linear programming, critical path method, Markov models).

To value the predicted outcomes, analysts try to choose objective (often quantitative) standards or criteria against which policy choices can be evaluated. Since some alternatives will be superior with respect to certain goals and inferior with respect to others, evaluators need to decide how different combinations of valued goals are to be compared with one another. When valuation problems do not lend themselves easily to quantification, analysts may have to address the issue descriptively.

In selecting the alternative, analysts draw all aspects of the analysis together to identify the preferred course of action. Sometimes, the analysis is straightforward and the alternative that is best may be selected. At other time, the analysis may be so complex that researchers have to rely on a computer to keep track of all the options and their possible outcomes. In most situations, the choice among competing policy alternatives is never easy, for the future is always uncertain and the inescapable tradeoffs painful.

PROCESS

There are many ways to conduct health care evaluation. The following is a description of a six-step process to conduct evaluation.

Determine the Scope of Evaluation

To determine the scope of evaluation, first, the evaluator reviews all pertinent information about the evaluation assignment. Then, some background investigation is conducted to find out more about the nature of the assignment. Finally, the evaluator negotiates and reaches agreements with the sponsor about their mutual expectations of the evaluation assignment.

Get Acquainted with the Program

Once the contract has been signed, the evaluator officially proceeds with the evaluation. The first task is to get fully acquainted with the program you are going to evaluate. Specifically, you should find out about the program's goals and objectives, its principal activities, organizational arrangements, staffing, roles, and responsibilities, relationships between program operations and outcomes, profiles of the clients and services provided, financial performance, and its primary problem(s). In addition, some contextual information about the organization that administers the program may be useful. Such information includes organization's mission, history, services or products, characteristics of staff and clients, etc.

Often, your collaborators or evaluation coordinators are the best source of information and can provide you with materials containing the abovementioned information. Or they may suggest places you can obtain the information. Some common sources that the above information may be obtained include the program proposal written for the funding agency, the request for proposals (RFP) written by the sponsor or funding agency to which this program's proposal was a response, brochures of the program, program curricular or other materials, program implementation directives and requirements, administrative manuals, annual reports, an organizational chart or description of the administrative, decision-making, clinical, and service roles played by various people in the program, patient or client records, daily schedules of services and activities, the program's budget and actual spending reports, memos, meeting minutes, newspaper articles, document describing the program's history or the social context which it has been designed to fit, legislation, administrative regulations, completed evaluation studies, and perspectives and descriptions from

program managers, participants, sponsors, or users. If feasible, you may want to personally “scout” some or all of the program components and activities. At a minimum, you should conduct one site visit to obtain first-hand impressions of how programs actually operate.

The evaluator then directs attention to the goals, objectives, and outcomes of the program and their measurements. The goals and objectives specified for the program will be used as a benchmark. Program staff and planners will be consulted with to make sure these indeed are the goals and objectives of the program. They may be asked to write a clear rationale describing why the particular activities, processes, materials, and administrative arrangements in the program will lead to the goals and objectives specified for the program. Some adjustments may be necessary to add or delete some of the objectives. Reasons for the adjustments should be clearly documented. However, the program’s major goals may not be changed. The evaluator may look for additional sources for program goals and objectives. For example, are there written federal, state or local guidelines about program processes and goals to which this program must conform? What are the needs of the community or constituency which the program is intended to meet? Utilizing the above information, the evaluator can recreate a detailed description of the program, including statements identifying program goals and objectives, and cross-classifying them with program elements or components, comparing how the program is supposed to operate with how it actually operates.

Information about the outcomes of the program may be obtained from the program’s published documents, performance records, productivity indicators, patient or client data base, and cost data such as financial performance records, insurance claims records, workers’ compensation claims records. Often it may be necessary to conduct additional studies to find out more about the program and its performance. For example, participant and/or staff survey or interview may be conducted to obtain additional or supporting data to back up depiction of program events, operations and outcomes. Past evaluations of this or similar program, if available, will provide insight into how measurements can be constructed. Books and articles in the evaluation literature that describe the effects of programs such as the one in question are also valuable. If feasible, you may want to personally observe or monitor the program outcomes.

Finalize the Methodology for Evaluation

Although you should have considered the evaluation plan at the start, now that you have become

acquainted with the program, it is time to finalize the method(s) of evaluation. Specifically, you will decide the evaluation design, data instruments and measures, data collection methods, and data analysis techniques. You may consider involving the primary potential users in planning for the evaluation to facilitate their ownership of the study, and encourage trust and cooperation. As you plan for the evaluation activities, you should draw a detailed time schedule indicating when each activity will be performed by whom using what types of resources, and the duration of these activities. The schedule will be used to monitor the progress of the tasks so that the evaluation can be completed in a timely fashion.

Collect Data

Once the evaluation methodology has been finalized, the evaluator then proceeds to perform the tasks as scheduled. Much of the initial research will involve implementing the planned methodology including design, sampling, data collection and processing. Usually, there will be different kinds of data and different ways of collecting them. Data may be collected cross-sectionally or longitudinally. They may be collected from program staff and/or participants themselves. Staff survey might indicate that services have been delivered. Participants’ survey is valuable for a number of reasons. First, it provides evidence to corroborate the staff’s response. If not, different perspectives can be obtained. Second, securing participant data enables providers and program planners to know what is important to clients, including their satisfaction with and understanding of the intervention. Finally, it is an important way to find out not only whether services were delivered, but what was actually delivered, how they are perceived, and whether they are utilized as intended. Collecting a variety of information enables you to gain a thorough look at the program, and obtain more indicators of program effects.

The evaluator should make sure that the instruments are administered, interviews and observations conducted and coded, secondary data gathered and processed, and the scheduled deadlines met. If necessary, the evaluator should see to it that proper training has been given to those responsible for data collection (sending out questionnaires and monitoring their return, or conducting telephone or face-to-face interviews, observations, field work, etc.).

Analyze Data

Data analyses are conducted according to the techniques specified in the methodology section. The major objective is to find out the net outcome of the program,

i.e., whether or not a program produces desired levels of effects, reflected by the outcome measures, over and above what would have occurred either without the program or with an alternative program. Net outcome may be expressed as gross outcome subtracting effects from non-program-related, extraneous confounding factors and design effects. The net outcome may then be compared with that from other programs or some objective standards.

Report Findings

Evaluators have many ways to report the findings of their evaluation, including informal meetings with program and evaluation sponsors, memos, newsletters, formal presentations, formal written reports, and scholarly publications.^[5] To program sponsors, a formal written report perhaps is the most important product of evaluation and is required from the evaluator. To enhance their utility, valuation results should be presented in a way that non-researchers can easily understand. The evaluator may share any draft reports with the clients for review before turning in the final product.

CONCLUSION

In sum, assessment and evaluation of health care is the systematic application of scientific research methods for assessing the conceptualization, design, implementation, impact, and/or generalizability of health

services programs. Health care evaluation is technical, applied, and should be objective. The process to conduct evaluation consists of determining the scope of evaluation, getting acquainted with the program to evaluate, finalizing the methodology for evaluation, collecting data, analyzing data, and reporting findings. Health care evaluation is likely in high demand given the increase in publicly funded health care programs and the need to establish their efficacy for continued funding.

REFERENCES

1. Stecher, B.M.; Davis, W.A. *How to Focus an Evaluation*; Sage Publications: Newbury Park, CA, 1987.
2. Herman, J.L.; Morris, L.L.; Fitz-Gibbon, C.T. *Evaluator's Handbook*; Sage Publications: Newbury Park, CA, 1987.
3. Patton, M.Q. *Qualitative Evaluation and Research Methods*; Sage Publications: Newbury Park, CA, 1990.
4. Stokey, E.; Zeckhauser, R. *A Primer for Policy Analysis*; W. W. Norton & Company Inc.: New York, NY, 1978.
5. Morris, L.L.; Fitz-Gibbon, C.T.; Freeman, M.E. *How to Communicate Evaluation Findings*; Sage Publications: Newbury Park, CA, 1987.

BIBLIOGRAPHY

1. Rossi, P.H.; Freeman, H.E. *Evaluation: A Systematic Approach*; Sage Publications: Newbury Park, CA, 1993.

Health Policy in Singapore

M. Ramesh

Lee Kuan Yew School of Public Policy, National University of Singapore, Singapore

INTRODUCTION

The *World Health Report 2000* ranked Singapore sixth overall in terms of health care quality, cost-effectiveness, and accessibility among the 192 countries evaluated. The interest in the island state's health policy that followed has generated a striking debate among analysts and practitioners in countries contemplating reform. While much of the discussion has focused on its financing arrangement, particularly the Individual Medical Savings Account, Singapore's system of providing health care is no less remarkable.

The objective of this entry is to describe the arrangements for providing and financing health care in Singapore and highlight relevant policy lessons. It will show that public provision of hospital care, combined with private financing, offers a viable reform alternative for other nations contemplating reform. The case is particularly compelling for public provisions because of the equity implications of private financing.

POLICY HISTORY

In the final years of colonial rule, the British established a reasonably well-developed health care system in which public and private care coexisted. In the 1950s, the colonial government became increasingly involved in the provision of inpatient care through public hospitals, partly reflecting the recently established National Health Service in Britain, through which hospital care was available to all without a means test.

The People's Action Party (PAP) government that took office in 1959 continued and expanded the arrangement. In addition, it massively expanded its commitment to promoting public health and hygiene, to the point that they were of First World standards by the 1970s. The large financial implications of publicly provided hospital care became increasingly apparent in the 1970s, as trends of population aging and epidemiological transition became apparent, prompting the government to search for alternative mechanisms for health care financing.

After briefly toying with the idea of establishing public health insurance, the government firmly rejected it in 1982 because of fears of third-party payment

would lead to over-supply by producers and over-use by consumers. The National Health Plan, announced in 1983, instead initiated a range of privatization measures designed to reduce the share of expenditures borne by the government and to encourage consumers to assume greater responsibility for their health care. At the same time, the government reiterated its commitment to providing heavily subsidized service to those who could not afford it.

The establishment of Medisave within the Central Provident Fund (CPF) in 1984 was the first in a series of measures aimed at reducing government involvement in the financing of health care. The scheme compulsorily covers the entire workforce, including the self-employed. It is a compulsory individual savings scheme that requires the working population to set aside a portion of their income for medical care. However, it cannot be used to pay for outpatient care, except for some specialist care, and there are limits on the amount that can be withdrawn each day and for each procedure. The exclusions and ceilings are intended to allow accumulation of funds to pay for expensive inpatient care when needed. For further information on the health system in Singapore, see Singapore, 2006.^[6]

In the same year that Medisave was established, the government announced a plan to grant greater autonomy to government hospitals in managing their operations. Operational autonomy was intended to lead to greater competition, which in turn was expected to lead to greater efficiency, lower costs, and higher standards.

In 1990, the government established a publicly managed, but voluntary, health insurance scheme called Medishield. It covers hospitalization expenses for surgery and outpatient treatment for specified catastrophic illnesses. Its establishment, despite the government's opposition to social insurance, was in response to the realization that most Medisave accounts did not have sufficient funds to pay for illnesses with expensive treatment costs. Medishield's premiums are kept low by imposing a large number of exclusions and a severe cost-sharing requirement to avoid the moral hazards associated with third-party payment schemes. The exclusion of the elderly (currently above 85 years), who need more health care, serves the same purpose.

The Health Review Committee's report of 1992, issued as a white paper the following year, supported expansion of the role of the private sector and greater competition among health care providers, both public and private. At the same time, it acknowledged, "Market forces alone will not suffice to hold down medical costs to the minimum The government has to intervene directly to structure and regulate the health system."^[5] It called for greater government involvement in controlling costs, managing the supply of doctors, and governing key operations of subsidized hospitals. In the following years, measures were taken to control: (1) the introduction of technology and specialist disciplines in government hospitals; (2) the imposition of price caps in public hospitals, excluding expensive treatment from public schemes; and (3) tightening the supply of doctors.^[2]

The realization that there was a segment of the population that could not afford even highly subsidized health care led the government to establish a public assistance scheme called Medifund in 1993. It serves as an endowment trust fund built on a S\$1 billion contribution from the government. To prevent dissipation of the fund, income from the fund is used only to pay the bills of those unable to afford hospital care. Under the scheme, patients in the lowest class wards at public hospitals and outpatients needing expensive services may apply for complete or partial waiver of their bills. To receive assistance, applicants need to pass a means test administered by the hospital.

PROVISION

The government is heavily involved in providing health care in Singapore and owns and operates over 80% of all hospital beds.

Public hospitals are nominally private firms, though they are owned entirely by a government-owned holding company. The hospitals and holding companies enjoy considerable operational autonomy. However, the government continues to play a vital role as a regulator, buyer, and, no less significantly, owner.

Public hospitals in Singapore offer a choice of different classes of ward accommodations, ranging from a one-bed room (Class A) to a dormitory with ten or more beds (Class C), at different prices. The different wards provide equivalent clinical services but different levels of comfort and physician choice. More significantly, patients in Class A wards pay the full cost, whereas charges in other classes range from 80% in Class B1 to 20% in Class C; the rest is subsidized by the government. Thus, the choice of ward class involves not only a choice of level of comfort but also the level of subsidy, because the two are linked. While there is no required means test for the subsidized

wards, it is expected that those who can afford better class wards will choose them in order to receive better nonclinical services.

The government provides only a quarter of outpatient care through polyclinics, and the rest is provided by the private sector. In addition to providing subsidized outpatient services, referral points exist for accessing subsidized specialist medical services. Through polyclinics, users pay a nominal charge at government clinics, and the elderly and students pay even less. The number of visits to government clinics, as a share of all visits, has declined since the early 1980s. However, these clinics serve as a price benchmark for private clinics and offer a viable alternative to those who may find private services unaffordable.

While Singapore's health care system is based on Western medical sciences, a significant portion of the population relies on traditional Chinese medicine and, increasingly, Indian Ayurvedic medicine. Such medicine serves to complement and supplement Western medicine rather than substitute for it. The government regulates this sector lightly, controlling only poisonous and dangerous substances.

FINANCING

Singapore has an elaborate financing system that consists of direct expenditure by the government, regulated private expenditure from Medisave and Medishield schemes, and out-of-pocket payments.

Table 1 shows that Singapore, and especially its government, spends little on health care in general. Despite comparable income levels, its national health care spending of US\$898 per capita and 4.3% of GDP is less than half of the Organisation for Economic Cooperation and Development (OECD) average. No less remarkable is that the government's share of the total is around one-half of the share in other high income countries.

The government's share of national health care expenditures is 30%. Remarkably, the government's share stood at 75% of the total in 1980, indicating a massive shift from public to private funding over the last two decades.^[4] Singapore's government spending on health care is only around 1% of its GDP, compared to an average 6% in OECD.

Government expenditures on health care include subsidy of public hospitals and outpatient clinics, capital expenditures of the Ministry of Health, and the cost of providing health care to state employees. Class C wards in public hospitals are the main vehicles for delivering subsidized health care to the general population. Patients pay only 20% of the cost of service in Class C wards, rising to 80% for Class B1 wards, and they pay the entire cost of Class A wards. Government subsidies

Table 1 Health care expenditures

	OECD		Singapore	
	1990	2002	1990	2002
GDP per capita, PPP (current international \$)	18,810	29,510	12,230	23,980
Total health expenditure per capita, current US\$	1,417	1,979	444	898
Total health care expenditure, % of GDP	7.1	8.5	3.3	4.3
Share of total which is public, %	61	59	30	30

World development indicators, 2005, <http://devdata.worldbank.org/wbquery/>.

on average form 60% of a hospitalization bill; the remaining 40%, on average, is recovered from patients.

Those unable to pay for lower classes and highly subsidized service in public hospitals may request a partial or total waiver, which was once funded from the government's general revenues, but is now funded indirectly through Medifund. In 2003, 18,000 Class B2/C patients received assistance totaling S\$19.2 million, or an average of S\$1,067 per person. The government also provides the bulk of public outpatient clinics' operating costs. Individual patients pay a modest sum for outpatient treatment, and the elderly and students pay even less.

Private health expenditures in Singapore take the form of indirect payment through Medisave or Medishield or directly out-of-pocket. Medisave requires a compulsory saving of 6%–8% of one's monthly income, depending on age, to be used for payments for the hospital care of the account holder and his or her immediate family, except siblings. In 2004, total Medisave balances amounted to S\$32.1 billion, or an average account balance of S\$17,321 per active member. The government claims that the average amount is sufficient for most people if they choose the lowest class ward for admission. Nevertheless, Medisave forms only 7% of total health expenditures, a figure that has remained largely constant despite the increase in average account balance.^[1]

Medishield premiums are modest for most people, but they can be substantial for the low-income, especially if they are also elderly. Premiums are small (S\$30–80) for those under the age of fifty, but then rise rapidly for those who are older. Despite the relaxation of conditions for claiming Medishield benefits, almost 50% of inpatient admissions do not rely on

it to pay for any portion of their hospitalization bill. The government estimates that, following changes introduced in 2005, Medishield pays for 70% of the average hospital bills for the lower class (previously only 40%).^[3]

OUTCOMES

Medicare, the compulsory savings scheme for medical expenses, covers two-thirds of the labor force; the excluded are, for the most part, foreign workers. Subscribers to Medishield, which is a public, but voluntary, health insurance scheme, account for just over one-half of the labor force, suggesting that more than one-half of the population is uncovered. Non-subscription to Medishield is particularly high among the elderly, the segment of the population who needs it most.

Targeting of subsidy, without administration of a costly means test, and extensive government involvement in the provision of health care has kept a lid on Singapore's health care costs. The government also plays a significant role in what doctors and hospitals charge for their service. The government has a plethora of arrangements in place to reduce costs at public hospitals.

Table 2 shows that Singapore has excellent health status compared to other wealthy countries.

Its infant mortality rate (IMR) is less than half the rate found in OECD, and its residents' average life expectancy is superior to the OECD average. Yet more remarkably, its IMR has declined and its life expectancy has increased much more rapidly than others in the OECD. Singapore's health system also performs

Table 2 Health status

	Infant mortality rate (deaths per 1000 registered live births)		Life expectancy at birth (years)	
	1970	2003	1970	2003
Singapore	19.7	2.9	68	79
OECD	22	5.8	71	78

World development indicators, 2005, <http://devdata.worldbank.org/wbquery/>; http://www.undp.org/hdr2003/indicator/indic_289.html.

very well when measured for Healthy Life Expectancy (HALE) at birth—67 years for men and 69 years for women—which ranks it significantly higher than the United States, the world's largest spender on health care.

The heavy reliance on private financing in Singapore is likely to have adversely affected equity and access, but no reliable information is available to confirm or deny the claim. This problem is likely to be especially severe for households whose health care needs are great but their income is higher than the very low income necessary to qualify for public assistance.

CONCLUSIONS

Singapore has an elaborate health care system consisting of a range of measures for providing and financing health care. The provision of inpatient care is dominated by the public sector, while outpatient care is dominated by the private sector. Financing for both inpatient and outpatient care is derived largely from private sources, though the government is intricately involved through its Medisave and Medishield schemes.

The fine health care status that Singapore has achieved at a rather modest cost reflects its systems of providing, as well as financing, health care. The dominance of inpatient care by public hospitals, who enjoy considerable operational autonomy while still under close government watch, allows the realization of the benefits of planning, as well as market competition.

While the health care financing arrangements have worked so far, their long-term viability remains a matter of concern. As the population ages, Singapore's

health care expenditures are projected to rise to levels similar to the OECD average by 2030. A clear need still exists for a more comprehensive risk-pooling arrangement, because households will be unable to carry the heavy financing burden expected under the current arrangements.^[1]

REFERENCES

1. Asher, M.G.; Nandy, A. Health financing in singapore: a case for systemic reforms. *Int. Soc. Secur. Rev.* **2006**, *59* (1), 75–92.
2. Barr, M.D. Medical savings accounts in singapore: a critical inquiry. *J. Health Polit. Policy Law* **2001**, *26* (4), 709–726.
3. Loy, Chua Boon. *Is Class B2/C Hospitalisation Affordable?* MOH Information Paper No. 2005/09, 2005.
4. Phua, K.H. Attacking hospital performance on two fronts: network corporatization and financing reforms in singapore. In *Innovations in Health Service Delivery: The Corporatization of Public Hospitals*; Preker, A.S., Harding, A., Eds.; The World Bank: Washington, DC, 2003; 451–484.
5. Singapore, Ministry of Health. *Affordable Health Care*, Singapore, author, 1993.
6. Singapore, Ministry of Health. *Statistics, Publications and Resources*, Singapore, author, 2006. <http://www.moh.gov.sg/corp/publications/index.do>.

BIBLIOGRAPHY

1. Taylor, R.; Blair, S. Financing healthcare: Singapore's innovative approach. *Public Policy for the Private Sector* **2003**, May.

Home Rule

Dale Krane

*Department of Public Administration, University of Nebraska at Omaha,
Omaha, Nebraska, U.S.A.*

INTRODUCTION

Home rule refers to the authority of a local government to make and implement a policy decision without first obtaining specific permission from a superior tier of government. In the United States, home rule is defined legally as the authority of a local government to exercise those powers granted by the state legislature, except where the exercise of a power would violate state law or the state constitution. Embedded in the idea of home rule is an important aspect of democracy and citizen sovereignty—policy decisions affecting a locality ought to be made by the citizens and officials of the locality. Of course, cities, counties, and other forms of local government exist within a larger framework of state governments and are expected to accommodate their actions to the interests of other localities as well as the state government. This is especially the case when the effects of a local decision “spill over” into another jurisdiction. Home rule, therefore, is a critical dimension of state–local relations in the United States, and as such, varies among the 50 states in terms of the types and extent of discretionary authority available to local governments. In some states, local governments possess little or no autonomy from state government control, while in other states, local governments exercise a broad range of policy making in areas such as their structure, functions, finances, and administration. The result of this inter-state variation is that the capacity of local governments to act in ways desired by their residents or to solve problems in a manner congruent with local conditions depends directly on the degree of home rule possessed by local governments in a particular state.

CREATURES OF THE STATE

The U.S. Constitution, while it contains numerous specific statements on the relationship between the national government and state governments, is silent on the place of local governments in the federal framework. The absence of any constitutional language referring to local governments means that the position and powers of local governments have been left to decisions by the states. Instead of granting local

governments substantial authority and capacity, states have acted to make local governments legally inferior jurisdictions.

In contrast to the federal relationship between the national and state governments, American local governments exist and function as units of government under the hierarchical control of state government. A local government, be it a municipality, a county, a special district, or a public authority, may act only in areas and in ways specified by state government. American legal theory, at least since the late 1860s, has held that local governments are “the creatures of state government.” Any action or means not permitted by state government is prohibited. If any doubt exists as to which level of government has the authority to act, this doubt must be resolved in favor of the state government. Furthermore, the existence and continuation of a local government rests with state government.

This “creatures of the state” theory of local government authority was promulgated in a precedent-setting decision by the Iowa Supreme Court in 1868. In the case of the *City of Clinton v. Cedar Rapids and Missouri Railroad Company*, Judge John F. Dillon ruled that municipalities “owe their origin to, and derive their powers and rights wholly from the [state] legislature,” and, therefore, are “mere tenants at will of the legislature.” This decision became known as Dillon’s Rule, and after it was upheld by the U.S. Supreme Court in 1903 and again in 1923, it has served as the principal legal opinion governing state–local relations.^[1]

Since the promulgation of Dillon’s Rule, the legal interpretation of home rule has become relatively standardized around the country. While the idea of home rule implies vigorous local government autonomy and self-determination, generally, state courts have held that local governments have only as much freedom and authority to govern their own affairs as the state, through its constitution and statutory laws, chooses to give to local governments. Unfortunately, Dillon’s Rule “often is interpreted improperly by observers who conclude the Rule centralizes all exercisable authority in the state legislature and local governments possess very limited discretionary authority. What these observers fail to recognize is the fact that the state legislature is free to grant broad discretionary

powers to local governments and several have done so.”^[2] That many state governments have granted substantial discretionary authority to local governments is an important trend that flies in the face of prevailing legal theory. In fact, the over-time accretion of significant powers by local governments has altered and even negated Dillon’s Rule in many states. So, it is important not to view the legal position of local governments as exactly the same in all 50 states.

THE HOME RULE MOVEMENT

Dillon’s Rule settled the question of the balance of powers between state and local governments in favor of state government. Other reformers who believed in more independence for local governments argued for a different solution to the question of just what was the proper balance of state and local government authority. These reformers campaigned to enact state constitutional changes that would permit cities to write and amend their own charters. This reform strategy was known as the Home Rule Movement and originally sought to end the interference of local state legislative delegations in municipal affairs by creating municipal charters for local governments. Eventually, the Home Rule Movement became associated with the idea of broad grants of local government autonomy.

In the 1840s and 1850s, local delegates to state legislatures used their position to write statutes that permitted the local legislative delegation to interfere in municipal decisions. So-called “local privilege” legislation often led to battles between the city council and the local legislative delegation. This state–local conflict sometimes provoked the state legislature to transfer control of certain activities of municipal government (e.g., police protection) to state-appointed officials. These laws became known as “ripper laws.” Beginning in the 1850s, some states such as Ohio and Iowa acted to prohibit “local privilege” legislation and to grant a modest degree of home rule to cities. But the state courts followed the “creatures of the state” theory, so reformers had to push for constitutional changes that altered the state–local relationship.^[3]

After the Civil War, those who sought to provide cities with constitutional protections from state legislative abuse understood that it would be impossible to obtain grants of substantive powers to cities without also conferring on cities the power to write their own charters. When Missouri wrote its 1875 constitution, it included the first-ever state constitutional provision to permit the drafting of municipal charters. Several other states soon followed Missouri and gave positive grants of authority to municipalities via various forms of general enabling laws and the option for home rule

charters. By 1912, constitutional language in the 13 states shared the common feature of establishing home rule through the use of a locally drafted charter that was to be ratified by local voters or by the city council.^[4]

Advocates of home rule benefited from an association with the Progressive reformers who sprung up on the Great Plains after the Civil War. The Progressive Movement’s emphasis on business-like government became the vehicle for further efforts to reform government after World War I, and the home rule idea continued to spread. In the 1920s, “Good Government” associations such as the National Municipal League disseminated model constitutional provisions which included home rule ideas. By 1937, 21 states had adopted some form of home rule charter authority.

The Home Rule Movement in its campaign for municipal and county charters raised the crucial question of “where was the line between state and local powers to be drawn?” New York State in 1923 tried to answer this question with a constitutional amendment that included a long list of items regarded as “local,” and thus, would be under local government control. Other states also adopted this strategy to enumerate “local affairs,” because this approach was thought to clarify the line between state and local interests, and as a consequence, would reduce the number of lawsuits initiated after the adoption of a home rule charter. This enumeration of a local affairs solution to state–local relations suffered from the inflexibility of constitutions. Once the list of state versus local powers was written into the state constitution, it was difficult to amend. State courts soon faced numerous lawsuits over whether some activity such as public health, street repair, or liquor licenses was a matter of local or state affairs. Finding a clear line was frustrating, and many state supreme courts found it easier to continue to adhere to Dillon’s Rule.^[5]

The American Municipal Association, forerunner to the National League of Cities, hired Dean Fordham in 1952 to devise a new model of state–local relations. He proposed a plan in which a state legislature would permit municipalities to write a charter with a broad list of powers. This new charter would allow municipalities to supersede special state laws and many general laws applicable to municipal corporations. In its essence, the Fordham “devolution of powers” plan proposed a reversal of Dillon’s Rule. Instead of the position that municipalities may exercise only those powers explicitly granted to them, municipalities under Fordham’s plan could act unless explicitly prohibited by state law. This model of home rule authority is sometimes referred to as “legislative home rule.” Several states have adopted this model.^[6]

Home rule for county government also made progress after World War II. In 1945, Missouri made

home rule charters available to its largest counties, and Washington enacted county home rule in 1948. By the end of the 1960s, county home rule had been established in 15 states. By the mid-1990s, 37 states (of the 47 states with viable counties) provided for some form of county home rule.

THE COMPLEXITY OF HOME RULE

Most treatments of home rule discuss the topic from a legal perspective. Yet, home rule refers to the actual discretion local officials exercise over various dimensions of local government. Usually, the powers granted by states to municipal or county governments are placed in three main classes: 1) governmental powers, which include the authority to enact ordinances, levy taxes, and exercise police powers; 2) corporate powers, which include the ability to enter into contracts, buy and sell property, and sue as well as be sued; and 3) proprietary powers, which allow local governments to own and operate commercial enterprises such as a utility. Which of these three classes of powers is granted to which type of local government varies from state to state.^[7]

The history of the Home Rule Movement shows that proponents of more local government authority focused their efforts primarily on amending state constitutions in order to create charters, or lists of power and authority possessed by a local government. The earliest charters typically provided local residents with the opportunity to select their own form of government (e.g., mayor-council, council-manager, or commission form). While the first charters granted by state governments were limited to the structural features of local government, over time, home rule advocates convinced state legislatures to broaden the scope of municipal (and county) charters to include other dimensions of local government. As state legislatures found that “local affairs” consumed more and more legislative time, state legislatures increasingly assigned responsibility for designated policy areas to local governments or permitted local governments more discretion over certain public functions. Likewise, state legislatures, often responding to demands from local government officials, granted more authority over administrative and financial matters to local governments. In recent years, states have even granted significant new authority to allow local governments to pursue economic growth.

All of these changes in the scope and types of discretionary authority that local governments possess and exercise demonstrate that a legal perspective to home rule is a narrow view of home rule. Any portrayal of local government authority that continues to paint a picture of local governments as “creatures

of the state” ignores the changes that have occurred in the relationship between state and local governments since 1970. Just as state governments modernized their structures and administration, so too have they enhanced the authority of local governments to make choices across a wide range of important activities. The litmus test of local government autonomy is discretion over fiscal matters. While it is true that states have continued to impose certain constraints on local fiscal authority, primarily restrictions on the use of the property tax, it is also true that many states have diversified the types of revenue sources local governments may choose to tap. Furthermore, states have granted local governments more discretion in how they spend their resources. As long as a legal lens is used to study home rule, the description will miss the complexity of the state-local relationship, and thus, miss much of the actual behavior of local governments.^[8]

CURRENT STATUS OF HOME RULE

Despite the enlargement of local government discretion during the last half of the twentieth century, the actions of state legislatures have often been negated by state courts that have been reluctant to overturn Dillon’s Rule. Since 1970, however, several states have acted to redefine the scope of local government authority. Illinois, for example, modified the devolution-of-powers model by moving in the direction of the enumeration strategy that had been tried during the “good government” era. That is, the Illinois constitution set forth complex decision rules related to the establishment of home rule, the writing and adoption of charters, the specific areas of local affairs including powers to regulate health, safety, morals, and welfare, and powers to tax and incur debt. Equally important, the Illinois constitution went to the heart of Dillon’s Rule by telling state legislators and judges how to interpret the list of local powers—“the powers and functions of home rule units shall be construed liberally.” By contrast, non-home-rule units and limited purpose governments (e.g., special districts, townships) “shall have only powers granted by law.”^[9]

The Illinois approach is significant, because it established a new model of home rule—the “liberal construction” model. By sorting out the list of local powers and stating them in language that the state courts and the legislature had to follow, Illinois lawmakers specified the conditions under which the state government could preempt a local government charter. If the specified conditions did not exist or pertain, then the state government (legislature or courts) could not preempt actions taken by a local government. Other states copied the Illinois language in recent revisions of their constitutions. The Alaska constitution is

No Home Rule n = 9

Alabama, Connecticut, Idaho, Indiana, Nebraska,
Nevada, New Hampshire, Virginia, West Virginia

Constitutional Home Rule Charter n = 22

Arizona, Arkansas, Colorado, Florida, Georgia, Hawaii*,
Iowa, Kentucky, Maine, Maryland, Massachusetts, Minnesota,
Missouri, New York, Oklahoma, Oregon, Pennsylvania,
Rhode Island, South Dakota, Tennessee, Texas, Utah

Legislative Home Rule Charter n = 6

California, Delaware, Mississippi,
North Carolina, Vermont, Washington

Liberal Construction n = 13

Alaska, Illinois, Kansas, Louisiana, Michigan,
Montana, New Jersey, New Mexico, North Dakota,
Ohio, South Carolina, Wisconsin, Wyoming

*Hawaii does not have municipalities; four counties constitute local government.
Source: Author's review of state chapters in (Ref. 11).

Fig. 1 Municipal home rule, by type.

direct: "... a liberal construction shall be given to the powers of local government units ..." and "[a] home rule borough or city may exercise all legislative powers not prohibited by law or by charter." Amendments to constitutions in Colorado, Louisiana, Montana, and New Jersey also added similar language.^[9]

The usual classification of the home rule authority found in the 50 states identifies three distinct models of local government authority: 1) Dillon's Rule states where no home rule exists; 2) home rule charter states where cities and often counties may adopt a charter with specific powers; and 3) "devolution of powers" states where local governments may act unless something is specifically prohibited.^[10] A fourth model of "liberal construction" should be added to these three different distributions of authority between state and local governments. This fourth model would include states where constitutional language has been written so as to instruct the state courts to interpret broadly the authority of local governments.

A recent compilation of the status of home rule in the 50 states describes in detail the variation in local government discretionary authority.^[11] Fig. 1 classifies state governments by the type of home rule available to municipal governments. Only nine states currently restrict their municipalities to a narrow interpretation of Dillon's rule. In 22 states, the state constitution

provides for the drafting of home rule charters by municipal governments. Six states make home rule power available to municipalities via grants by the state legislature. Thirteen states have adopted some version of the "liberal construction" approach to home rule. Other than those states in the first category [no home rule], the other 41 states permit municipal governments to exercise various combinations of discretionary authority, most typically discretion over the form of local government and over functional responsibilities. About a dozen states also grant their municipalities some degree of fiscal discretion. The existence of this considerable degree of variation in the extent and types of authority available to local governments means that local governments in many states are no longer "creatures of the state," but rather, these local governments are increasingly becoming "partners of the state" in the provision and delivery of services to citizens.

REFERENCES

1. Frug, G.F. The city as a legal concept. *Harvard Law Rev.* **1980**, 93 (6), 1059–1154.
2. Liebschutz, S.F.; Zimmerman, J.F. Fiscal dependence and revenue enhancement opportunities for local

- governments in the United States. In *Future Challenges of Local Autonomy in Japan, Korea, and the United States: Shared Responsibilities Between National and Sub-National Governments*; Horie, F., Nishio, M., Eds.; National Institute for Research Advancement: Tokyo, 1997; 186–207.
3. Zimmerman, J.F. *State-Local Relations: A Partnership Approach*, 2nd Ed.; Praeger Publishers: Westport, 1995.
 4. McBain, H.L. *The Law and the Practice of Municipal Home Rule*; Columbia University Press: New York, 1916.
 5. Alderfer, H.F. *American Local Government and Administration*; MacMillan: New York, 1956.
 6. Fordham, J.B. *Local Government Law: Text, Cases, and Other Materials*; Foundation Press: Brooklyn, 1949.
 7. Martin, D.L. *Running City Hall: Municipal Administration in America*, 2nd Ed.; University of Alabama Press: Tuscaloosa, 1990.
 8. Krane, D. Local government autonomy and discretion in the United States. In *The Challenge to New Governance in the Twenty-First Century: Achieving Effective Central-Local Relations*; Hoshino, S., Fosler, R.S., Eds.; The National Institute for Research Advancement and The National Academy of Public Administration: Tokyo, 1999; 267–298.
 9. Krane, D.; Rigos, P.; Hill, M., Jr. Introduction. In *Home Rule in America*; Krane, D., Rigos, P., Hill, M., Jr., Eds.; CQ Press: Washington, DC, 2001; 1–22.
 10. Weeks, J.D.; Hardy, P.T. The legal aspects of local government. In *Small Cities and Counties: A Guide to Managing Services*; Banovetz, J.M., Ed.; International City/County Management Association: Washington, DC, 1984; 25–41.
 11. Krane, D.; Rigos, P.; Hill, M., Jr. Appendix: home rule across the fifty states. In *Home Rule in America*; Krane, D., Rigos, P., Hill, M., Jr., Eds.; CQ Press: Washington, D.C., 2001; 471–491.

Human Relations Management Theory

Gerald T. Gabris

*Division of Public Administration, Northern Illinois University,
DeKalb, Illinois, U.S.A.*

INTRODUCTION

Human relations management theory encompasses a rich and diverse tradition of models, techniques, research findings, and ideas that often trace their roots back to the Hawthorne Experiments conducted during the late 1920s. Researchers implementing experiments at the Hawthorne plant of Western Electric (in Cicero, IL) placed two groups of employees doing the same work into separate rooms. One group was treated as the control, and the second was exposed to various experimental stimuli such as increased lighting, decreased lighting, rest pauses, and so on. The researchers, F.J. Roethlisberger of Harvard and W.J. Dickson of Western Electric management, expected the experiments to lead to different levels of performance for the experimental group.

This thinking was in vogue for the day, where classical organization theory conceived the organizational system as a mechanism that could be manipulated or readjusted to affect employee performance. To the amazement of the researchers, both groups increased their performance. Subsequent analysis of these unexpected results led Roethlisberger and Dickson to conclude that the experimental design was problematic, which allowed extraneous factors to enter the design that led to these unanticipated results.^[1]

What happened was that employees in the control and experimental groups were treated as special. They were given attention by management, separated from other employees, and encouraged to perform. The fundamental lesson that emerged from this early research suggested that employees who are given attention by management, who are treated as special, and who perceive their work as significant can become highly motivated and thus become more productive. While this insight today may seem rather commonsensical, back then, it was novel. Employees were not conceived as special but rather as expendable cogs in a machine who either performed their work or were terminated. The need to motivate employees to perform better was not part of the managerial equation at this point in time.

Needless to say, the results of the Hawthorne Experiments, once the reasons for them were better understood by managers and academicians, served as

a springboard for a panoply of new theories and approaches to management. This new way of thinking that gained impetus in the 1930s and 1940s can be broadly labeled “human relations management.” Subsequent research and theory development in the 1950s and 1960s provided further conceptual grounding to this management school of thought to a point where today, using human relations techniques and models is considered an almost taken-for-granted skill that effective managers have mastered. This essay will provide a brief snapshot of the kinds of human relations theories that have emerged over the years by discussing their commonalities and areas of difference. But first, we will endeavor to define this rather broad approach to management by identifying its core values.

HUMAN RELATIONS MANAGEMENT DEFINED

In fairness, there is no simple, all-encompassing definition for the human relations school of thought. Nonetheless, there are certain values and themes that tend to permeate the theories, techniques, and models connected with this way of thinking. These can be stated as follows.

Human relations management acknowledges the centrality of the employee as one of the important and, indeed, perhaps most important strategic factors within the organization.^[2] Here, it is management’s responsibility to treat employees with respect, dignity, and value in order to motivate and develop the full performance potential of employees. Often, but not always, this involves utilizing employees in decision making, expanding job responsibilities, delegating authority, and increasing employee autonomy, all aimed at maximizing an employee’s perception of self-worth. In other words, by striving to satisfy the higher psychological needs of their employees, organizations find that employees exhibit better performance and commitment to organizational goals and needs. Both win. On balance, employees are considered mature, self-directed individuals who will perform high-quality work if given the chance. Supervisors are expected to avoid rigid and overbearing invasive control methods

and to rely more on coaching, mentoring, and participative styles.

This broad description of human relations management incorporates elements of many different models and theories. All tend to support and affirm the concepts highlighted above. At this point, we will briefly consider several seminal contributions to human relations theory and how these ideas have become part of the day-to-day vernacular of public administration.

THE OHIO STATE UNIVERSITY STUDIES

In the 1950s, researchers at Ohio State were striving to finding a balance between tasks and relationships insofar as to how managers dealt with their subordinates. Much of this was based on the Leader Behavior Description Questionnaire (LBDQ).^[3] Later, Stogdill^[4] condensed this work into the concepts of “initiating structure” and “consideration.” The importance of this research involves its emphasis on structure and human relationships that effective managers need to master. Managers who solely stress structure, tasks, and production will not be as effective as managers who also strive to motivate employees by appealing to their human needs and emotions. What the Ohio State research was finding, was that the most effective managers blended and balanced an emphasis on task structure with consideration of the human needs of their subordinates. In short, effective managers needed to appeal to the higher psychological needs of employees.

Although now considered passé, the work of Abraham Maslow^[5] provided perspective on this point. Maslow’s “hierarchy of needs” made the case that what really motivates people the most is not basic, extrinsic, physical needs, but instead, higher psychological needs such as esteem and actualization. These are intrinsic. Managers should keep this in mind when trying to motivate employees. Strategies that only emphasize physical and extrinsic factors may satisfy basic needs for a short while but will not likely cultivate deeper motivational drives that keep employees going when they are not closely supervised.

DOUGLAS MCGREGOR’S THEORY X AND THEORY Y

Douglas McGregor’s *The Human Side of Enterprise* captured, perhaps better than most “human relations theories,” the basic core of this particular notion.^[6] This model dichotomized management style into two basic strategies called “theory X” and “theory Y.” This simple duality builds on Maslowian motivation theory and points out that if given the chance, most

employees will become reasonably productive because they want to, not because they are forced to.

Theory X assumes employees are only motivated by basic needs and, consequently, have relatively little motivation to grow and mature in their work environment. Theory X suggests that most employees are naturally lazy, docile, indolent, and incapable of performing complex tasks without direct supervision. Thus, supervisors need to utilize a variety of “control” techniques to keep employees in line so they have little option other than to comply with managerially determined objectives.

Theory Y provides an alternative model based on quite different assumptions. Here, employees are understood to be motivated by higher psychological needs. Employees want to grow, to be given challenging tasks, and to take on new responsibilities. By providing a work environment that enables employees to meet these higher needs, the organization not only satisfies the goals of the employee but also motivates the employee to a higher level of performance in achieving organizational objectives. Theory Y suggests that most employees want to work hard, and want to experience challenging jobs, but all too often, are treated like children and thus conform to the lower expectations placed upon them. Based on these concepts, McGregor concluded that by humanizing management, we create better employees and better organizations. This is the normatively correct path to follow.

BLAKE AND MOUTON: THE MANAGERIAL GRID

A somewhat related yet conceptually distinct theory developed by Robert Blake and Jane Mouton is referred to as the “managerial grid.”^[7] These authors provided a more practical application of the Ohio State Studies’ concepts of “initiating structure” and “consideration.” The managerial grid consists of two simple axes. The vertical axis measures “concern for people,” while the horizontal axis measures “concern for results.” A nine-point scale measures each axis, creating a grid.

Blake and Mouton then labeled specific grid intersects with benchmark descriptors to facilitate how a person’s score on their leadership questionnaire could be interpreted. For example, managers who have a high concern for results but a low concern for people (1, 9) are labeled “authority compliance” managers. Those concerned with people and not interested in results (9, 1) are called “country club” managers. The best managers in all situations are those who evidence a simultaneous concern for people and results (9, 9), and they are referred to as “team” managers.

This is the kind of management system to which all organizations should aspire.

RENSIS LIKERT AND THE HUMAN ORGANIZATION

Rensis Likert, a researcher at the University of Michigan, wrote a book in 1961 that further expanded our knowledge of human relations in organizations.^[8] As a survey researcher, Likert surveyed large numbers of employees in many different organizational settings. He is the developer of the “profile of organizational characteristics” survey questionnaire that serves as the basis for his theory and data.

Based on expansive data, Likert divided organizations into four systems. These were as follows:^[9]

System 1—Exploitative authoritarian.

System 2—Benevolent authoritarian.

System 3—Consultative.

System 4—Participative group.

Each system is composed of several factors, such as motivation techniques, supervisory style, incentive systems, and so on. Likert referred to these as managerial systems, because how an organization scores on one factor is predictive of how they score on other factors. Hence, organizational factors are interrelated parts. What Likert found was that most organizations score somewhere between systems 1 and 2 on average, but that when asked where they would like to be, most organizations respond systems 3 or 4. Thus, he infers that system 4, participative group, must be normatively the best, because that is where employees want to be if they can, almost all of the time.

In a vein similar to McGregor and Blake and Mouton, Likert made a strong case that humanizing the organization by increasing participation, expanding communications, appealing to higher psychological needs, and increasing responsibility contributes to the responsible freedom of employees and, therefore, should stimulate higher performance. As a consequence, organizations may wish to incorporate human relations management philosophy into their repertoire, not simply out of altruism but out of self-interest.

CONTINGENCY MANAGEMENT THEORY: HERSEY AND BLANCHARD

This essay will provide one final twist on the human relations perspective. Some scholars suggest that style of management is contingent on the situation, and that

perhaps there is no one best way to manage all of the time. Paul Hersey and Ken Blanchard^[10] utilized the same axis descriptors as Blake and Mouton, but they came to quite different conclusions. Their vertical axis is labeled “relationship” behavior, and the horizontal axis is labeled “task” behavior. How one scores on their Leadership Effectiveness Adaptiveness Description (LEAD) survey determined where one would be placed on their grid.

Hersey and Blanchard suggested that in some cases, employees are immature or ignorant on how to perform certain tasks. Here, participation makes little sense. Supervisors should simply tell the employees what to do. They call this S1 or “telling” management. Sometimes, employees may not know how to perform a task, but there is a need for them to work together cooperatively. Here, an S2 or “selling” style of management would be better. Alternatively, employees may know how to perform a task or have high task maturity, but they need to build relationships to work cooperatively on a project. In this case, an S3 or “participative” style might work best. Finally, if employees know how to perform a task, and they do not need to work with others to get it done, then a supervisor should use an S4 style or simply “delegate” the work.

The point is that Hersey and Blanchard realized that supervisory style and approach are somewhat contingent on conditions at work. They felt that a human relations approach may be useful at times but not necessarily in all situations. This approach represents a more sophisticated approach to understanding how human relations theory might be utilized in public organizations. It is a tool, an approach that encourages more flexibility and adaptativeness, in contrast to the earlier models discussed.

CONCLUSION

Human relations management is alive and well today. We probably take it for granted in most of the organizations we work in and for. Bosses today who exude little regard for people are typically considered old-fashioned autocrats who, due to insecurity, strive to bully people around. It is not surprising that in research conducted by Jay Hall^[11] that asked employees to rate their favorite supervisor and to specify their behavioral traits, those who received the highest “management achievement quotients” or MAQs (one’s hierarchical rank divided by chronological age) were those who practiced human relations management. Human relations management has become an ingrained assumption about how we should manage public organizations. While it may at times be difficult to define, you can tell it when you see it.

REFERENCES

1. Roethlisberger, F.; Dickson, W. *Management and the Worker*; Harvard University Press: Cambridge, MA, 1947.
2. Barnard, C. *The Functions of the Executive*; Harvard University Press: Cambridge, MA, 1935.
3. Hemphill, J.; Coons, A. Development of the leader behavior description questionnaire. In *Leader Behavior: Its Description and Measurement*; Stogdill, R.M., Coons, A.E., Eds.; Ohio State University Bureau of Business Research: Columbus, OH, 1957.
4. Stogdill, R.M. *A Handbook of Leadership: A Survey of Theory and Research*; The Free Press: New York, 1974.
5. Maslow, A. *Motivation and Personality*; Harper and Brothers: New York, 1954.
6. McGregor, D. *The Human Side of Enterprise*; McGraw-Hill: New York, 1960.
7. Blake, R.; Mouton, J. *The Managerial Grid*; Gulf Publishing: Houston, TX, 1964.
8. Likert, R. *New Patterns of Management*; McGraw-Hill: New York, 1961.
9. Likert, R. *The Human Organization*; McGraw-Hill: New York, 1967.
10. Hersey, P.; Blanchard, K. *Management of Organizational Behavior*; Prentice-Hall: Englewood Cliffs, NJ, 1982.
11. Hall, J. To achieve or not to achieve: the manager's choice. *Calif. Manage. Rev.* **1976**, *18* (4), 5–18.

Human Resource Management

Stephen E. Condrey

Ilka M. Decker

Carl L. McCoy

Carl Vinson Institute of Government, University of Georgia, Athens, Georgia, U.S.A.

INTRODUCTION

The field of public human resource management has endured a tumultuous yet challenging decade, with the emergence of reforms such as outsourcing, privatization, decentralization, reinventing government, new public management, and the like. At the center of these reform activities has been the call for human resource management to become a relevant and viable partner in the actual management of public organizations.

In a sense, human resource management is always in a state of reform; competing values such as responsiveness, economy, efficiency, and social equity play an important role in shaping the field. Lynn and Klingner noted that as the profession has evolved, so too has the role of the human resource manager:

(H)istorical traditions emphasize the technical side of personnel management, with less emphasis on policy-related analytical work, relationships with outside organizations, and decision making among conflicting values... A more contemporary view emphasizes different activities and relationships... For example, the modern personnel director might be called upon to prepare cost-benefit analyses of alternative pay and benefit proposals related to collective bargaining with employees in the solid waste department. At the same time, he or she might also be asked to evaluate the cooperative feasibility, productivity, and cost of privatizing or contracting the entire function out, thus making the collective bargaining analysis obsolete.^[1]

Human resource managers must adapt to these changes in the profession, all the while implementing effective and responsive human resource management systems in their organizations. In *Human Resource Management in Public Service*, Berman et al. defined human resource management as “a titanic force that shapes the conditions in which people find themselves... From deciding how individuals will be recruited to how they are then compensated, trained, and evaluated, human resource administration has a significant, even definitive, effect on the careers of all employees.”^[2]

OVERVIEW

Human resource management affects individual employees and their internal calculus as to whether and to what extent they will contribute toward organizational goals. Collectively, the system of human resource management present within an organization will positively or negatively affect the internal functioning of the organization and, ultimately, the quality of services it delivers to its clients (citizens).

This article is divided into two sections. The first section briefly discusses significant elements of an effective system of human resource administration. The second section provides a listing of reference sources in the field.

DESIGNING AN EFFECTIVE HUMAN RESOURCE SYSTEM

What makes an effective system of human resource management? While there is no simple answer to this question, it is nonetheless a critical issue to be explored. Indeed, an effective system of human resource management can enhance every area of administrative operations. Likewise, an ineffective or nonresponsive human resource management system can hinder administrative performance and have debilitating effects on an organization's climate, its performance, and its ability to attract and retain a quality workforce.

The following eight elements can be combined to help develop an effective public human resource management system:

- Systematic and meaningful employee involvement.
- Competitive compensation and benefits system.
- Responsive position classification system.
- Professional human resource management staff.
- Political and bureaucratic support.
- Integral part of the organization's management infrastructure.
- Aggressive employee recruitment and professional selection techniques.
- Effective and systematic training and development activities.

While this is a nonexhaustive listing, it begins to bring substance to the concept of an effective human resource management system. Following is a brief discussion of each of these elements.

Systematic and meaningful employee involvement in human resource decisions and processes can build the trust and support necessary for employee buy-in concerning organizational mission and goals. Additionally, a competitive compensation and benefits system will help attract and retain a quality workforce. An effective compensation system is predicated upon a flexible and responsive system of job classification. The classification system should assure “equal pay for equal work,” while also providing mechanisms to

recognize professional growth and development through the use of career ladders or broadbanding.

Necessary to any effective human resource management system is professional staff devoted individually or collectively to various aspects of the personnel system. Additionally, the organization’s political and bureaucratic leadership must be committed to a professional human resource management system and allow adequate funding for its development and maintenance. As a corollary, human resources should be viewed as an integral part of the organization’s management infrastructure. This entails the human resource management staff having input into important general management concerns. Because three-quarters

Table 1 Selected public human resource management reference sources

Texts:

<i>Human Resource Management in Public Service: Paradoxes, Processes, and Problems</i> (Sage, 2005)	Berman, Bowman, West, and Van Wart
<i>The New Public Personnel Administration</i> (Wadsworth Publishing, 2006)	Nigro, Nigro, and Kellough
<i>Public Personnel Management: Contexts and Strategies</i> (Prentice-Hall, 2002)	Klingner and Nalbandian
<i>Personnel Management in Government: Politics and Processes</i> (Marcel Dekker, 2001)	Shafritz, Riccucci, Rosenbloom, Naff, and Hyde

Readers:

<i>Classics of Public Personnel Policy</i> (Wadsworth Publishing, 2002)	Thompson
<i>Public Personnel Administration and Labor Relations</i> (M.E. Sharpe, 2006)	Riccucci
<i>Public Personnel Administration: Problems and Prospects</i> (Prentice-Hall, 2002)	Hays and Kearney
<i>Public Personnel Management: Current Concerns, Future Challenges</i> (Longman Publishing, 2005)	Ban and Riccucci
<i>Radical Reform of the Civil Service</i> (Lexington Books, 2001)	Condrey and Maranto

Handbooks:

<i>Handbook of Human Resource Management in Government</i> (Jossey-Bass, 2005)	Condrey
<i>Handbook of Public Personnel Administration</i> (Marcel Dekker, 1995)	Rabin

Journals:

<i>Review of Public Personnel Administration</i> (Sage Publications)	Section on Personnel Administration and Labor Relations, American Society for Public Administration
<i>Public Personnel Management</i> (IPMA-HR)	International Public Management Association for Human Resources

Websites:

Section on Personnel Administration and Labor Relations	www.aspaonline.org/spalr
International Public Management Association for Human Resources	www.ipma-hr.org
<i>Review of Public Personnel Administration</i>	rop.sagepub.com
Society for Human Resource Management	www.shrm.org
National Association of State Personnel Executives	www.naspe.net
United States Equal Employment Opportunity Commission	www.eeoc.gov

of most public organization's budgets are personnel related, it makes good managerial sense to involve the human resource staff in managerial decisions.

An aggressive system of recruitment and selection can help an organization recruit quality employees into the organization. Coupled with professional selection techniques such as assessment centers, an effective recruitment and selection program can build and strengthen an organization. Finally, effective and systematic training and development activities can further enhance organizational capacity.

DISCUSSION

Taken individually, each of the areas discussed above will not result in an outstanding human resource management department or function. Collectively, however, they signal a human resource department that is attuned to the needs of the organization's employees, its management purpose, and the citizens it serves. Building an effective human resource management system cannot be accomplished overnight; it requires internal competence and external support. As such, it is necessarily an incremental process, whereby major actors in the organization (governing body, management, and employees) gain confidence and respect in the human resource department and its ability to facilitate an effective and responsive system of personnel management. Those seeking an effective human resource system should not look to the latest management fad or quick fix, but rather concentrate on the proper resources (monetary and human) and buy-in of the organization's major stakeholders.

SOURCES AND RESOURCES

Table 1 lists various resources for those interested in further information concerning human resource management. The table lists useful texts, readers, handbooks, journals, and websites. While not exhaustive, this listing provides a sound starting point for discovering further information about public human resource management.

CONCLUSIONS

This article has provided a brief introduction to the challenging and evolving field of human resource management in public organizations. It is hoped that the reader will refer to the reference sources herein as a roadmap for further study in the field.

REFERENCES

1. Lynn, D.B.; Klingner, D.E. Beyond civil service: new roles and new tools for public personnel managers. In *Radical Reform of the Civil Service*; Condrey, S.E., Maranto, R., Eds.; Lexington Books: Lanham, MD, 2001; 198–199.
2. Berman, E.M.; Bowman, J.S.; West, J.P.; Van Wart, M. *Human Resource Management in Public Service: Paradoxes, Processes, and Problems*; Sage Publications: Thousand Oaks, CA, 2005; xiii.

BIBLIOGRAPHY

1. Condrey, S.E. *Handbook of Human Resource Management*; Jossey-Bass: San Francisco, 2005.

Human Resource Management for Productivity

Seok-Hwan Lee

*Department of Public Administration, Catholic University of Korea,
Puchon City, Republic of Korea*

INTRODUCTION

Human resource management for productivity refers to a comprehensive strategy that is designed to achieve productivity improvement by managing human resources in public organizations. Although productivity improvement in the public sector is also comprehensive both in concept and scope, it is possible to define productivity improvement in human resource management. From a perspective of effective human resource management, productivity is defined as the production of a series of efforts that encourage public employees to work hard in an efficient manner.

In addition, because productivity improvement is affected by various factors, ranging from top management support to feedback and correction on budget management decisions, productive human resource management requires human resource managers to pay special attention to four major elements in an open system perspective.

SELECTION AND COOPERATION AMONG RELATED AGENCIES: RIGHT PEOPLE, RIGHT TIME, AND RIGHT PLACE

Productive human resource management must start with recruiting the best and brightest. Because certain negative images (for example, corruption, red tape, and buffoonery) about public services and adverse circumstances prevail among citizens, attracting highly qualified people into the public sector may be harder than into the private sector. However, streamlined and flexible selection procedures can provide effective tools to hire highly talented people at the right time and at the right place.

Structured Application Forms

A structured application form that is focused on specific selection criteria and competencies provides an agency with a major head start in the short-listing stage of the recruitment process. Application forms must be tailored to a specific competency model that reflects agencies' professional needs.

Effective Cooperative Systems Among Related Agencies

So often, ineffective cooperation systems among related agencies hinder hiring agencies from carrying out their recruitment plan. Given that productive human resource management must be focused on the premise of "right people, right time, and right places," a lack of effective cooperation and the failure to eliminate unnecessary involvement by related agencies will delay the process and cause hiring agencies to lose the opportunity of having highly qualified and talented people in the competitive market.

Effective cooperation can benefit from using Internet technology that enables related agencies to communicate important information in a timely manner. Internet technology can certainly make a streamlined recruitment process possible by cutting time and costs, as well as raising an agency's ability to locate and identify qualified people in the market.

Continuous Recruitment Process

Flexibility in recruitment is essential to attract viable candidates from the limited pool. Furthermore, diversification of public employees will also be facilitated by continuous recruitment. Whenever there is a job vacancy, an agency should be able to hire the right people immediately. The longer the gap between vacancy and replacement, the more likely an agency is to lose competitiveness.

In an effective continuous recruitment process, hiring department managers will submit a "request to recruit" to the human resource department and develop a written rationale that the availability of qualified candidates is limited. The approval notice must be made as soon as possible.

Professional Needs and Process Needs: Shared Responsibility

A streamlined and flexible recruitment process also encourages greater strategic use of the agencies' people, as well as a better balance between the professional needs of employees and the process needs of

the agencies.^[1] When there is a need for human resources, the recruitment process requires both the human resource department and related departments to work together. The reason is that working together not only helps to identify whose qualification would most fit a specific department's need, but it also facilitates budgetary approvals to create the position in a timely manner.

In addition, although the human resource department is in charge of the whole recruitment process, it is also important to have related department managers involved in the process. Department managers can assist the recruitment effort through disseminating letters or announcements to colleagues in the field and soliciting similar assistance from current employees to attract the best-qualified people.^[2] Therefore, shared responsibility among departments is a necessary first step for a streamlined recruitment process.

NEED FOR CONTINUOUS EDUCATION AND TRAINING

Productivity also depends on those who have appropriate knowledge, skills, and abilities (KSAs). As missions and environments change, KSAs become outdated. Therefore, organizations must continually renew their human resources.^[3] Training is critical to improving performance and reduces the likelihood of employee obsolescence in a changing environment. Staff training and development efforts are important to achieving manpower plans, as well as to improving competencies of individuals. Successful continuous training and development programs that require all members to actively participate typically include providing:

- An opportunity for employees to meet on a regular basis, thus enhancing team spirit.
- A forum for discussion of problems encountered on the job.
- A means of continuous reinforcement of the core mission of the organization.
- Valuable information for use on the job.
- Necessary tools and feedback for personal development.

IDENTIFYING MULTIPLE MOTIVATIONAL BASES

Perhaps, motivating employees to work hard is the most important job of human resource managers. Because productive organizations depend heavily on productive workforces, it is essential to identify employees' motivational bases and encourage them to work at their peak. Among other motivational bases,

organizational commitment has been said to be the key to enhancing productivity in any organization. The importance of employee commitment is found in its relationship to performance and productivity. Much of the literature on organizational commitment finds that commitment to organization is positively associated with job satisfaction, motivation, and attendance, and negatively related to such outcomes as job performance, absenteeism, and turnover.^[4,5] It should be noted, however, that employees could also be differentially committed to occupations, top management, supervisors, co-workers, and customers.^[6,7]

Among other foci of commitment, commitment to a supervisor is an important motivational base for increasing individual performance. Because extrinsic rewards controlled by one's supervisor are a major means for directing and reinforcing managerial and executive behavior,^[8] those committed to their supervisor are likely to be motivated to improving productivity. It is also possible to expect that where a teamwork-based organizational culture prevails, people are likely to be committed to their work group, thereby increasing their willingness for group-based performance. Therefore, human resource managers should realize that employees' motivational bases are different from individual to individual. This indicates that simply focusing on how to enhance organizational commitment (commitment to organization) may not guarantee a high level of individual performance and productivity.

Productivity improvement so often requires innovative behaviors among employees. Existing literature points out that innovation in the public sector is a function of strong leadership, openness, freedom, and public service spirit. However, because innovation often requires risk-taking behaviors among employees, many public employees are not inclined to suggest innovative ideas, even if they have some ideas that may help to improve productivity in the workplace. From such a point of view, this study assumes that employee trust, along with employee commitment, is the key to encouraging innovative behaviors among employees.

On a general level, employee trust is defined as an expectancy of positive outcomes that one can receive based on the expected action of another party in an interaction characterized by uncertainty. In the meantime, employee commitment is defined as employees' psychological attachments to their workplaces. It has also been said that the level of trust increases that of employee commitment. Therefore, to encourage innovative behaviors among employees, it is essential to have a high level of trust, thereby increasing the level of commitment.

This study argues that previous research has mainly centered on a unidimensional approach to both

employee trust (organizational trust) and employee commitment (organizational commitment), and that distinguishing among different dimensions of trust and commitment will help to explain variations in innovative behaviors above and beyond those that are explained by both organizational trust and organizational commitment. In this regard, this study proposes that both employee trust and commitment are multidimensional constructs.

As for employee trust, employees can differentially trust their supervisor, their co-workers, and their organization. Each dimension of trust also has three subdimensions: competency trust, work-related trust, and family-related trust. With reference to employee commitment, employees can also be differentially committed to their supervisor, co-workers, and organization.

ORGANIZATIONAL JUSTICE

Productive human resource management should be able to deal with various issues associated with workplace and personal life. It is the employers' responsibility to maintain a working environment free from discriminatory harassment. Any form of unlawful discrimination, including harassment based on race, color, religion, gender, national origin, age, disability, or any other characteristic protected by applicable law, is strictly prohibited.

Employee Assistance Program

Crises in employees' personal lives obviously affect their performance at work. When organizations help employees to overcome these problems, they can improve employee morale and loyalty and avoid incurring costs to the organization.^[1] The Employee Assistance Program (EAP) can help employees with the challenges of balancing work and personal life. The EAP is designed to provide easy access and assistance with resolution for a wide variety of personal issues, such as alcohol and drug abuse, financial counseling, and domestic and workplace violence. The EAP directors must be able to coordinate the EAPs offered by the organization or by contract providers, as a response to personal employee problems that become workplace issues.^[3] An effective EAP requires easy access to professional, confidential service 24 hours per day, 7 days per week, on either phones or Internet sites.

Due Process

The Fifth and Fourteenth Amendments to the Constitution ensure that the state and national government

cannot deprive individuals of life, liberty, or property, except by established judicial procedure. These due process clauses provide the people with several means for the procedural protection of their fundamental rights from arbitrary administrative action. Its ultimate goal is to secure the fairness in a feasible and reasonable way.

Due process in the legislative rule-making setting probably requires a fair-minded administrative official and adequate advance notice of the content of the rule to permit compliance. However, the agency that chooses to develop its substantive policies through case-by-case adjudication must conduct itself, in each case, so as not to violate the applicable procedural due process requirement that courts have defined for varying adjudicative situations.

In this regard, the adjudicative action does require the due process, such as a hearing, because:

1. The government action is not general and affects a relatively small number of persons.
2. The immediate effect on the class is significant and direct.
3. The government action is directed at specific person.

In the meantime, legislative action does not require due process. The characteristics are as follows:

1. The government action is general and affects equally all persons within the class.
2. Immediate effect on the class is significant but indirect.
3. The government action is not directed at a specific person.
4. The government action is "general determination dealing only with the principle" on which all assessments are based.

CONCLUSION

From an open system perspective, productivity improvement requires comprehensive efforts, ranging from internal top management support to feedback from citizens. Among others, effective utilizations of human resources are crucial for productivity improvement.

An organization without human commitment is like a person without a soul: skeleton, flesh, and blood may be able to consume and to exert, but there is no life force. Government desperately needs life force.^[9] Thus, recognizing and responding to emergent human resource management demands is directly connected to improving productivity in public organizations.

Productive organizations are humane and show genuine concern for meeting employees' needs.

Productive human resource management in the twenty-first century is the management that focuses on balancing both employees' and the organization's needs.

REFERENCES

1. Holzer, M.; Callahan, K. *Government At Work: Best Practices And Model Programs*; Sage: Thousand Oaks, CA, 1998.
2. Hickman, G.R.; Lee, D. *Managing Human Resources in the Public Sector: A Shared Responsibility*; Harcourt College: Orlando, FL, 2001.
3. Klingner, D.E.; Nalbandian, J. *Public Personnel Management: Context and Strategies*, 4th Ed.; Prentice Hall: Upper Saddle River, NJ, 1998.
4. Mathieu, J.E.; Zajac, D.M. A review and meta-analysis of the antecedents, correlates, and consequences of organizational commitment. *Psychol. Bull.* **1990**, *108*, 171–194.
5. Balfour, D.L.; Wechsler, B. Organizational commitment: antecedents and outcomes in public organizations. *Public Prod. Manage. Rev.* **1996**, *19*, 256–277.
6. Becker, T.E. Foci and bases of commitment: are they distinctions worth making? *Acad. Manage. J.* **1992**, *35*, 232–244.
7. Reichers, A.E. A review and reconceptualization of organizational commitment. *Acad. Manage. Rev.* **1985**, *10*, 476–485.
8. Perry, J.L.; Wise, L.R. The motivational bases of public service. *Public Adm. Rev.* **1990**, *50*, 367–373.
9. Minzberg, H. Managing government, governing management. *Harvard Bus. Rev.* **1996**, *74*, 75–83.

Human Resource Planning

Joan E. Pynes

Department of Government and International Affairs, University of South Florida,
Tampa, Florida, U.S.A.

INTRODUCTION

Human resource planning (HRP) provides agencies with realistic information on the capabilities and talents of their current staff, as well as what knowledge, skills, abilities, and other characteristics (KSAOCs) will be needed by the organization in the future. Agencies with information that enables them to quickly adapt to changes can be provided by HRP.

Over 46% of local government employees are 45 years or older, while 30% of state government employees nationwide are eligible for retirement in 2006, and by 2008 more than 50% of federal employees will be eligible for retirement.^[1] Montgomery County, VA estimates that 50% of its senior managers will be eligible for retirement in 2010. By 2007, more than 70% of Fairfax, VA's top officials will be able to leave and collect benefits.^[2]

HRP refers to the implementation of human resource activities, policies, and practices to make the necessary ongoing changes to support or improve the agency's operational and strategic objectives. Agency leaders need to understand how their workplaces will be affected by impending changes and prepare for the changes accordingly. Agency objectives should be formulated after relevant data on the quantity and potential of available human resources have been reviewed. Are there human resources available for short- and long-term objectives? To be competitive, organizations must be able to anticipate, influence, and manage the forces that impact their ability to remain effective. In the service sector, this means that they must be able to manage their human resource capabilities. All too often agencies have relied on short-term service requirements to direct their human resource management (HRM) practices. Little thought is often given to long-term implications. By invoking HRP, agencies are better able to match their human resource requirements with the demands of the external environment and the needs of the organization. The human resource focus is not just an individual employee issue; it also focuses on integrating human resources into the organization's strategy. It becomes part of the visionary process. Strategic planning, budgeting, and HRP are linked together.^[3]

HUMAN RESOURCE PLANNING

In *The Case for Transforming Public-Sector HRM*, the National Academy of Public Administration identified five steps that were imperative in aligning human resource management to an agency's mission.^[4]

Step 1: Include human resource in strategic plans: HRM strategies needed to address agency goals must be discussed; and the resources needed to implement those strategies must be identified.

Step 2: Define human resource requirements: Workforce planning requires line managers and HRM experts working as partners to identify gaps that exist in the current workforce when compared to future requirements. A workforce planning system must be in place that enables the agency to understand the competency gaps to be addressed so that the agency can achieve its strategic goals. The assessment process will typically include documenting workforce demographics; calculating the current and projected attrition rates for key occupations; assessing the efficacy of primary sources of labor supply in both qualitative and quantitative terms; developing strategies for building the human capital needed to successfully achieve future mission goals; and evaluating progress, making mid-course corrections, and fixing accountability for achieving the desired outcomes.

Step 3: Develop action plan for implementing human resource strategies: An action plan should establish accountability for implementing the plan and identify the resources needed to complete the plan. The plan should also provide solutions to close any competency gaps that have been identified. To what extent is the agency able to attract and retain the expertise it needs to accomplish its strategic goals? Does the agency have the right people to meet mission and strategic goals?

Step 4: Evaluate progress: It is important to determine progress, make mid-course corrections, and assign accountability for achieving the desired outcomes. An evaluation provides a regular flow of information about how well HRM strategies are working. Information should be provided on a regular basis so informed decisions can be made about an agency's staff. Information about recruitment, retention, morale, and training and development should be analyzed.

Step 5: Manage the change process: Top management and human resource professionals' work together to develop a successful change process. Top management identifies the desired change, the reason for change, the expected benefits of the change, and the impact the change will have on employees. The leadership of the agency provides the necessary resources and allows time for the changes to occur, but holds staff accountable for performance.

In an effort to be proactive, the U.S. General Accounting Office developed *Human Capital: A Guide for Assessing Strategic Training and Development Efforts in the Federal Government*.^[5] During the planning and analysis stage, agencies review the following questions: 1) Does the agency have training goals and related performance measures that are consistent with its overall mission, goals, and culture? 2) To what extent do the agency's strategic and annual performance planning processes incorporate human capital professionals in partnership with agency leadership and other stakeholders in addressing agency priorities, including training and development efforts? 3) How does the agency determine the skills and competencies its workforce needs to achieve current, emerging, and future agency goals and missions and identify gaps, including those that training and development strategies can help address? 4) How does the agency identify the appropriate level of investment to provide for training and development efforts and prioritize funding so that the most important training needs are addressed first?

HRP is the process of analyzing and identifying the need for and availability of human resources to meet the organization's objectives. Forecasting is used to assess past trends, evaluate the present situation, and project future events. Forecasting and planning complement each other as forecasts identify expectations, while plans establish concrete goals and objectives. Forecasting has become increasingly important as a large segment of the public work force is inching toward retirement.

Agencies must consider how to allocate people to jobs over long periods. Attempts must be made to anticipate expansions or reductions in programs or other changes that may affect the organization. Based on these analyses, plans can be made for the recruitment and selection of new employees, the shifting of employees to different programs or units, or the retraining of incumbent employees.

A *demand forecast* anticipates the workforce that will be needed to accomplish future functional requirements and carry out the mission of the organization. In this step a visionary staffing assessment against future functional requirements is conducted. The result is a forecast of the type of competencies, numbers, and locations of employees needed in the future. An important part of the demand forecast is examining not only

what work the agency will do in the future, but also how that work will be performed. Some things to consider include:

- How will jobs and workload change as a result of technological advancements, economic, social and political conditions?
- What are the consequences or results of these changes?
- What will be the reporting relationships?
- How will divisions, work units, and jobs be designed?
- How will work flow into each part of the organization? What will be done with it?

Once the above questions have been answered, the next step is to identify the competencies employees will need to carry out that work. The set of competencies provides management and staff with a common understanding of the skills and behaviors that are important to the agency. Competencies play a key role in decisions on recruiting, employee development, personal development, and performance management.

Forecasting human resource requirements involves determining the number and types of employees needed by skill level. First, agencies need to audit the skills of incumbent employees and determine their capabilities and weaknesses. Positions must also be audited. In most organizations there are likely to be jobs that are vulnerable, that technology or re-engineering are ready to replace. Job analyses must be conducted to provide information on existing jobs. The basic requirements of a job should be defined and converted to job specifications that specify the minimum KSAOCs necessary for effective performance. The skill requirements of positions do change so any changes that occur must be monitored and reflected in the job specifications. It is not enough to monitor changes in positions. Organizations must also keep abreast of the skills that their employees possess. It (HRP) uses data inventories to integrate the planning and utilization functions. Data inventories compile summary information, such as the characteristics of employees, the distribution of employees by position, employees' performance, and career objectives. Specific data that are typically catalogued are age, education, career path, current work skills, work experience, aspirations, performance evaluations, years with the organization, and jobs for which one is qualified. Expected vacancies due to retirement, promotion, transfer, sick leave, relocation, or termination are also tracked. Using a computerized human resource information system to compile these data makes the retrieval of information readily available for forecasting work force needs.

When forecasting the availability of human resources, agencies need to consider both the internal

and the external supply of qualified candidates. The internal supply of candidates is influenced by training and development, and by transfer, promotion, and retirement policies. Assessing incumbent staff competencies is critical. Agencies should undertake an assessment of employees' competency levels. This will provide information for determining the number of those available and capable of fulfilling future functional requirements. It will provide salient information as to what recruitment, training, and other strategies need to be deployed to address workforce gaps and surpluses. The County of Los Angeles Department of Human Resources has developed a Strategic Workforce Planning Guidebook.^[6] The Guidebook is intended to provide departments with a common definition of Strategic Workforce Planning as it relates to the County's Strategic Plan; a case study to illustrate an actual Strategic Workforce Plan; a framework for departments to conduct a four-part needs assessment of current and future requirements; and developments of their department and a list of strategies, methods, and tools that departments can utilize when developing their workforce plans and be equipped to complete their Needs Assessment Workbook.

A *succession analysis* should be prepared that forecasts the supply of people for certain positions. Succession plans should be used to identify potential personnel changes, to select backup candidates, and to keep track of attrition. The external supply of candidates is also influenced by a variety of factors, including developments in technology, the actions of competing employers, geographic location, and government regulations. A number of local governments have begun to examine succession planning for their executive positions. They have put together talent banks listing incumbent employees' current skills and areas of expertise, job preferences, and possible new skills the employees may need in order to be promoted. Succession planning becomes more important as the labor supply tightens and the unemployment rates hit low levels. Los Angeles County has developed a Training Academy to enhance the skills of its current staff. A variety of certificate programs are offered that encompass leadership, problem solving, and creativity training.

HRP attempts to match the available supply of labor with the forecasted demand in light of the strategic plan of the organization. A gap analysis is the process of comparing the workforce demand forecast with the workforce supply projection. The expected result is the identification of gaps and surpluses in staffing levels and competencies needed to carry out future functional requirements of the organization. A gap occurs when the projected supply is less than the forecasted demand. It indicates a future shortage of needed employees. Strategies such as recruitment, training, succession planning will need to be developed and implemented.

A surplus is when the projected supply is greater than forecasted demand. This indicates future excess in some categories of employees that may also require action to be taken. The surplus data may represent occupations or skill sets that will no longer be needed in the future or at least not to the same degree. Retraining, transfers, or separation incentives may need to be implemented to address surplus situations.

HRP involves the development of strategies to address future gaps and surpluses. There are a variety of factors that will influence which strategy or which combination of strategies should be used. Washington State has identified some of the factors:^[7]

- *Time*: Is there enough time to develop staff internally for anticipated vacancies or new competency needs, or is special, fast-paced recruitment the best approach?
- *Resources*: The availability of adequate resources will likely influence which strategies are used and to what degree, as well as priorities and timing.
- *Internal depth*: Does existing staff demonstrate the potential or interest to develop new competencies and assume new or modified positions, or is external recruitment needed?
- *"In-demand" competencies*: How high the competition is for the needed future competencies may influence whether recruitment vs. internal development and succession is the most effective strategy, especially when compensation levels are limited.
- *Workplace and workforce dynamics*: Whether particular productivity and retention strategies need to be deployed will be influenced by workplace climate (e.g., employee satisfaction levels), workforce age, diversity, personal needs, etc.
- *Job classifications*: Do the presently used job classifications and position descriptions reflect the future functional requirements and competencies needed? Does the structure of the classification series have enough flexibility to recognize competency growth and employee succession in a timely fashion? Does it allow compensation flexibility?

If necessary skills do not exist in the present workforce, employees will need to be trained in the new skills or external recruitment must be used to bring those skills to the organization. The employer must identify where employees with those skills are likely to be found and recruitment strategies must be developed.

IMPLEMENTING HUMAN RESOURCE PLANNING

To implement HRP, it is important for HRM departments to expand their scope of activities beyond the

traditional HRM tasks and functions, and enter partnerships with managers and employees. Working together provides managers and employees a better understanding of HRM issues. Likewise, HRM staff becomes more informed about the needs of the employees and departments.

To facilitate the partnership, HRM Departments should establish a human resource planning task force composed of managers and/or staff from a variety of departments, and staffs from the HRM department who are responsible for identifying the trends and challenges that will impact the agency. The task force should involve other employees seeking their input and observations. This can be done through the use of surveys, focus groups, agency publications, or a needs assessment instrument. After the task force receives information, it should prioritize the most critical issues. HRM staff should be directed to develop strategies, solutions, and alternatives to address the issues.

Questions that should be investigated and planned for include:

- Is the top administration committed to quality and excellence?
- Are changes necessary?
- How can we meet employees' perceptions and concerns?
- Which employees will be affected?
- What barriers might be there to successful implementation of HRP?
- Once program cost estimates are developed, will dollars be provided for training and development?
- Who will be responsible for the implementation of HRP?
- Who will evaluate and adjust the planning process?

Once these questions have been answered, authority and resources must be assigned to the person(s) responsible for the planning and implementation of HRP. Failing to put human and financial resources into HRP will relegate it to just another fad. Organizational leaders must demonstrate their commitment to HRP and champion it.

CONCLUSION

HRP determines the human resource needs of the agency and ensures that qualified personnel are recruited and developed to meet organizational needs. Turnover, including retirements, must be anticipated and planned for. The HRM departments must track the skills of incumbent employees and keep skill inventories. Recruitment and training must be tied to the organization's mission. Should there be a shift in demand for services; agencies must know whether there are potential employees with the requisite skills available to provide these services and whether the agency's finances can afford the costs associated with additional compensation and benefits. Forecasting an agency's human resource supply reveals the characteristics of its internal supply of labor; it also helps to assess the productivity of incumbent employees, implement succession planning and salary planning, and identify areas where external recruitment and/or training and development are necessary.

REFERENCES

1. Ibarra, P. The myths and realities of succession planning. *IPMA-HR News* **2006**, August, 13, 15
2. Turque, B. Graying of workforce troubles county governments. *The Washington Post* **2006**, February 26, C06.
3. IPMA-HR. *Workforce Planning Resource Guide for Public Sector Human Resource Professionals*; Author: Washington, DC, 2002.
4. National Academy of Public Administration. *The Case for Transforming Public-Sector Human Resources Management*; Author: Washington, DC, 2000.
5. U.S. Government Accounting Office. *Human Capital: A Guide for Assessing Strategic Training and Development Efforts in the Federal Government*; Author: Washington, DC, GAO-04-546G, 2004.
6. County of Los Angeles, Department of Human Resources. *Strategic Workforce Planning Guide*; Author: Los Angeles, CA, 2005.
7. Washington State Department of Personnel. *State of Washington workforce planning Guide: Right people, right jobs, right time*; 2000; Retrieved 9/11/2006, from www.wa.gov, <http://hr.dop.wa.gov/workforceplanning>.

Human Resources Administrators in State Government, Changing Roles of

Curtis R. Berry

*Department of Political Science, Shippensburg University,
Shippensburg, Pennsylvania, U.S.A.*

INTRODUCTION

Numerous challenges and expectations confront human resources administrators in state government today. The “devolution” of governmental responsibilities by the federal government has shifted increased responsibility for domestic policy to the states. A sluggish economy has created, in several states, intense budgetary pressures and the need to cut expenditures or impose tax increases. Terrorist threats and workplace violence increasingly make the headlines and contribute to increased anxiety in the workplace. Today more than ever state personnel agencies are experiencing increased pressure to provide effective, meaningful, and timely assistance to agencies of state government.

GROWTH IN STATE BUREAUCRACIES

In May 2002, the seasonally adjusted number of persons employed by the 50 states totaled some 4,930,000. Increases in state employment levels were recorded in 9 of the 11 years between 1990 and 2000 and by 10.2% between 1991 and 2000. These increases, however, were not evenly distributed across the 50 states or among agencies or programs within the states. Between October 1990 and April 2000, 36 states experienced a net increase in full- and part-time employment, with 26 states experiencing growth of 10% or greater and six states experiencing growth of 20% or more in their level of employment. Since the fall of 2001, numerous states experienced fiscal stress brought on by the recession (and the events of September 11, 2001). Commentators were once again focusing on the phenomena late hired, first fired and the implications that it holds for minority and female employment (Table 1).^[1,2]

ORGANIZATION OF STATE PERSONNEL AGENCIES

With the exception of Texas, all states have a centralized personnel function, although significant differences exist in their structure and scope. In 47 states, a single

personnel office has primary responsibility for most activities associated with the personnel function. In some, one or more functions (e.g., labor relations or appeals of adverse actions) may be assigned to a separate organization. Some offices have responsibility for serving only state agencies or departments; other have additional responsibilities for educational institutions and/or local government administrative units. Two states, Pennsylvania and California, divide responsibilities associated with directing the state merit system into separate agencies. In about one-half of the states, the personnel function operates as a stand-alone administrative entity, whereas in the remaining states it functions as a division of a larger administrative structure, such as an Office of Administration. The chief personnel official in most states (26) is selected by the governor; officials in 15 states are appointed by the department head of the agency in which the personnel office is located; in the remaining states, a personnel board makes the appointment. Finally, the number of employees the office is responsible for ranges from about 10,000 in New Hampshire to 188,000 in New York.^[3–5]

FUNCTIONS OF THE CENTRAL PERSONNEL AGENCY

Functions of the central personnel agencies alone or shared with operating agencies included activities such as recruitment, test development and administration, issuing hiring rules, classification and compensation administration, labor-management relations, and appeals of personnel actions. The range of functions and roles handled by the central personnel agency in the states is shown in [Table 2](#).

Recruitment

Recruitment of personnel is a vital function with central personnel agency normally developing and disseminating recruitment literature, and maintaining a pool of applicants certified for appointment for the various state agencies, educational institutions, and local governments served by the office. The Internet

Table 1 State government employment growth (decline), 1990–2000

State	No. of state employees 1990	No. of state employees 2000	Percentage change	No. per 10,000 population 1990	No. per 10,000 population 2000	Percentage change
1 Kentucky	84,177	86,756	3.06	228	214	−6.14
2 Arizona	60,674	77,372	27.52	165	150	−9.09
3 Idaho	22,818	28,993	27.06	227	224	−1.32
4 Texas	258,905	305,957	18.17	152	146	−3.95
5 Utah	43,012	58,689	36.45	250	262	4.80
6 Missouri	86,507	110,243	27.44	169	197	16.57
7 Hawaii	57,595	67,184	16.65	520	554	6.54
8 Washington	112,110	139,936	24.82	230	237	3.04
9 Colorado	68,854	79,546	15.53	209	184	−11.96
10 Louisiana	99,572	113,456	13.94	236	253	7.20
11 Connecticut	66,939	74,942	11.96	204	220	7.84
12 Arkansas	49,245	59,954	21.75	209	224	7.18
13 Mississippi	52,854	60,402	14.28	205	212	3.41
14 North Dakota	20,081	20,541	2.29	314	319	1.59
15 Delaware	24,878	27,330	9.86	374	348	−6.95
16 North Carolina	122,535	142,791	16.53	185	177	−4.32
17 Florida	180,597	202,883	12.34	140	126	−10.00
18 Pennsylvania	150,008	149,753	−0.17	126	121	−3.97
19 New Hampshire	21,011	24,395	16.11	189	197	4.23
20 California	389,805	434,856	11.56	131	128	−2.29
21 Georgia	123,249	136,908	11.08	190	167	−12.11
22 Ohio	171,742	176,694	2.88	158	155	−1.90
23 South Carolina	87,724	92,621	5.58	252	230	−8.73
24 Tennessee	91,811	93,706	2.06	188	164	−12.77
25 Alabama	92,124	95,841	4.03	228	215	−5.70
26 Oregon	62,221	67,791	8.95	219	198	−9.59
27 Indiana	106,536	104,667	−1.75	192	172	−10.42
28 West Virginia	39,407	39,085	−0.82	220	216	−1.82
29 Oklahoma	78,006	87,807	12.56	248	254	2.42
30 Montana	22,807	24,046	5.43	285	266	−6.67
31 Virginia	141,309	142,469	0.82	228	201	−11.84
32 New Mexico	51,535	53,012	2.87	340	291	−14.41
33 South Dakota	17,300	16,588	−4.12	314	219	−30.25
34 Rhode Island	24,274	24,291	0.07	242	231	−4.55
35 Iowa	62,445	63,330	1.42	225	216	−4.00
36 Minnesota	84,898	79,982	−5.79	194	162	−16.49
37 Wyoming	12,679	13,506	6.52	279	273	−2.15
38 Nevada	21,705	25,752	18.65	181	128	−29.28
39 Kansas	57,824	53,381	−7.68	233	198	−15.02
40 Nebraska	35,751	36,002	0.70	227	210	−7.49
41 New Jersey	125,430	146,480	16.78	162	174	7.41
42 Alaska	25,021	26,011	3.96	455	414	−9.01
43 Vermont	14,743	14,276	−3.17	261	234	−10.34

(Continued)

Table 1 State government employment growth (decline), 1990–2000 (*Continued*)

State	No. of state employees 1990	No. of state employees 2000	Percentage change	No. per 10,000 population 1990	No. per 10,000 population 2000	Percentage change
44 Illinois	170,438	167,761	–1.57	149	135	–9.40
45 Maine	26,659	24,692	–7.38	217	193	–11.06
46 Michigan	177,721	172,515	–2.93	191	173	–9.42
47 New York	305,475	275,780	–9.72	170	145	–14.71
48 Maryland	101,522	95,523	–5.91	212	180	–15.09
49 Massachusetts	107,901	105,623	–2.11	179	166	–7.26
50 Wisconsin	90,367	75,399	–16.56	185	140	–24.32
Year average	1990 No. of state employees 90,056	2000 No. of state employees 95,950	Percentage change 6.55	1990 No. per 10,000 population 225.74	2000 No. per 10,000 population 212.26	Percentage change –5.97

Source: Adapted from Refs.^[43] and ^[44].

has substantially increased the potential audience of their recruiting efforts in recent years. Central personnel agencies' web sites linked to their state's "Home Page," as well as nationally focused web sites, such as "U.S. Government Info/Resources/Jobs in State Government: State Employment and Job Openings Web Sites," "State Job Search," and "State Government Jobs," to name a few, have opened up an unparalleled audience for state recruitment efforts in the twentyfirst century.^[6–8]

Test Development and Administration

Decentralized testing and assessment of applicants has been widely discussed; however, it generally remains the responsibility of the central personnel office. Although Wisconsin has received a lot of attention for decentralizing testing, the actual extent of that decentralization is somewhat limited. Test development and validation is a highly technical, time-consuming, and expensive endeavor. It must be undertaken in a manner consistent with the requirements of federal civil rights laws, as well as specific and perhaps unique requirements of the agencies being served. In addition, the "client" agency may assist in developing procedures or standards used by panels who conduct oral interviews of applicants for vacant positions. The central personnel agency has responsibility for certifying test results and applying the veteran's bonus points to passing test scores of qualifying applicants. Finally, civil service lists of "eligibles for appointment" must be continually updated.^[3]

Hiring Process Constraints

Political choices play a significant role in the policies associated with hiring employees. Hiring constraints

include the desire for a socially representative bureaucracy at a time when there is a backlash against the use of affirmative action, which for some implies "reverse discrimination." Popular initiatives, legislation, or gubernatorial executive orders have restricted or banned the use of affirmative action, causing states to rethink their strategies for achieving a workforce representative of the society it serves. Many have adopted policies that are less aggressive or threatening to the interests of majority group males, effectively stunting efforts to achieve a representative workforce in the short term. Today, the focus is on "workforce diversity," a concept open to interpretation as to meaning and extent of effort in pursuing a representative bureaucracy.^[9–11]

Another hiring constraint can be found in statutory rules mandating reliance on the so-called "Rule of 3." A number of jurisdictions have sought ways to expand the pool of persons eligible for consideration through interviews. States have expanded the pool of applicants under consideration by adopting a "Rule of 5" (e.g., Alaska, Kentucky, Maryland, Nevada, South Carolina); a "Rule of 6" (Maine); a "Rule of 7" (Washington), a "Rule of 10" (Alabama, Arkansas, Idaho, Ohio, Oklahoma, West Virginia); or a "Rule of 15" (Delaware, Missouri). In Connecticut and Kansas, anyone who passes the examination may be interviewed. In some states, agencies may hire applicants based on the specific requirements of the position.^[3]

"Veterans' preference" systems represents yet another major constraint on the hiring process used by agencies to fill vacancies. Typically, the preference entails awarding of "bonus points" to the test scores of qualifying veterans. At least 42 states award bonus points of 2, 5, or 10 points to qualifying veterans. Disabled veterans or veterans with service/combat-connected disabilities may receive bonus points of 10,

Table 2 Functions and responsibilities of central personnel agencies in the states

- Establishes minimum qualifications
- Merit system testing
- Provides HRIS
- Human resources planning
- Classification
- Position allocation
- Compensation
- Recruitment
- Selection
- Performance evaluation
- Position audits
- Personnel function audit of agencies
- Employee assistance and counseling
- HR development
- Training
- Employee health and wellness programs
- Affirmative action
- Labor and employee relations
- Collective bargaining/labor negotiations
- Grievance and appeals
- Alternative dispute resolution
- Employee incentive programs
- Productivity systems
- Conducts employee attitude surveys
- Dependent care
- Workers' compensation
- Group health insurance
- Deferred compensation
- Drug and alcohol testing
- Retirement
- Employee promotions
- Provide budget recommendations to the legislature

Source: Adapted from Ref.^[4]

15, or 20 points, depending on the state in which they are seeking employment. At least seven states have an “absolute preference” in the appointment of personnel.^[3,12]

Position Classification

Classification is typically a function that the central personnel agency shares with operating agencies or those who contract with it for services. The number of separate classifications found in a state can be enormous, with some 6169 existing in New Jersey in the

mid-1990s. Several factors contribute to this situation. States with large workforces responsible for a multitude of activities typically have larger numbers of classes. Growth in the number of classes is also linked to “grade creep” and efforts to circumvent the restrictive nature of the compensation plan linked to the classification system. The frequency with which the overall system undergoes a comprehensive updating significantly affects the number of classifications. Finally, absolute veterans’ preference may cause agencies to prepare narrow position descriptions and classification specifications to mitigate against the negative impact such systems have on sound personnel selection procedures.^[12,13]

Managing the Compensation Plan

Revisions to the compensation plan require that the agency address several important concerns. It must carefully consider how many steps or levels will exist in the plan. An appropriate balance must be struck between longevity and qualifications. Further, if employees are guaranteed step increases plus cost of living adjustments during their career with the organization, they may become complacent. Consequently, the organization needs to prudently consider how much of a percentage change there should be between increments in the salary plan. Generally, the larger the increments, the more time an employee should remain within a pay grade.^[9,14]

To minimize the future implications of negotiated compensation increases, the agency may seek payment of a one-time only “bonus” to employees in lieu of an increase in the base pay structure. Such a strategy allows management to claim that it has addressed the needs of employees to keep pace with inflation in a particular year, but without making the long-term commitment of modifying the base. By limiting growth in the base, the employer’s related expenses for social security, Medicare, pensions, health and welfare benefits, and the like are held constant at the current level. Employee unions will likely resist proposed bonuses because future compensation levels and other benefits are not positively enhanced or adjusted to keep pace with inflation.^[15]

Controlling Benefit Costs and Increasing the “Value” of Benefits

Employers are increasingly asking employees and/or their unions to choose between improvements in salary/wage levels or preserving existing benefits and minimizing the shift to employee contributions (e.g., higher deductibles or copayments) or reductions

in levels of coverage. Imposing higher deductibles represents a significant cost containment strategy. The greater the deductible, the lower the premium the employer and/or employee pays. Another tool involves use of an offset, which represents a benefit paying something less than 100% of the lost income or expense replacement. Coordination of benefits has become commonplace and entails ascertaining if a scheduled payment can be reduced or rejected because it is covered by a spouse's plan. Finally, a coinsurance/copayment benefit provision in which the employee generally must pay a fixed percentage of the cost of service provided by a benefit on each occurrence of use is another significant cost containment mechanism.^[16,17]

"Cafeteria benefits" programs allow employees to personalize the range of benefits and levels of coverage contained in their compensation packages. "Flexible spending accounts" represent yet another option to assist employees in tailoring a benefits program to meet their unique needs. This form of "cafeteria benefits" allows an employee to set aside "pretax" dollars into a specific fund for recurring or routine medical, dental, and dependent care expenses. Funds are then deducted from the account to pay the incurred expenses. The principle advantage of these accounts is that neither income nor social security taxes are paid on the funds set aside.^[16,18]

Early Retirement Incentives and Cost Savings

Early retirement incentives and "window" periods have been created to enable employees to separate from the personnel system prior to the normal retirement age, with fewer years of service than normally required to qualify for full pension entitlements and/or with "bonus" credits allowing the retiree to enjoy a higher monthly pension. In considering such a strategy, several issues must be considered in determining its appropriateness. First, determining which employees are to be targeted and how the incentive will affect their decision to leave the system. Second, deciding whether the objective is to permanently eliminate positions or merely replace more "expensive," older personnel with younger, less expensive ones. Additional administrative expenses and resources associated with implementing the early retirement effort must also be anticipated and accounted for. Finally, these incentives may not yield the cost savings expected when initially considered. Careful analysis must be undertaken to avoid a situation in which, rather than producing substantial financial savings, the effort ends up haunting the jurisdiction for years to come and creates nearly insurmountable barriers to future early retirement incentive efforts.^[14,19,20]

Labor Relations and Collective Bargaining

Collective bargaining rights for state employees exists in some 28 states plus the District of Columbia. Strikes by public employees are strictly prohibited, with severe penalties imposed in the event of an illegal strike or job action in all but 13 states that authorize a "limited right to strike." The central personnel agency may bear responsibility for contract negotiations with a number of separate bargaining units (and unions), representing a wide range of employees, negotiating in succession, with each attempting to gain a better deal than the one arrived at previously by other units. Once negotiated, the office advises line agencies on matters related to the implementation and enforcement of the "master" collective bargaining agreement in a wide array of units that may have local agreements or "side letters," which have an impact on the bargaining relationship.^[21-24]

Appeals of Adverse Actions

Adverse action appeals may be handled through administrative reviews in the operating agency and/or the central personnel agency. When the facts are in dispute, hearings before an impartial decision-maker leads to factual determinations and recommendations for resolution. These "preliminary" determinations are subject to final review by the agency head and/or merit protection panel operating in the jurisdiction. Critics of the merit system cite the amount of time and documentation necessary to terminate public employees which appears to be generally low. Delays are often a function of a backlog of cases and the inability to schedule hearing until months after petitions are filed. Requests for continuances by one or both of the parties and the need for multiple hearing dates may also push back the final decision date.^[3,12,25]

Outsourcing and Privatization

Privatization efforts in the state government can range from "modest"—such as reliance on clerical employees provided by a temporary employment service while a permanent employee is on maternity or military leave—to private sector construction and lease back of facilities, to construction, staffing, and operation of correctional facilities. Appropriate administrative controls need to be developed prior to implementation of privatization initiatives. Promised savings should be closely scrutinized to ensure that they are real and that private delivery of such services are not more expensive than the public systems they replace. Additional staff may have to be hired and trained to draft complex

contracts that cover the specific services for which government is seeking private sector bids. Poorly crafted requests for proposal and contract language may result in vendors delivering far less than what advocates of privatization expected. Contract compliance officers must be employed to monitor and oversee the implementation of contracts to ensure that government is getting that for which it has paid. Agencies need to develop tools to effectively and accurately measure performance, and account for all relevant costs to realistically determine if tax dollars are being saved and the public is being served better.^[26]

Preserving Merit System Principles

The federal government has been a driving force behind the implementation of merit-oriented personnel systems in state government. Today, such standards are set forth for recipients of grants-in-aid. The standards cover recruitment and selection of personnel promotions, equitable compensation, access to training opportunities, continued employment based on adequate performance, identification of prohibited partisan abuses, and fair treatment in all aspects of personnel administration.^[12]

Beginning in 1976, a trilogy of U.S. Supreme Court cases,^[27-29] addressing “freedom of association” protections began to have a profound impact on the manner in which elected officials and political appointees carry out or influenced the process of staffing and maintaining bureaucracy. These decisions make it essential for personnel departments to clearly identify those positions that are legitimately policy making in character and are therefore among those organizational slots that may be properly filled on the basis of political sponsorship.

A final factor restricting political sponsorship or spoils patronage are the standards for imposing discipline in collective bargaining agreements. Such agreements often cover persons occupying non-managerial positions that were once the prize to be claimed after the election was won. Contracts specifically limit dismissals to a “just cause” determination subject to appeal through the grievance procedure culminating in binding arbitration. These provisions coupled with the union’s “Duty of Fair Representation” create a powerful incentive to challenge the legitimacy of terminations, which smack of patronage dismissals.^[19,30,31]

Several steps can be taken to minimize organizational exposure to judicial challenges on the basis of alleged patronage employment practices. The personnel agency should develop and disseminate clear statements concerning patronage practices and nepotism in the workplace. Such pronouncements should clearly state the organization’s commitment to “neutral

competency.” Positions which are non-policy making should be clearly identified, and the organization should state its commitment to identifying qualified personnel for such vacancies.^[32]

EMPLOYEE RELATIONS ISSUES CONFRONTING THE CENTRAL PERSONNEL AGENCY

Nepotism and Workplace Romances

Nepotism policies prohibit relatives and spouses from being employed in the same organization, the same office, or in a supervisor-subordinate relationship. Such policies exist to protect the “merit principle” by preventing the appointment of employees on the basis of family connection rather than competence. In addition, such policies are intended to prevent the potentially disruptive consequences that may arise when married couples work together. Restrictive policies may adversely affect the employment opportunities of dual career couples. Organizations may minimize inequities created by such policies by clearly defining the family relationships [i.e., spouses (including “partners or companions”), children, parents, brothers, sisters, blood relatives] covered by the policy. The policy should then stipulate that these persons are not be permitted to work in the same department, under the direct line of supervision of each other (first and second line), or be involved in decisions affecting each other in areas involving compensation, benefits, promotion, performance appraisal, etc.^[33,34]

Personnel offices have also increasingly had to deal with employees who date one another or who cohabit. Promulgating policies that balance the employer’s legitimate interest and concern in maintaining an efficient and effective workplace while respecting an employees right to privacy is problematic. Office romances may undermine employee morale if coworkers believe a colleague is receiving special treatment as a result of their relationship with a superior. The employer has legitimate concerns about the possibility of the romance turning sour and the subordinate employee raising allegations that they are being sexually harassed or punished for ending the relationship. Non-fraternization policies will not prevent office romances from developing, but well-thoughtout policies can minimize the negative potential of such relationships for the organization.^[35,36]

Sexual Harassment Policy

The central personnel agency must develop and promulgate specific policies that assist organizations to

identify and root out sexual harassment whenever it arises in the workplace. Central personnel offices should, in consultation with employee representatives, develop and disseminate to all agencies, departments, and bureaus under its jurisdiction a clear and unequivocal policy against all forms of sexual harassment. Education and training programs for supervisors and managers are a key component in the success of the program. All employees should be informed of what constitutes harassment, their right to raise the issue, and how they may file a complaint. An in-house grievance or complaint procedure, with appropriately trained individuals to hear and investigate fully and fairly all claims of sexual misconduct, should be developed. Staff responsible for conducting the investigation should be sensitized to the special aspects of sexual harassment cases. Investigators should listen and take complaints seriously. Immediate steps should be taken to stop the harassment and rectify any wrongs that have been documented. When the evidence warrants, transfers, training and education or disciplinary action should be directed against persons engaging in harassing conduct.^[37,38]

Addressing Violence in the Workplace

Various steps can be taken to minimize threats to employees in the workplace. The personnel office must develop and promulgate a comprehensive policy on workplace violence that provides clear direction to agencies. Included would be information that helps employees to identify the warning signs that an employee, former employee, client, or vendor is potentially violent. The policy should also provide direction to employees to help them to minimize personal threats they may encounter and to deescalate potentially violent situations. In addition, the policy should provide a detailed reporting mechanism to document and communicate to superiors threats or incidents that employees observe, receive, or encounter on the job. Finally, a “zero tolerance” policy with respect to employees bringing weapons to the workplace should be adopted, which includes a provision for a “measured response.”^[39,40]

Potential threats to the workplace may be addressed by preemployment background checks on prospective employees and inquiries about prior incidents of violence from previous employers. Persons with a history of workplace threats or violence obviously should not be employed if a potential for similar behavior in the future exists. In addition, the central personnel agency, in consultation with those responsible for facilities management should evaluate security measures and use of barriers, security devices, etc., to minimize threats to employees.^[39,40]

Qualified Immunity/Civil Suits for Damages

Central personnel offices are responsible for disseminating clear statements to all personnel that violating the constitutional or statutory rights of public employees, agency clients, or persons incarcerated or subject to the authority of the state’s paroling authority is intolerable. If employees knew or should have reasonably known that the actions taken were in violation of someone’s rights, they may be found liable for damages. Employees should receive appropriate training regarding their job-related actions that may create legal liability. In the current legal environment, employees enjoy what might be characterized as “qualified immunity.” As such, if a violation of protected rights occurs, they may be sued personally for actual and punitive damages associated with rights violated. When allegations arise regarding employee involvement in violations of protected rights, the employer should conduct an independent internal investigation to determine if any of its policies or rules were violated that might subject the employee to discipline or discharge.^[41,42]

THE INTERNET: MASTERING AN EMERGING RESOURCE

The Internet has the promise of making it considerably easier for the agency to fulfill its mission. Recruitment efforts, for example, can and have been enhanced as agencies post information about position vacancies, civil service position announcements, testing requirements, submission of applications, and the like. Test development and validation efforts have the potential for significant improvement due to the ability to have nearly instantaneous access to information from operating agencies, professional associations, personnel agencies in other states, and the federal government. The central personnel agency staff has instant access to information from federal databases, and professional counterparts in other states without ever having to telephone anyone or send a written request for information. As the knowledge explosion continues, the Internet affords personnel agencies a low cost means of keeping pace with available information.

REFERENCES

1. Bureau of Labor Statistics, Bureau of Labor Statistics data. *Nonfarm Payroll Statistics from the Current Employment Statistics, National Employment, Hours, and Earnings, Seasonally Adjusted: State Government*; U.S. Department of Labor: Washington, DC, 1998. <http://146.142.4.24/cgi-bin/surveymost>, 7-13-98, 12:43 PM.

2. Bureau of Labor Statistics, Bureau of Labor Statistics data. *Table B-1 Employees on Nonfarm Payrolls by Industry*; U.S. Department of Labor: Washington, DC, 2002. <http://www.bls.gov/news.release/empsit.t11.htm>, 6-12-02, 1:29 PM.
3. Sheibley, L.B. *Survey of Other States: Decentralization of Personnel Functions*; Pennsylvania Civil Service Commission: Harrisburg, PA, October 21, 1997.
4. National Association of State Personnel Executives. *State Personnel Office: Roles and Functions*, 3rd Ed.; The Council of State Governments: Lexington, KY, 1996.
5. Council of State Governments. *The Book of the States, 1996-1997 Edition, Volume 31*; The Council of State Governments: Lexington, KY, 1996.
6. U.S. Gov Info/Resources, "State Government Job Resources"; @ <http://usgovinfo.about.com/blstjobs.htm>, 6-12-02, 1:49 PM. Copyright © 2002 About, Inc.
7. *Employment Spot*, "State Job Search" @ <http://www.employmentspot.com/state/>, © 2001, StartSpot Mediaworks, Inc., 6-12-02 @ 2:55 PM.
8. *The Internet Job Source*, "State Government Jobs" @ <http://www.statejobs.com/gov.html>, 6-12-02 @ 2:45 PM.
9. Sylvia, R.D.; Meyer, C.K. *Public Personnel Administration*, 2nd Ed.; Wadsworth Publishing Company, Inc.: Belmont, CA, 2002.
10. Ban, C., Riccucci, N.M., Eds.; *Public Personnel Management: Current Concerns, Future Challenges*, 3rd Ed.; Addison, Wesley Longman, Inc.: New York, NY, 2002.
11. Ridge, T. Governor Commonwealth of Pennsylvania. *Executive Order 1996-9, Equal Employment Opportunity*; Office of Administration, Governor's Office, Commonwealth of Pennsylvania: Harrisburg, PA, December 20, 1996.
12. PA Legislative Budget and Finance Committee. *Study on Civil Service Reform: Pursuant to Senate Resolution 1977-14*; PA Legislative Budget and Finance Committee: Harrisburg, PA, April 1998.
13. States Update Personnel Systems. *State Trends Bulletin*; April/May, 1995; Vol. 1 (3), pp. 1, 6-7.
14. Klingner, D.E.; Nalbandian, J. *Public Personnel Management: Contexts and Strategies*, 4th Ed.; Prentice-Hall, Inc.: Englewood Cliffs, NJ, 1998.
15. Mayer, D.F., *Comments of Dr. Donald F. Mayer, Chairperson Negotiating Team, 1998*; Association of Pennsylvania State College and University Faculties, June 1998.
16. McCaffery, R.M. *Employee Benefit Programs: A Total Compensation Perspective*; PWS-Kent Publishing Company: Boston, MA, 1988.
17. Rosenbloom, J.S.; Hallman, G.V. *Employee Benefit Planning*, 2nd Ed.; Prentice-Hall, Inc.: Englewood Cliffs, NJ, 1986.
18. Henderson, R.I. *Compensation Management: Rewarding Performance*, 4th Ed.; Reston Publishing Company: Reston, VA, 1985.
19. Rabin, J., Vocino, T., Bartley Hildreth, W., Miller, G.J., Eds.; *Handbook of Public Personnel Administration*; Marcel Dekker, Inc.: New York, NY, 1995.
20. PA Public Employee Retirement Commission. *Feasibility of Early Retirement Incentives in the Public Sector*; Public Employee Retirement Commission, Commonwealth of Pennsylvania: Harrisburg, PA, March, 1995.
21. Kearney, R.C.; Carnevale, D.G. *Labor Relations in the Public Sector*, 3rd Ed.; Marcel Dekker, Inc.: New York, NY, 2001.
22. Public Employee Department, AFL-CIO. *Public Employees Bargain for Excellence: A Compendium of State Labor Relations Laws*; Public Employee Department, AFL-CIO: Washington, DC, 1997.
23. Kenyon, C.F. *Comments of related to personnel and collective bargaining issues within the Commonwealth of Pennsylvania*; Spring 1998.
24. Office of Administration, Bureau of Personnel. *Pay Plan 1996: M525.2*; Commonwealth of Pennsylvania, Office of Administration, Bureau of Personnel: Harrisburg, PA, July 12, 1996.
25. Ridge, T. Governor Commonwealth of Pennsylvania. *Governor's Annual Workforce Report 2001*; Governor's Office, Commonwealth of Pennsylvania: Harrisburg, PA, 2001.
26. U.S. General Accounting Office. *Report to the Chairman, House Republican Task Force on Privatization: Privatization: Lessons Learned by State and Local Government*, GAO/GGD-97-48; U.S. General Accounting Office: Washington, DC, March, 1997.
27. *Richard J. Elrod, et al. v. John Burns, et al.*; 427 U.S. 347, 96 S. Ct. 2673; 49 L. Ed. 2d 547; 1976 U.S. LEXIS 109; 1 BNA IER CAS 60 (1976).
28. *Peter Branti, as Public Defender of Rockland County v. Aaron Finkel and Alan Tabakman*; 445 U.S. 507; 100 S. Ct. 1287; 63 L. Ed. 2d 574; 1980 U.S. LEXIS 4; 1 BNA IER CAS 91 (1980).
29. *Cynthia Rutan, et al., v. Republican Party of Illinois, et al.*; 497 U.S. 62, 110 S. Ct. 2729; 111 L. Ed. 2d 52; 1990 U.S. LEXIS 3298; 58 U.S.L.W. 4872; 5 BNA IER CAS 673 (1990).
30. Act 1970-195, The Public Employee Relations Act, 43 PA Statutes Annotated, Section 1101.101 et seq.
31. Elling, R.C. Bureaucracy—maligned yet essential. *Politics in the American States: A Comparative Analysis*, 6th Ed.; Gray, V., Jacob, H., Eds.; Congressional Quarterly Press: Washington, DC, 1996.
32. PA Legislative Budget and Finance Committee. *PA Turnpike Commission Performance Audit: Issues Related to Turnpike Organization, Management and Operations: Report II*; PA Legislative Budget and Finance Committee: Harrisburg, PA, July 1997.
33. Tompkins, J. *Human Resource Management in Government*; HarperCollins College Publishers: New York, NY, 1995.
34. PA Fish and Boat Commission, (March 27, 1995). *Nepotism Policy, Memo to all Managers and Supervisors*, from Executive Director, Pete Colangelo, PA Fish and Boat Commission, Harrisburg, PA.
35. Winning, E.A. *Romance in the Workplace*; E.A. Winning Associates: Walnut Creek, CA, 1998, <http://www.ewin.com/articles/romance.htm>, 8-11-98 @ 10:54 AM.
36. Levine, Daniel. S. *Dangerous Liaisons: Why Your Employer Wants To Tell You With Whom You Can Sleep*

- And Why They Probably Don't, Disgruntled Feature*; 1995; Vol. 1, No. 4. <http://www.disgruntled.com/love.html> (June 15, 1998 @9:48 AM).
37. Greenbaum, M.L. Arbitrator/Mediator and former President Society of Professionals in Dispute Resolution, Essex, MA. Sexual Harassment in the Workplace, In *Presentation at the 14th Annual Summer Program of the American Arbitration Association, Ogunquit, Maine*; July, 20, 1992.
 38. U.S. Merit Systems Protection Board. *Sexual Harassment in the Federal Workplace: Trends, Progress, Continuing Challenges*; U.S. Merit Systems Protection Board: Washington, DC, November 1997.
 39. Williams, H.A.H. Violence in the workplace: a reality for men and women. Calif. Labor Employ. Law Q. Spring 1997, <http://www.calbar.org/2sec/3lab/labs7-11.htm> 6-25-98 @ 1:50 PM.
 40. Defense Personnel Security Research Center for the Private Sector Liaison Committee of the International Association of Chiefs of Police. *Combating Workplace Violence—Guidelines for Employer's and Law Enforcement*; The International Association of Chiefs of Police: Alexandria, VA, 1995, <http://www.amdahl.com/ext/iacp/pslcl.toc.html> 6-25-98 @ 12:32 PM.
 41. Rosenbloom, D.H.; Carroll, J.D.; Carroll, J.D. *Constitutional Competence for Public Managers: Cases and Commentary*; Peacock Publishers, Inc.: Itasca, IL, 2000.
 42. *Bryce Harlow and Alexander Butterfield v. A. Ernest Fitzgerald*; 457 U.S. 800, 102 S. Ct. 2727, 73 L. Ed. 2d 396 (1982).
 43. PA Legislative budget and finance committee. In *1992 Statistical Digest*; PA Legislative Budget and Finance Committee: Harrisburg, PA, 1992.
 44. PA Legislative budget and finance committee. In *2001 Statistical Digest*; PA Legislative Budget and Finance Committee: Harrisburg, PA, 2001.

Human Resources Management in Local Government

J. Edward Kellough

School of Public and International Affairs, University of Georgia, Athens, Georgia, U.S.A.

INTRODUCTION

In the broadest sense, human resources management encompasses everything connected with the management of people in organizations. More specifically, it is that area of activity associated with the recruitment, selection, training, compensation, and discipline of employees. These functions are central to effective management; in the public sector, which is heavily labor intensive, that significance is magnified. Furthermore, in local governments, jurisdictional size and the nature of certain public services provided will shape the context of human resources management in important ways. Small local governments, for example, may lack the capacity necessary for in-house delivery of critical human resource management functions, and in some locations where the legal environment is permissive, public employee unions play an important role in shaping the nature of human resource management policies. Some local government services, including police and fire protection, have long histories of union activity.

certified as eligible for employment and have their names placed on a list or register in descending order by examination score. Those selected for employment are required to come from the top of the list. In many jurisdictions, rules will limit selection to among the top three (or four or five) on the register.

Of course, the integrity of selection processes organized in this fashion depends in large part on the validity of the examinations used. In this context, validity refers to the extent to which the examination used actually measures job qualifications. The most widely used approach to assessing examination validity is based on the notion that an examination for selection should reflect job content and should do so proportionately. For example, an examination for a clerical position might assess a job candidate's word processing skills, provided that the job requires word processing as part of its ordinary activity. It is not unusual for smaller local governments to use examinations developed and validated by consultants or other larger public organizations, given the limited resources often available in smaller jurisdictions.

FUNCTIONS IN HUMAN RESOURCES MANAGEMENT

Employee Selection

One of the most significant human resources management functions is employee selection. Obviously, it is in the interest of local governments, or any organization, to select employees who are capable and well qualified. In the public sector, of course, a responsibility to the citizenry makes that objective even more imperative. Typically, public jurisdictions structure selection processes through merit systems designed to ensure that employees and applicants for employment are screened on the basis of their ability to perform the jobs at issue. The focus of activity is on open competitive examinations of qualifications. Examinations may be written or non-written, and may include assessments of job knowledge, skills, abilities and, in certain instances, physical capabilities. Often, the selection process involves a combination of several examination techniques. Individuals who pass an examination are

Job Analysis, Job Evaluation/Classification, and Compensation

Job analysis, job evaluation/classification, and compensation are also major human resources management functions. Job analysis is the process of systematically examining and determining job content. The purpose is to identify the tasks and activities involved in a particular job, the amount of time devoted to specified tasks, and the kinds of knowledge, skills, and abilities required to perform the job effectively. Once this information is collected, it may be used to inform selection procedures and the construction of examinations, employee performance appraisal systems, and compensation decisions. Questionnaires completed by employees and supervisors, employee interviews, and direct observation are a few of the techniques used to collect data on job content. The information is organized and presented in a formal written document known as a job description, which specifies the nature of the job, the various tasks involved, and the qualifications required.

Once job descriptions are developed, jobs may be compared with one another so that the relative value of each may be determined. This is known as job evaluation, or sometimes as job classification. It is based on the assumption that an employee's rank within an organization should be determined by the nature of the job or position held—a concept known as rank in job. The most common approach to job evaluation is one in which jobs are assessed based on the presence or absence of a set of specific job characteristics or factors. Typical job factors used, include knowledge, skills, contacts with other employees, responsibility for materials and equipment, working conditions, and physical effort required. The factors are weighted and various levels of each are identified. Points are then assigned to each level of each factor, and jobs are then assigned points on the basis of the presence in the jobs of the specified levels of each factor. Once points are assigned to a job on the basis of each factor, a summary score is then developed by adding the factor scores. The summary scores or point totals are then assumed to represent the relative value of jobs analyzed.

After job evaluation is completed and a determination has been made regarding the relative value of the jobs in an organization, the stage is set for the development of a compensation or pay structure. An underlying purpose here is to ensure that employees are treated fairly with respect to compensation. This objective will require that employees with different jobs be paid differently and employees with similar jobs be paid similarly. Jobs that have been determined to have lower relative values will be paid less than those that have higher values. In this manner, the compensation structure, which is usually based on the identification of specific pay grades, will provide equity internally among employees within an organization. It is also important, however, that an organization achieve what is known as external equity, which is the idea that employees be paid fairly relative to what employees in other organizations are paid. Obviously, this objective requires that the organization collect and analyze salary information from the labor market within which it operates.

Clearly, job analysis, job evaluation, and compensation are based on complex and time-consuming processes. Significant skill is necessary to properly complete these functions. Again, smaller local governments may not always have the resources in terms of skilled staff members who can perform these functions. There is an industry of private consultants available, however, to assist in the provision of these services, and smaller jurisdictions may turn to them for such assistance. It is true, though, that consultant's services are often very expensive; as a result, human resources functions associated with job analysis and evaluation

are sometimes not adequately performed in small governments. Large municipalities and urban county governments are usually in a much stronger position to take advantage of professional human resources services, either from well-known consulting firms or from professionals employed in their own human resources management staffs.

LOCAL GOVERNMENT FRAMEWORK

The legal framework within which local government human resources management occurs is also extremely important. As is the case in any public sector setting in the United States, constraints imposed on government by the Constitution significantly influence the nature of human resources practices in local government. In particular, the constitutional guarantee of equal protection of the laws found in the Fourteenth Amendment will restrict the government's ability to draw distinctions among employees and may be particularly important in limiting the nature of affirmative action programs used within selection and promotion systems. Constitutional requirements of due process also significantly shape human resources policies, especially in terms of adverse actions taken against employees. Due process requires, for example, notice and a hearing prior to the dismissal of most (although not all) local government employees. Employees' First Amendment rights to freedom of speech and association must also be protected.

State law also provides further limitations on the operation of human resources management practices in local government settings. States may mandate specific kinds of procedures or practices, and of course, these laws can vary considerably from one state to the next. A good example is provided by laws governing the operation of public employee unions in local governments. Some states permit collective bargaining on a range of issues, including salary and benefits, whereas other states prohibit bargaining in any form with unions and their representatives.

CONCLUSION

In sum, human resources management in local government is a dynamic field. Although the core functions associated with the effective management of human resources are the same in all jurisdictions, the manner in which those functions are carried out varies considerably, depending on the size of local governments and the availability of resources. Variation across the states in the legal framework as specified in state law will also determine the nature of local government human resources practices.

BIBLIOGRAPHY

1. Ban, C.; Riccucci, N.M. *Public Personnel Management: Current Concerns, Future Challenges*, 3rd Ed.; Longman: New York, 2002.
2. Berman, E.M.; Bowman, J.S.; West, J.P.; Van Wart, M. *Human Resource Management in Public Service: Paradoxes, Processes, and Problems*; Sage Publications: Thousand Oaks, CA, 2001.
3. Buford, J.A.; Lindner, J.R. *Human Resource Management in Local Government*; Southwestern Publishers: Cincinnati, 2002.
4. Daley, D.M. *Strategic Human Resources Management: People and Performance Management in the Public Sector*; Prentice Hall: Upper Saddle River, NJ, 2002.
5. Gatewood, R.D.; Field, H.S. *Human Resource Selection*, 4th Ed.; The Dryden Press: Fort Worth, TX, 1998.
6. Hays, S.W.; Kearney, R.C. *Public Personnel Administration: Problems and Prospects*, 4th Ed.; Prentice Hall: Englewood Cliffs, NJ, 2002.
7. Shafritz, J.M.; Rosenbloom, D.H.; Riccucci, N.M.; Naff, K.C.; Hyde, A.C. *Personnel Management in Government: Politics and Process*, 5th Ed.; Marcel Dekker: New York, 2001.

Human Rights, Ethical Analysis of and in Public Policy

Cyrus Ernesto Zirakzadeh

Department of Political Science, University of Connecticut, Storrs, Connecticut, U.S.A.

INTRODUCTION

Today, politicians and political activists of different partisan bents and ideological leanings cite “human rights” as a warrant for extraordinary action. Protesters sometimes justify civil disobedience on the grounds that they are protecting or promoting human rights. Citing the need to punish violators of human rights, superpowers sometimes intervene in other states’ affairs and, on occasion, inflict economic boycotts, war, and other forms of suffering on foreign populations.

Of course, protesters, nation states, and other political actors sometimes are mistaken, insincere, and dishonest when they declare that are acting to protect or promote human rights. Still, human rights are cited often enough to merit an investigation into its meaning. What do we mean when we sincerely talk about human rights? What are the implications of human rights for public policy?

DUAL CHARACTER OF HUMAN RIGHTS

“Human rights” is a compound concept with philosophical and political connotations.

The concept is partly rooted in a Western tradition of moral philosophy—sometimes called the natural-rights tradition—that stretches back at least as far as the ancient Greeks. This tradition derives moral prescriptive judgments from a presumption that underneath the flux of daily life, a timeless natural order exists independently of humans’ preferences and wishes. One element of the timeless order is a fixed human nature.

According to Aristotle, systematic observation of the world reveals that humans are a distinctive species. Among all animals, humans are uniquely able to make reasoned judgments about competing values and about right and wrong behavior. Human-rights philosophers since Aristotle have contended that this choice-making characteristic of human beings is self-evidently valuable and should be protected by all governments. Moreover, its protection ought to outweigh competing government duties, except under truly extraordinary circumstances, such as military invasion. In the language of contemporary human-rights philosophers,

human rights normally should “trump” all competing political interests, values, and goals.

The notion of human rights has another connotation that arises from often blood-drenched histories of political struggles against tyranny. For many political observers and scholars, “human rights” refers to those quasi-constitutional arrangements that have been established to minimize the danger of arbitrary and violent action by rulers against the ruled. Here, the concept is closely connected to specific civil rights and legal processes, such as rule of law and the principle of habeas corpus. A government’s “human-rights record” refers to the government’s palpable treatment of citizens and to the citizens’ practical rights to dissent and contest.

The choice-making and antityrannical connotations of human rights are equally true. Neither meaning is false, and one is not superior or more important than the other. The two meanings are, in a sense, different sides of the same conceptual coin. The philosophic meaning celebrates opportunities for individuals to be creative. The legalist tradition celebrates constraints on the power of the state.

HUMAN RIGHTS AS THE OPPORTUNITY TO MAKE CHOICES

Many natural-rights and human-rights philosophers, such as Jack Donnelly, H. L. A. Hart, and Ronald Dworkin, used evocative phrases, such as treating people as “ends in themselves” or as “subjects, and not objects,” to convey their public-policy orientations. Such pithy mottos assume that all humans are supremely valuable because of their capacities to establish goals, to dream of alternative conditions, and to deliberate about and prioritize aims. Stated differently, the policy implications of human rights arise from the assumptions that humans are meaning-creating creatures who are capable of choice, and that this creative, self-directing capacity should not be curtailed except for extraordinary reasons. Therefore (argue human-rights philosophers), public policy should be guided foremost by the principle that each person should be as free as possible to choose how to dispose of time, energy, and other possessions. Any government

that infringes on its citizens' opportunities to be autonomous (or self-directing) must demonstrate that such actions are needed to protect citizens from a clear, immediate danger of significant proportions. Otherwise, the government's infringement is illegitimate, and citizens have the right to defy, rebel, and (if necessary) overthrow political authorities.

Some human-rights advocates, such as Milton Friedman and Ayn Rand, make almost no further recommendations about legitimate government action. In effect, they say that a government should resemble the proverbial night watchman, whose sole task is to insure that enough peace exists within the community for individual residents to pursue their private dreams. On strictly human-rights grounds, a government should not enforce or implement a particular theory of economic justice on citizens, because this may limit individuals' opportunities to reason morally for themselves. Perceiving the laissez-faire implications of human rights for public policy, Karl Marx criticized the concept's celebration of greed, competitiveness, and self-concern.

Some political philosophers and theorists, however, have deduced more ambitious lists of government responsibilities from the working assumption that all humans are endowed with a capacity to choose. These authors argue that the capacity to choose requires that the government directly or indirectly promote specific social and economic conditions.

One condition that is frequently cited in the human-rights literature is economic security. Some writers believe that without protection from hunger, homelessness, and joblessness, a person cannot have the peace of mind needed to reflect carefully and systematically on values, options, and priorities. Donnelly, for one, argued that governments in advanced capitalist societies must assure all citizens a minimum standard of living (including minimum levels of education, health care, and housing), although Donnelly denied that there need be either a strict equality of condition among all citizens or state ownership of the means of production. In Donnelly's opinion, the exercise of human choice presupposes, whenever possible, an extensive but never totalitarian welfare-capitalist order (akin to the social democratic practices of northern Europe).

Some human-rights philosophers contend that because of contemporary sexual stereotyping, the development of the moral reasoning capacity of human beings requires that the state regulate gender relations and extinguish popular prejudices that may deter some citizens from aggressively defining and pursuing their own goals. Martha Nussbaum, for instance, contended on human-rights grounds that a government should protect its gay and lesbian citizens from housing discrimination, from exclusion from the military, and from daily verbal harassment—hardly a laissez-faire

agenda. Arati Rao similarly argued on human-rights grounds that it is the government's responsibility to protect women from the internalization of feelings of worthlessness and inferiority, and that this may require governmental monitoring of and intervention in citizens' conventionally understood "private affairs"—for example, aggressive enforcement of laws against marital rape.

Donnelly's, Nussbaum's, and Rao's policy recommendations illustrate how the philosophic assumptions of human rights need not lead to the set of laissez-faire policy prescription envisaged by Friedman, Rand, and Marx. The number and kinds of public-policy recommendations that human-rights advocates draw depend in large part on their implicit theories of psychological growth. Broadly speaking, the larger the number of social preconditions that one considers necessary before citizens are emotionally ready to exercise their independent judgments to choose, the larger will be the number of the duties that one wants a government to undertake in the name of promoting human rights.

HUMAN RIGHTS AS THE MINIMIZATION OF TYRANNY

Unlike the philosophic approach to human rights that deduces political prescriptions from a mental vision of all humans as choice makers, Edmund Burke, in the late eighteenth century, championed a historical approach to appreciating rights. He argued that rights should not be viewed mathematically, as logical deductions from timeless principles. Instead, rights should be viewed as the outgrowth of collective experience in overcoming concrete dangers, tyranny in particular. Rights are those arrangements that have been demonstrated to protect groups and individuals from significant political harm.

According to Burke and the later followers of his historical approach, human rights are a logically messy collection of pragmatic rules of thumb that have been tested in actual political struggles and that have been shown to be valuable in constructing ramparts against authoritarianism and illegal rule. Perhaps the first collection of guidelines was the Magna Charta, which in 1215 prohibited the English King from capriciously convicting and punishing subjects. Since 1215, documents outlining proper restrictions on the state have become more commonplace across the globe and also have gradually grown longer. Consider the 1689 English Bill of Rights that specified limits to the King's authority to suspend laws, keep a standing army, and tax subjects. The U.S. Constitution's Bill of Rights (adopted in 1791) similarly enumerated several bulwarks against tyranny, such as citizens' right to bear arms.

Many historically oriented students of human rights believe that the 1948 United Nations' Universal Declaration of Human Rights marked a turning point in the evolution of these quasi-legal documents. Unlike previous documents, the Universal Declaration broadened the concept of tyranny to include social conditions that deprive human beings of different forms of happiness and respect. Later twentieth-century documents, such as the 1993 Vienna Declaration and Programme of Action, continued the practice of seeing social, and not just governmental, threats to personal freedom.

The Universal Declaration of Human Rights did not only explicitly prohibit cruel and unusual punishment by governments and their use of torture. It also proclaimed that dignified work, motherhood, free elementary education, the opportunity to participate in trade unions, and periodic holidays with pay are "human rights" that, if not politically protected and promoted, may compel people to engage in "rebellion against tyranny and oppression." Article 23 proclaimed that "everyone has the right to work" and a right to "free choice of employment," to "just and favorable conditions of work," to "just and favourable remuneration insuring for himself and his family an existence worthy of human dignity." Article 24 added that "everyone has the right to rest and leisure, including reasonable limitation of working hours and periodic holidays with pay." According to Article 25, "Everyone has the right to a standard of living adequate for the health and well-being of himself and of his family, including food, clothing, housing and medical care and necessary social services."

Some critics of the Universal Declaration, the Vienna Declaration, and other similarly worded twentieth-century documents say that these proclamations are little more than laundry lists composed by dewed-eyed reformers, and that the lists are for the most part disconnected from the concrete problems of tyranny that made earlier human-rights doctrines so important. After all, is it obvious that the absence of periodic holidays with pay is a form of tyranny and oppression that must be eradicated immediately? What about the absence of free elementary education?

Another common criticism is that these twentieth-century lists are disconnected from practical sensitivity to the constraints and possibilities posed by historical conditions. How can everyone be assured of "dignified work," in any non-trivial sense of the phrase, given the inevitability of monotonous and servile service-sector jobs even in the wealthiest of nations? And when we look outside the industrialized West, do these proclamations amount to much more than pie-in-the-sky thinking? Can decent housing and modern health services be provided for all citizens in most African and Asian countries, to take but one example?

Brian Barry is among the scholars who are wary of the usefulness of recent lists of human rights. He has openly wondered whether the conventional notion of human rights has drifted so far from the original anti-tyrannical concerns of the Magna Charta that the notion now is little more than shorthand for a particular group's image of desirable social relations—a topic on which opinions differ and about which there are no authoritative "right answers." Barry, moreover, found most of the specific rights listed in the Universal Declaration of Human Rights to be obscure and to be too vaguely stated to be useful even in debates. Barry mused that the cryptic statement in the Universal Declaration about everyone having a right to an "adequate" standard of living might mean no more than that the state is obligated to keep citizens at least barely alive. Any serious declaration of human rights ought to explain what grand phrases mean, insisted Barry. Otherwise, the declaration can have no practical, prescriptive force.

Maurice Cranston, going a step beyond Barry's critique, has argued that even if we give the current lists of human rights clear, substantive meaning, we soon will discover that the concept has become wildly utopian and therefore meaningless. Can we truly expect (Cranston asked) that impoverished nations in Africa and Asia will in the foreseeable future provide free elementary education to everyone? Where will money for teachers' salaries and professional training come from? What about school buildings, books, and electricity? Similarly, "special care and assistance" to mothers and "periodic holidays with pay" will seem utterly impossible goals as soon as we try to give these entitlements some concrete meaning.

Cranston may have a point. Can universal paid holidays, employment security for everyone, and "just and favourable remuneration" (to mention just a few provisions found in the Universal Declaration) be guaranteed to all residents in any major country, including those with the most prosperous, industrialized, and rapidly growing national economies of our day?

But, Cranston could also be wrong. One might, after all, believe in a more optimistic theory of economics and argue that sufficient economic resources already exist to meet almost all the social demands that the Universal Declaration of Human Rights enumerates. Donnelly, for example, has insisted that today, economic surpluses are adequate to realize the social rights listed in the Universal Declaration. The failure to secure these rights reflects a lack of political will, not a shortage of economic wherewithal (or so Donnelly contended).

A somewhat different line of criticism of the current political-historical notions to human rights draws on Sir Isaiah Berlin's seminal distinction between positive and negative freedom. A negative freedom (or negative

right) refers to a person's freedom not to be interfered with, to be left alone. As an example, consider the freedom to pray to the god of one's choice without pressure from the state. Negative freedom is, Berlin maintained, valuable, because it allows for idiosyncrasy, originality, and creativity. A positive freedom (or positive right) refers to the opportunity to affect the external world and thus master one's fate. Consider the freedom to petition a government official or to vote. Critics, such as Cranston, sometimes argue that recent declarations of human rights improperly mix negative and positive freedoms. This mixture is confusing, because it combines two different postures toward political rule: one highly critical and oppositional, and one highly supportive and almost supplicating. The notion of human rights pushes one, paradoxically, toward wanting to hem in the state and wanting to enlarge its powers. How can one derive a coherent and consistent view of public policy when one's basic attitude toward government is so bipolar?

Finally, many commentators have noted that the long list of rights that currently comprise human-rights documents fail to address the messy question of prioritizing policy claims. Of the many proclaimed rights in existence, which should the government take more seriously when conflicts in rights inevitably occur? Article 26 of the Universal Declaration of Human Rights, for example, states that all levels of education are to include a specific ethical curriculum that "promotes understanding, tolerance and friendship among all nations, racial or religious groups, and shall further the activities of the United Nations for the maintenance of peace." But, Article 26 in the same document states, "parents have the prior right to choose the kind of education that shall be given to their children." What if parents wish to teach their children to be wary and critical of the beliefs and practices of other nations, racial groups, or religious groups? Consider the following two hypothetical cases. What should happen if Catholic parents in southern California do not want their child to be exposed to the Islamic ideas of nearby Iranian-Americans, or Amish parents in Ohio do not want their child to be exposed in school to the more agnostic and materialistic values of nearby city dwellers? Which of these two rights governing education of children—the right to learn religious tolerance or the right of parents to choose the kind of education that their children will receive—takes precedence?

Similar questions arise when we juxtapose the human right to "own property alone" (Article 17) with the long list of employees' human rights to a livable wage, to job security, to a paid holiday, and to dignified treatment in a healthy workplace (Articles 23–25). Which, if any, of these rights is to be privileged—the rights of business owners to use their property as

they see fit or the rights of employees to particular wages, holidays, and styles of work?

PUNISHMENT OF RADICAL EVIL

One policy area where the notion of human rights has been applied with minimal conceptual criticism or confusion involves what the late legal philosopher Carlos Santiago Nino called the punishment of "radical evil." This phrase refers to large-scale acts of government violence against unarmed people, such as the Nazis' enslavement and extermination of millions of Jews, homosexuals, gypsies, and ethnic minorities; the "Young Turks" massacre of one-third of Turkey's Armenian population in 1914; and the Argentine military's torture and murder of thousands of students, lawyers, journalists, and peasants during the 1960s and 1970s. Regardless if one adheres to the philosophic understanding of human rights or the political understanding, such acts seem to be unquestionable violations of human rights on a large scale. What policy response is appropriate?

Nino, Hannah Arendt, and some other political philosophers and theorists have concluded that reasonable assessments of proper punishment are hard to reach, partly because it is difficult to determine who should be held responsible for atrocities on a such a scale: the soldiers who are expected in times of war to carry out orders without question; the bankers and industrialists who financially backed the tyrannies and therefore might be considered as complicit in the atrocities; or just a handful of top government officials who signed the orders to murder? Moreover, how does one determine the proper punishment for perpetrators of radical evil—what punishment could ever be commiserate with the scale of the inhumanity of the act? Lastly, because acts of radical evil usually are formally legal—that is, when the despicable violence was carried out, the perpetrators scrupulously adhered to the letter of existing laws and seldom technically violated them—how do outsiders who want to punish someone for the crime against human rights avoid charges of arbitrarily imposing one country's moral standards upon others? How can one reconcile the rule of law, which denies the legitimacy of *ex post facto* laws, with conscientious concern for human rights?

A few famous human-rights trials, such as the 1945–1946 Nuremberg trial of Nazi leaders, have been undertaken since World War II. Although Nino and Arendt were outraged by the Nazis' violence (Arendt died before the Argentine trial was held in the 1980s), both were greatly troubled, because at times, the victors seemed to arbitrarily impose their own standards of justice and openly trample on legal principles. Arendt and Nino were further troubled, because an unusually large

number of local onlookers (other than members of the victims' families) appeared disinterested in the course of the trials. Arendt and Nino reported that if anything, many citizens of Germany and Argentina expressed admiration for the perpetrators of radical evil because of their atypically high status and material wealth. Conversely, most local citizens expressed little outrage over the fate of the victims of radical violence, especially when the victims were poor, gypsies, or belonged to some other socially stigmatized group.

The apparent disinterest of local populations during trials of radical evil raises troubling questions about the extent to which the moral concept of human rights is held worldwide and among all social strata. Although some human-rights advocates, such as Donnelly and Dworkin, insist that arguments about human rights carry considerable weight among policy makers and citizens at large, the observations of Arendt and Nino suggest otherwise. Could it be that in the current pecking order of political goals and social values, the concept of human rights is not as highly placed as Donnelly and Dworkin contended?

Whether in the foreseeable future human rights will become more influential in domestic or foreign policy making depends in part on whether political theorists can satisfactorily answer current questions about conceptual vagueness, resolve debates about the social preconditions for human choice, address concerns about the economic practicality of twentieth-century human rights, and solve puzzles about prioritizing the many rights that human-rights documents currently proclaim. Perhaps if the meaning of the concept becomes more refined and its policy implications less ambiguous, the importance of human rights will become more salient to both government officials and everyday citizens.

BIBLIOGRAPHY

1. Arendt, H. *Eichmann in Jerusalem*; Viking Press: New York, 1963.

2. Barry, B. *Political Argument*; The Humanities Press: New York, 1965.
3. Berlin, I. *Two Concepts of Liberty*; Clarendon Press: Oxford, 1958.
4. Burke, E. *Selected Writings and Speeches*; Doubleday: Garden City, 1963.
5. Cranston, M. *What Are Human Rights?* Basic Books: New York, 1964.
6. Donnelly, J. *The Concept of Human Rights*; St. Martin's Press: New York, 1985.
7. Dworkin, R. *Taking Rights Seriously*; Harvard University Press: Cambridge, 1970.
8. Friedman, M. *Capitalism and Freedom*; University of Chicago Press: Chicago, 1962.
9. Hart, H.L.A. Are there any natural rights? *Philos. Rev.* **1955**, 64, 175–191.
10. Marx, K. On the Jewish question. In *The Marx-Engels Reader*, 2nd Ed.; Tucker, R.C., Ed.; W. W. Norton and Company, Inc.: New York, 1978; 26–52.
11. Nino, C.S. *Radical Evil on Trial*; Yale University Press: New Haven, 1996.
12. Nussbaum, M.C. Lesbian and gay rights. In *The Philosophy of Human Rights*; Hayden, P., Ed.; Paragon House: St. Paul, 2001; 574–597.
13. Rand, A. *Atlas Shrugged*; Penguin Putnam Inc.: New York, 1957.
14. Rao, A. Right in the home: feminist theoretical perspectives on international human rights. In *The Philosophy of Human Rights*; Hayden, P., Ed.; Paragon House: St. Paul, 2001; 505–525.
15. The bill of rights of the United States of America. In *The Philosophy of Human Rights*; Hayden, P., Ed.; Paragon House: St. Paul, 2001; 347–349.
16. The English bill of rights. In *The Human Rights Reader*; Laqueur, W., Rubin, B., Eds.; Temple University Press: Philadelphia, 1979; 104–106.
17. The Magna Charta. In *The Human Rights Reader*; Laqueur, W., Rubin, B., Eds.; Temple University Press: Philadelphia, 1979; 102–104.
18. The universal declaration of human rights. In *The Human Rights Reader*; Laqueur, W., Rubin, B., Eds.; Temple University Press: Philadelphia, 1979; 197–201.
19. Vienna declaration and programme of action. In *The Philosophy of Human Rights*; Hayden, P., Ed.; Paragon House: St. Paul, 2001; 641–649.

Human Subjects Research: United States Public Policy

Patricia M. Alt

Department of Health Science, Towson University, Towson, Maryland, U.S.A.

Marcie Weinstein

College of Health Professions, Towson University, Towson, Maryland, U.S.A.

INTRODUCTION

While the benefits of scientific research on humans have been considerable, it has also raised many ethical questions, most notably about the treatment of the people participating in the research. Public policy makers have tended to produce regulations governing the conduct of human subject research in the aftermath of scandal. In the United States today, federal regulations apply only to research using certain types of public funds or in institutions that choose to apply them to all their activities. Some states have begun regulating research within their boundaries. The Health Insurance Portability and Accountability Act of 1996 (HIPAA) regulations have added a new layer of complexity to the use of medical information, and emerging areas such as stem cell research have become politically charged. Meanwhile, Federal agencies are encouraging publicly and privately funded research to spread beyond academic centers, with widely varied oversight; there is pressure for the use of central Institutional Review Boards (IRBs) (including for-profit organizations) instead of local ones; and some faculty organizations are calling for relaxed IRB review for social and behavioral research.

BRIEF HISTORY OF HUMAN SUBJECTS RESEARCH POLICY

One of the most notorious examples of policy making in crisis was the Nuremberg War Crime Trials, conducted at the end of World War II, which resulted in the drafting of the Nuremberg Code as a set of international standards for judging physicians and scientists who had performed biomedical experiments on concentration camp prisoners. The World Medical Association's Declaration of Helsinki, issued in 1964, further elaborated on these principles. The United States only developed strict policies about human subject protection in research when information about exploitation of subjects in this country became public knowledge in the 1960s and 1970s, painfully exemplified

by the famous Tuskegee Syphilis Study and the Milgram experiment.

Signed into law in 1974, the National Research Act (PL 93-348) charged the Department of Health, Education and Welfare (HEW) with codifying its human subject protection policies into federal regulations. It also created the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, whose charge was to identify the basic ethical principles that should underlie the conduct of research involving human subjects, and to develop guidelines to assure that research is conducted in accordance with those principles. The Commission issued the Belmont Report in 1979, identifying three basic ethical principles, which when applied by scientists, subjects, and reviewers, were intended to provide an analytical framework for the resolution of ethical problems. In 1979, HEW [renamed the Department of Health and Human Services (DHHS)] started the revision of the 1974 regulations and gave final approval in 1981 to 45 CFR 46, referring to Title 45, Part 46 of the Code of Federal Regulations, Protection of Human Subjects.^[1]

The principles discussed in the Belmont Report are: *respect for persons*, which requires the acknowledgment of individual autonomy and the protection of individuals with diminished autonomy; *beneficence*, which translates into "do no harm," seeking to maximize potential benefits and minimize potential harms; and *justice*, referring to fairness in distribution of research benefits and burdens, so that no one group unduly bears the costs, while other groups unduly reap the benefits.^[2] These three principles form the basis for six norms of scientific behavior that the Belmont Report mandated for the ethical conduct of research, and they are summarized below:

- *Valid research design*: The researcher must use a valid design that takes into account relevant theory, methods, and prior findings; design is selected to minimize risks.
- *Competence of researcher*: The researcher must have the necessary skills and experience to carry out all procedures; the researcher is responsible for

training all other investigators; and the researcher is responsible for guaranteeing appropriate facilities for safe implementation.

- *Identification of consequences:* The researcher must provide subjects with an assessment of benefits and risks, taking care to establish a favorable balance of benefits to risks; the researcher must respect privacy and ensure confidentiality.
- *Selection of subjects:* The researcher must offer additional protections to vulnerable populations and enroll them only in minimal risk studies; the researcher must use fairness in enrolling subjects so that no groups are unduly burdened and no groups are excluded who may benefit.
- *Voluntary informed consent:* The researcher must obtain consent beforehand that must be obtained freely after the subject is fully informed, has the opportunity to ask questions, and demonstrates comprehension; consent means that explicit agreement is reached, and it can be withdrawn at any point during the study.
- *Compensation for injury:* The researcher is responsible for what happens to subjects and is required by federal law to inform subjects of possible compensation; compensation is not required.^[3]

The 45 CFR 46 regulations require institutions that receive federal funds and conduct research involving human participants to establish Institutional Review Boards (IRBs) for the purpose of protecting the rights and welfare of human research subjects. Specific requirements govern the composition and operation of IRBs; they must be composed of a minimum of five members, and their membership should be sufficiently diverse to guarantee expertise in a broad range of scientific, and also ethical, considerations. Racial and cultural diversity must be considered, as well as community representation.^[2]

Researchers submit their research protocols to their institution's IRB prior to conducting research, and IRBs use the following criteria to evaluate whether the proposals appropriately safeguard the rights and welfare of subjects:

Is the study design consistent with sound scientific principles and ethical standards?

Have all risks to subjects been minimized, including their right to privacy and confidentiality of participation?

Are risks to participants reasonable relative to anticipated benefits and the importance of the potential knowledge generated?

Is subject selection equitable? If subjects are drawn from vulnerable populations, such as children, pregnant women, and prisoners, are additional appropriate safeguards built in? Are questions of undue influence and coercion considered?

(Continued)

Are all necessary elements of voluntary informed consent present? Is the consent form written in language that would be understandable to a layperson? Are subjects apprised of all of their rights as a research subject, such as the right to withdraw without penalty?

Does the study design make adequate provisions to ensure the safety of subjects during their participation, and to modify protocols as appropriate, including discontinuation?

Are adequate provisions in place to safeguard the data at study conclusion to protect subjects' rights to privacy and confidentiality?^[2]

The IRBs are responsible not only for the prospective review of all research involving human participants but also for the ongoing review of each approved research protocol, at least annually. In an initial review, an IRB may approve a study protocol, may approve the protocol contingent upon modifications requested from the Board, may declare a study exempt from IRB approval if specific criteria are met, or may deny approval. An IRB has the responsibility to modify, suspend, or terminate an ongoing study that is not being conducted in accordance with its initial requirements or that poses serious risk to subjects.^[2]

While institutionally based IRBs provide oversight and assurances of human subject protection in their respective facilities, the Office for Human Research Protections (OHRP), which was formerly called the Office for Protection from Research Risks (OPRR), is the body responsible for broadly overseeing the implementation of 45 CFR 46 at the federal level. Housed in the Office of the Secretary, DHHS, the OHRP also provides guidance on ethical issues related to biomedical and behavioral research. Similarly, the U.S. Department of Education has federal regulations governing the protection of human subjects, known as 34 CFR 97, and provides oversight through the Grants Policy and Oversight Staff. In fact, a number of federal agencies have their own policies and regulations for human subject protection, and while they maintain their individual oversight functions and regulations specific to vulnerable populations, there is substantial similarity among them. In 1981, 17 federal departments and agencies adopted a common set of regulations known as the Federal Policy for the Protection of Human Subjects, or "Common Rule." The "Common Rule" is based on the ethical principles and standards outlined in the Belmont Report. The "Common Rule" was reissued in 1991, without change, and was again endorsed by the federal agencies it governs. Since then, it has been only slightly amended, although under constant scrutiny by a series of Federal advisory committees on human research protections.^[4] However, not all federally funded human subjects research is covered by those agencies. Although the "Common

Rule” has provided for essential safeguards, changes to the policies and procedures for human subject protection are often encouraged to increase overall flexibility and ability to adapt to emerging ethical and scientific issues.^[1]

The IRBs have the mandate to educate researchers about human subject protection and will, therefore, work with researchers to design or revise protocols in order to include the necessary safeguards. Some IRBs have begun requiring proof from investigators that they have been trained in the responsible conduct of research (RCR), which includes not only human subjects protection but also plagiarism, animal protection, etc. The Office of the Inspector General, in a series of 1998 reports, found that investigators had inadequate knowledge of their responsibilities toward human subjects.^[5] The National Institutes of Health now require investigators to have had basic human subjects ethics training in order to be considered for grant funding. However, the Department of Health and Human Services’ Office of Research Integrity policy requiring broader RCR training of all investigators was suspended indefinitely during 2001, following a Congressional request that it be reviewed further, and has not been reinstated.^[6] There is no common standard among federal agencies regarding investigator training in human subject research or RCR at this time.

The National Bioethics Advisory Committee (NBAC) issued a draft report called “Ethical and Policy Issues in Research Involving Human Participants” in December, 2001, just as it disbanded. This report advocated the creation of a centralized National Office of Human Research Oversight (NOHRO) to supervise all federal and private studies. The “Common Rule” does not currently apply to private, non-federally funded studies, but the committee pointed out the tremendous growth in privately funded research, particularly pharmaceutical, and argued that it should be held accountable to the same standards.^[7] The proposed centralized office would have also eliminated other problems associated with the “Common Rule,” including the difficulty and confusion in enforcing multiple sets of regulations without a single authority overseeing uniform protections. The “Common Rule” is considered inflexible and unable to adapt to emerging ethical and scientific issues. However, there was also fear that creating a centralized oversight agency would stifle research creativity, as the NBAC report called for the NOHRO to have “final authority for enforcement,” although that “function should be shared with federal departments, institutions, and accrediting bodies.”^[7]

Upon the expiration of the NBAC’s charter in November, 2001, President George Bush created a successor President’s Council on Bioethics, with a similar mission. It is charged with “advising the President on

ethical issues related to advances in biomedical science and technology.” Its reports focus on such issues as human cloning, biotechnology, stem cell research, and ethical caregiving. As areas such as stem cell research have become controversial, with the ban on Federal funding of embryonic stem cell research and the passage of a variety of state laws designed to attract biotechnology researchers, human subjects research has faced the danger of politicization and public distrust.^[4]

In addition to the President’s Council on Bioethics, the National Human Research Protections Advisory Committee was created in 2000, to provide advice and recommendations to the Secretary of HHS, the Assistant Secretary for Health, and the Director of the Office of Human Research Protections on a range of issues connected with protecting human research subjects. When its charter expired, the Secretary’s Advisory Committee on Human Subject Research Protections was created in 2002, and still continues to operate. It has been carefully examining all aspects of the Common Rule, as well as considering requiring accreditation and training for IRBs.^[4]

Another pressure on research with human participants is coming from the implementation of the HIPAA. While compliance with the privacy regulations was not required until April, 2003, debate over their impact on medical research had raged since the law was passed. As an outgrowth of nationwide interest in the privacy of medical information, HIPAA’s main concern is not with human subjects research. However, the regulations sometimes do make research using medical records much more cumbersome. In order to protect individuals’ rights to control access to their medical records, these regulations establish much stricter guidelines for access to personal records than the “Common Rule” required. The Biotechnology Industry Organization and the Association of American Medical Colleges were among the groups arguing that the rules for eliminating identifying information are going to make the use of medical records much less meaningful.^[8] However, it can also be argued that the public’s increased concern for privacy needs to be reflected in the IRB system, and in all aspects of human subjects research. As Annas stated, “Public support of medical research is really a function of public trust. Providing meaningful protection of the privacy of medical records in research is an important goal in its own right and will also increase public trust in the entire medical-research enterprise.”^[9]

Additional human research policy issues have been raised by governmental and private agencies. Frequently heard are concerns that IRBs are not supplying adequate oversight and are often pressured to review and approve protocols quickly. Several well-known cases of ethical violations in medical research

resulting in serious adverse incidents have instigated repeated calls for heightened scrutiny and tighter regulation. A 1996 review by the General Accounting Office, among others, concluded that IRBs face serious shortages of staff and financial resources, undermining their ability to provide sufficient oversight to assure human subject protection.^[10] Other reports discussed the changing climate in medical research, which current regulations are inadequate to address, such as multisite research, conflicts of interest arising from privately sponsored research, new issues related to protection of vulnerable populations, and under-representation of women and children in research studies.^[11] The 2003 NIH "Roadmap" explicitly encourages innovative public-private research partnerships and increasing clinical research outside of academic medical centers, raising new questions about appropriate oversight.^[11] At the same time, non-medical university faculty are increasingly rallying to the concept that IRBs have expanded their oversight too far, especially in social and behavioral research.^[12]

The Children's Health Act of 2000 required that DHHS and FDA regulations for research with children be brought into agreement. It also required further study of the adequacy of IRB protection for children in clinical trials. As a result, the DHHS and the FDA issued new regulations during 2001 about the appropriate use of vulnerable populations in studies. While seeking to protect these populations, they are also trying to ensure their representation in studies, so that the results will be applicable to all groups of the population.^[13]

In addition, policy questions continue to arise about the appropriate balance between gathering information for public health purposes and for research. Should different privacy and informed consent standards apply if a county Health Department is gathering information while investigating an epidemic or potential bioterrorism than if it is "only" gathering the information to produce generalizable knowledge? How do we differentiate between public health practice and research, other than by the original intent of the data gathering?^[14] And, what about the concept of "group risk," which would call for IRBs to review research for its possible negative policy consequences for a protected group (such as Native Americans), as well as for the individual research subjects? "Community consultation" when research will disproportionately affect a particular neighborhood, ethnicity or racial group, is seen as one possible approach to identifying and dealing with these ethical concerns.^[15]

Finally, a perplexing policy question concerning the ethical principle of justice in human subject protection has emerged. Whereas justice was initially intended primarily as protection for vulnerable groups easily exploited in subject recruitment, a differing view

looks at justice in terms of access to clinical trials for vulnerable groups. Since the paradigm began shifting in the late 1980s and early 1990s, there has been a growing belief that research can offer real benefits, and the focus on potential benefits has started to over-ride the focus on potential risks. A classic example occurred in the HIV/AIDS community, which demanded opportunities to participate in studies of experimental drugs and protested limited enrollment in these studies, claiming injustice due to the lack of access to potential research benefits. Some cancer studies have produced similar results. The implications for human subject protection from a policy perspective concern the appropriate balance between protection from risk and access to benefits, the potential loss of distinction between research and treatment, and the adequacy of current informed consent standards.^[1] This is especially true when an increasing amount of research is being conducted in settings such as private doctors' offices.

CONCLUSION

Recognizing that many of our most important scientific discoveries and medical advances have resulted from the participation of human subjects in research, the United States continues to evolve its policies for the protection of those human participants. In the process, policy makers are seeking to balance the benefits of the research, to participants and to the world, against potential risks to each individual participant. Our policies seek to maintain the rights of all citizens to be treated ethically in the pursuit of science. To that end, current regulations, promulgated at the federal and state levels, and implemented at the local and institutional levels, aim to provide for the safety and protection of human subjects in research. Because scientific inquiry is a dynamic process, policy makers are attempting to ensure that regulations keep pace with the research they are charged to oversee. Most recently, this has included an emphasis on such areas as medical privacy rights, new genetic technologies and therapies, increased concern for the rights of children, and more stringent review of IRB operations.

REFERENCES

1. Mastroianni, A.; Kahn, J. Swinging on the pendulum: shifting views of justice in human subjects research. *Hastings Center Rep.* **2001**, *31* (3), 21–28.
2. <http://ohsr.od.nih.gov/guidelines/> (accessed October 2006).
3. Sieber, J.E. *Planning Ethically Responsible Research: A Guide for Students and Internal Review Boards*; SAGE Publications Inc.: Newbury Park, CA, 1992.
4. <http://www.hhs.gov/ohrp/sachrp> (accessed October 2006).

5. Yessian, M. Reflections of an Outsider. In *Institutional Review Board: Management and function*, 2nd Ed.; Amdur, R., Bankert, E., Eds.; Jones & Bartlett: Sudbury, MA, 2006; 9–12.
6. http://ori.hhs.gov/policies/RCR_Policy.shtml (accessed November 2006).
7. Appel, J.M. Research guidelines: changes urged. *J. Law Med. Ethics* **2001**, 29, 103–112.
8. Moreno, J.; Caplan, A.L.; Wolpe, P.R. Updating protections for human subjects involved in research. *J. Am. Med. Assoc.* **1998**, 280, 1951–1958.
9. Annas, G. Medical privacy and medical research—Judging the new federal regulations. *N. Engl. J. Med.* **2002**, 346, 216–220.
10. U.S. General Accounting Office. *Scientific Research: Continued Vigilance Critical to Protecting Human Subjects*; US General Accounting Office: Washington, DC (Publication GAO/HEHS-72), 1996.
11. <http://nihroadmap.nih.gov> (accessed November 2006).
12. Nelson, R. Research involving children. In *Institutional Review Board: Management and Function*, 2nd Ed.; Amdur, R., Bankert, E., Eds.; Jones & Bartlett: Sudbury, MA, 2006; 366–372.
13. <http://www.law.uiuc.edu/conferences/whitepaper> (accessed November 2006).
14. Speers, M. Epidemiology/public health research. In *Institutional Review Board: Management and Function*, 2nd Ed.; Amdur, R., Bankert, E., Eds.; Jones & Bartlett: Sudbury, MA, 2006; 415–417.
15. Dickert, N.W.; Sugarman, J. Community consultation: not the problem—An important part of the solution. *Am. J. Bioethics* **2006**, 6 (3), 26–28.

Humanitarian Intervention

Michelle Maiese

University of Colorado at Boulder, Boulder, Colorado, U.S.A.

INTRODUCTION

Humanitarian intervention is the threat or use of military force primarily for the purpose of protecting the nationals of the target state from gross and systematic human rights violations. These violations result from either the conscious policies of a central regime or the general breakdown of that regime. The use of force may come in a variety of forms, including economic sanctions, arms embargoes, restrictions on income-generating activities, aviation bans, restrictions on diplomatic representation, and suspension of membership or expulsion from international or regional bodies. The discussion here focuses on the use of military force by one or more outside states to relieve grave human suffering.

OVERVIEW

Controversy surrounding humanitarian intervention arises from widely accepted constraints on the use of force among states. According to international norms discussed below and provided they do not threaten the legitimate claims of other states, sovereign states have a right to pursue their own destiny free from interference. They also have primary responsibility for the protection of their people's lives. Some believe that military intervention, even to protect human rights, denies the rightful autonomy of the state and constitutes an unacceptable assault on state sovereignty.

Others point out that passivity in the face of gross and systematic human rights violations is likewise unacceptable. They suggest that a government that engages in substantial violations of human rights forfeits both its domestic and its international legitimacy. If a state is unwilling to halt serious harm to its people, the principle of non-intervention yields to the international responsibility to protect.^[1] To be legitimate under international law, military action to protect human rights should be authorized by the United Nations (UN) Security Council. Cases of "unilateral" intervention, undertaken by states without UN authority, are especially problematic.

There is a growing notion that the protection of human rights and dignity should be one of the fundamental objectives of modern international institutions. However, because full-scale military operations may

harm or kill many civilians, those who engage in intervention for human protection purposes should be careful to avoid doing more harm than good. The use of military force should produce substantial humanitarian benefits and be exercised according to principles of right intention, last resort, proportional means, and reasonable prospects.

POLICY OBJECTIVES

It seems clear that the issues faced in the 21st century differ from those the world faced in 1945 when the UN was founded. Cold War politics have imposed a brutal check on the political and economic development of many states. The amount of internal war and civil conflict has grown drastically over the last several decades and many countries have weak economic structures and government institutions.

Given the increase in intra-state conflict, the international community is faced with a dilemma: stand by while there is massacre, ethnic cleansing, and genocide; or take sides in civil conflicts, which may only further contribute to fragmentation of the state system.^[1] Furthermore, humanitarian intervention entails military operations that may harm or kill many civilians. To destroy human life for the benign purpose of protecting human rights seems unacceptably paradoxical. Humanitarian intervention may also seem to go against many of our moral intuitions about the conduct of foreign policy. After all, states are typically concerned only with their own security and well-being; and international relations are commonly thought to be first and foremost relations among states and not individuals.

On the other hand, it seems clear that justice and rights protection no longer stop at the border of the nation-state.^[2] When faced with gross human rights abuses, the international community should not accept that state sovereignty automatically trumps human rights claims. Instead, there is a need to balance the principle of state sovereignty against protection of human rights. Policy surrounding humanitarian intervention needs to meet at least four broad objectives:^[1]

1. Establishing clearer rules, procedures, and criteria for determining whether, when, and how to intervene.

2. Establishing the legitimacy of military intervention in those cases where all other approaches have failed.
3. Ensuring that military intervention is carried out for the purposes proposed, is effective, and is undertaken with proper concern for human costs and institutional damages that may result.
4. Helping to eliminate the causes of conflict while enhancing the prospects for a sustainable peace.

STATE SOVEREIGNTY AND HUMAN RIGHTS: A DILEMMA

The principles of sovereignty and non-intervention suggest that states can legitimately use armed force only in cases of self-defense. The international community, including the UN, is prohibited from intervening in matters that are considered wholly within the domestic jurisdiction of states. Article 2^[2] of the UN Charter prohibits the use of force against a state's territorial integrity or political independence and suggests that states have exclusive jurisdiction within their own territory. If the duty not to intervene is violated, the victim state has the right to defend its territorial integrity and political independence. This effectively rules out interference, whether direct or indirect, with a state's political, economic, or cultural elements. It also seems to rule out a legal right to use force against another state even to promote human rights.

For many states, sovereignty is their only line of defense against major powers. The notion of sovereign equality guarantees that despite differences in economic, social, or political structure, less powerful states have the same legitimacy, rights, and duties as the developed Western states, as well as equal membership in the international community.^[2] In general, sovereignty has meant immunity from external challenge and broad discretion with respect to methods of governance. Insofar as it serves as recognition of states' equal worth and dignity, their identity, and their right to self-determination, respect for sovereignty contributes to international stability.

Because the state is the ultimate repository of individual rights, a resolution should be sought within the state itself when states abuse human rights. If the people do not like what their government is doing, it is for them to rise up and reshape that government. When outside states impose their conception of rights on other countries, this may be viewed as a form of moral imperialism or intellectual arrogance. Moreover, even if human rights conditions are awful, an enduring remedy may require fundamental reordering of the whole society and its values; and an outside force is unlikely to be able to accomplish this. There is also a worry that widespread acceptance of humanitarian

intervention will lead aggressive governments to use it as an excuse for their expansionist policies.^[3]

However, others argue that humanitarian intervention does not represent cultural imperialism, but rather reflects the shared rights of humanity. Sovereignty has never meant the unquestionable right of government leaders and officials to do anything they please within their recognized space. Rather, a state's right of self-preservation is subordinate to individuals' rights of self-preservation.^[4]

The fact that a state has been established according to traditional legal norms and has a fixed territory is not enough for legitimacy. Rather, the community grants limited powers to a central governmental authority to serve the interests of that community. Because states gain their legitimacy from the consent of persons, it is unclear that states that abuse human rights and obstruct social justice should be regarded as legitimate. In fact, governments that behave badly or misuse their authority undermine the one thing that justifies their political power and thereby become delinquent within the international community.^[3] They should not be allowed to shield themselves behind international law and the principle of state sovereignty.

While Article 2^[2] of the Charter establishes the norm of non-intervention, other provisions of the Charter list the protection and advancement of human rights as a primary purpose of the UN system. For example, Article 1^[2] of the Charter commits the UN to promoting and encouraging respect for human rights and fundamental freedoms. The Universal Declaration of Human Rights (1948), the four Geneva Conventions, and the establishment of the International Criminal Court likewise set forth standards for state conduct. These agreements and mechanisms have changed expectations about what counts as acceptable state conduct and established a universal jurisdiction concerning matters of human rights. This emerging discourse implicitly recognizes a new kind of accountability and aims to protect oppressed minorities and excluded social groups against the domination of the powerful.^[2]

Insofar as nations have agreed to promote universal respect for human rights and fundamental freedoms, these rights are a matter of international obligation and subject to the supervision of the international community. But does this warrant forceful intervention to secure that such obligations are met? Chapter VII of the Charter suggests that the UN Security Council may pursue economic, diplomatic, or military measures to maintain or restore international peace and security. Because widespread human rights violations create large numbers of refugees and often escalate domestic civil strife that can spill over into other countries, some maintain that such violations constitute a threat to international peace and security. As

the international community becomes more and more connected, states that violate human rights come to pose a risk to people everywhere. According to this view, the UN has a right and a duty to respond to these situations. Passivity in the face of gross and systematic human rights violations is unacceptable.

However, there is no express linkage in the Charter between the maintenance of peace and security and the protection of human rights. While the Charter does include many human rights provisions, it does not outline concrete powers and functions of the UN in promoting and protecting these rights. It seems that even if an intervention is for the purpose of protecting core human rights, it violates Article 2^[2] by suppressing the political will of the target state. On the other hand, an expansive view of security sees it as extending beyond national or territorial security to include the security of individuals against threats to their life, safety, and well-being.^[1] Over the last several decades, the UN Security Council has empowered itself to consider humanitarian emergencies a threat to international peace and security.

International human rights covenants and UN practice have begun to conceptualize sovereignty in terms of responsibility. In signing the Charter, a state takes on a dual responsibility: externally, to respect the sovereignty of other states; and internally, to respect the dignity and basic rights of people within the state. When states are unwilling or unable to do so, this responsibility should be borne by the broader international community. The responsibility to intervene stems from a general duty to assist victims of grave injustice and help them regain their autonomy and dignity.

CRITERIA FOR LEGITIMATE HUMANITARIAN INTERVENTION

Humanitarian intervention is a matter of delivering aid to ordinary people whose lives are at risk because their states are unwilling or unable to protect them. The international community has yet to develop consistent, credible, and enforceable standards to guide state and intergovernmental practice.

Growing practice among states and regional organizations, human rights provisions, and Security Council precedent suggest that the “responsibility to protect” is an emerging guiding principle for humanitarian intervention policy. The responsibility to protect implies an evaluation of the issues from the point of view of those needing support. It acknowledges that primary responsibility rests with the state concerned, but suggests that if the state is unwilling or unable to fulfill this responsibility, it becomes the responsibility of the international community to act in its place. This raises questions about legitimacy, authority, and operational

effectiveness. The use of military force for human protection purposes should be guided by the principles listed in [Table 1](#).

Most turmoil within states does not require coercive intervention by external powers. The use of military force for human protection purposes should occur only in exceptional and extraordinary cases: situations of genocide, ethnic cleansing, forced expulsion or “disappearance,” enslavement, and widespread and systematic acts of terror or rape. Situations of state collapse, mass starvation or civil war, or overwhelming natural or environmental catastrophes might also be included. Actions that fall short of outright killing, on the other hand, may not warrant military intervention. This includes cases of systematic racial discrimination or political repression.

In those cases where state action (or inaction) “shocks the conscience of mankind,” military intervention should be guided by the principle of right intention. This means that the central purpose of the intervention should be the ending of gross human rights abuses and the restoration of peace. Because some level of self-interest may be required to justify expense of resources and troops, intervening states are likely to have mixed motives. In addition, there is often a gap between stated purposes and actual motivation. However, such cases should not be used as a rationale for abandoning the practice of humanitarian intervention altogether.^[3]

There are various ways to ensure right intention. First, intervention should involve multilateral rather than unilateral state action. Second, those in need of help should welcome and support the intervention. Third, other countries in the region should support the actions taken.^[1] These conditions reflect the general belief that interventions to safeguard commercial property or ideologically friendly regimes are impermissible.

Military intervention to protect human rights should be used only as a last resort, after all avenues for peaceful resolution of the crisis have been tried. Before sending troops into harm’s way, endangering local civilians, and spending valuable resources, leaders have an obligation to attempt other peaceful alternatives. This includes negotiations, economic pressure, peacekeeping, pressure from regional organizations, international legal bodies, and other diplomatic efforts. In addition, all reasonable local remedies should be tried before calling on the international community for assistance. The people of the region should have a voice and accountability in negotiating solutions to humanitarian crises.^[2]

In addition, the means used in humanitarian intervention should be proportional and limited to what is required to protect human rights. Foreign troops should not occupy the invaded territory longer than necessary, demand favors from a newly established

Table 1 Criteria for just humanitarian intervention

Just cause	Large-scale loss of life, with genocidal intent or not; or ethnic cleansing, whether carried out by forced expulsion, acts of terror, or rape. These atrocities may be actual or apprehended, and the product of either deliberate state action or inaction.	This excludes violations that fall short of outright killing: systematic discrimination, imprisonment, or political repression.
Right intention	Primary purpose must be to halt or avert human suffering. This is best assured by multilateral operations that are supported both by regional opinion and the victims.	This excludes the overthrow of regimes, the alteration of borders, or the advancement of ideologies or economic goals.
Last resort	Every non-military option for the prevention or peaceful resolution of the crisis must be explored; and gross human rights violations must be imminent or actual.	This excludes use of military force before negotiations, economic pressure, and other diplomatic avenues have been tried; or there are reasonable grounds to suspect such efforts would not succeed.
Proportional means	The scale, duration, and intensity of the intervention should be the minimum required to protect human rights. The effect on the country's political system should be limited.	This excludes military operations that result in massive civilian deaths, lengthy military occupations, and the use of destructive modern weaponry.
Reasonable prospects	There must be a reasonable chance of success in halting human suffering; consequences of intervention should not be worse than the consequences of inaction.	This excludes cases where protection cannot be achieved, or where intervention is likely to escalate or trigger a larger conflict.
Right authority	The UN Security Council is the most appropriate body to authorize intervention. Its authorization should be sought prior to any military action being taken. If the Security Council fails to act, the matter of intervention may be considered by the General Assembly or regional organizations.	Unilateral action by states is prohibited by international law; however, emerging state practice suggests that states may take unauthorized action in certain cases of extreme necessity.

Source: From Ref.^[1].

government, or seek to dominate the targeted country. Intervention should preserve the territorial integrity of the target state and interfere with the ruling structure only as necessary so as to promote an enduring peace. Whenever possible, military action should be directed not at the entire nation, but rather at those responsible for wrongdoing. Intervening states that contribute more to the deaths of innocents than the aid of the oppressed violate this principle of proportionality.

For states to engage in military action, there should be reasonable prospects of halting or averting human rights abuses. Intervening states need to calculate the damage that may be done to the target society and consider the possibility that the intervention will trigger a larger conflict. If humanitarian intervention leads to further escalation and war, it is likely to do more harm than good.

Note that humanitarian intervention differs from conventional warfare. Rather than seeking to defeat an enemy, it aims to protect and rescue populations. This calls for a new type of military operation, one that is preceded by preventive actions and followed by post-conflict rebuilding. The operational procedures used

should promise quick aid to those under attack and reflect the principle of proportionality. Forces should uphold military law governing the use of force and refrain from using destructive modern weaponry. Because the use of military power may lead to further human rights abuses, quick success through surprise or the use of overwhelming force may not be feasible. There is need for “incrementalism” with respect to the intensity of operations and “gradualism” with respect to phases of the operation. The use of military force should be carefully tailored to the objective of protecting human rights. Finally, intervening military forces should collaborate with civilian authorities and humanitarian organizations to bring assistance to populations at risk. This is the best way to ensure an effective intervention.

THE IMPORTANCE OF UN AUTHORIZATION

Many believe that to be legal and permissible, military action against a sovereign state requires UN authorization. Under the UN Charter, the Security Council has

primary responsibility for peace and security matters. Chapter VII of the Charter suggests that when the Security Council considers non-military measures to be inadequate, it may take military action to maintain or restore international peace and security. The Security Council has expanded the traditional conception of what constitutes a “threat” to peace so that human rights violations are increasingly on the agenda.

Reliance on the Security Council for authorization establishes a legal framework by which humanitarian intervention can and should be conducted. Collective intervention endorsed by the UN is widely regarded as legitimate because it is duly authorized by a representative international body. The Security Council is the principal institution for mediating power relationships and consolidating the authority of the international community. Advantages of relying on the UN include its universal membership, political legitimacy, administrative impartiality, convening and mobilizing power, and expertise.

Many believe that the UN is ineffective and point to inherent limitations in its capacity to conduct interventions. First, some question the democratic legitimacy of the Security Council and suggest that it does not have broad enough representation. The founders of the UN were not interested in creating an ideal system of collective security in which all states would participate equally.^[4] Instead, special responsibilities and privileges were given to the most powerful states. As a result, any of the Permanent Members can use their veto power to override the majority on matters of grave humanitarian concern. This sometimes leads to inertia on the part of the UN.

The UN does not have its own military force or the resources needed to carry out humanitarian intervention. Nor does it have facilities for the training of officers and troops in rules of engagement, command and control, or civic reconstruction. When the UN decides to deploy forces, they must be hastily recruited, and forces from different countries rarely have the opportunity to train with each other. It is likely that the UN will continue to authorize states to deploy forces under their own command in support of Security Council resolutions.^[4]

One alternative to Security Council authorization is to seek support for military action from the General Assembly in an Emergency Special Session under the established “Uniting for Peace” procedures. While the General Assembly lacks the power to authorize military action, it could help to legitimize an intervention that has taken place and put pressure on the Security Council to rethink its position. A General Assembly resolution would engage the views of a wide range of states to determine whether intervention was warranted. However, a General Assembly recommendation for military action still requires approval from the Security Council to be fully legally legitimate.

THE QUESTION OF UNILATERAL INTERVENTION

In cases where the Security Council fails to act, states may be inclined to conduct an intervention on their own initiative. Unilateral intervention has both prospects and problems.

If unified Security Council action is not possible, human rights imperatives will be pitted against the Charter’s rules governing the use of force. It seems we must choose between damage to the international order if the Security Council is bypassed or the slaughtering of human beings while the Security Council does nothing.^[1] If the Security Council cannot uphold its responsibility to protect, unilateral action by states becomes more likely.

However, without the discipline and restraint of the UN, such interventions may not be conducted for the right reasons or in accordance with the proper precautionary principles. Some believe that discounting UN restrictions on the use of unilateral force guarantees the legitimization of barbarism in the name of human rights. There is a danger that the stronger powers will assert a new “duty” or “right” to assert “global leadership” on behalf of the world’s vulnerable. Unauthorized humanitarian intervention poses a threat to the international constraints on the behavior of states and does serious damage to the UN’s reputation and authority. Many believe that intervening states do not really intend to change international law, but rather to create new, exceptional rights for themselves. Providing an exception for a powerful state to violate the rules that continue to apply to all other actors severely undermines respect for international law governing the use of force.^[5] The international legal system has a stake both in prohibiting unilateral intervention and preventing crimes against humanity.

Some point to the evolution of customary law and emerging state consensus as a legal justification for unauthorized humanitarian intervention in rare and exceptional cases.^[6] Such instances of “extreme necessity” arise when there are extremely severe violations occurring, force is necessary to stop immediate harm to civilian populations, and humanitarian intervention is indeed the lesser wrong.^[7] There should be widespread consensus that a situation of extreme necessity has in fact arisen, and states should proceed cautiously. Although the UN Security Council has not authorized action, the intervening states should maintain a close relationship with the council.

The intervention should be motivated primarily by humanitarian purposes, involve a coalition of states, and be welcomed by the population bearing the brunt of the atrocities. The consequences of the intervention should be more in keeping with the intent of the law than what would have ensued if no action had been

taken.^[7] This emerging consensus can help states to find a balance between what is lawful and what is right.

CONCLUSION

Military intervention for humanitarian purposes is controversial both when it happens and when it fails to happen. While there may be significant values of human dignity and justice advanced by intervention, there are also competing values favoring orderly and non-forcible resolution of such situations.^[4] It seems clear that the tension between UN authority to address human rights issues and the demand that it not interfere in the internal affairs of states has been resolved to some extent by an expansive view of peace and security. This expansion raises questions for the future. Will this expansive view of “threats to peace” unleash a new era of just wars? Is the international community seeing a return to warfare “for the right reasons” that it may ultimately regret? Will the Security Council act merely as a tool of imperialism and colonialism, dominated by the most powerful nations?

Military intervention should always remain an option of last resort. However, when humanitarian intervention is required, it should reinforce the collective responsibility of the international community. National leaders should collaborate to mobilize both domestic and international support for intervention, and the UN should respond with timely authorization.

Finally, it is important to note that policymaking has focused almost exclusively on international reaction human rights violations. Little attention has been paid to the prevention of catastrophe or the need for peace building after abuse has occurred. To genuinely protect human rights, the international community

should focus its attention both on conflict prevention and post-conflict rebuilding.^[1]

REFERENCES

1. Evans, G.; Sahnoun, M. The responsibility to protect. In *Report of the International Commission on Intervention and State Sovereignty*; International Development Research Centre: Ontario, Canada, December 2001; pp. 5–6, 15, 36, 55, 69.
2. Chandler, D. *From Kosovo to Kabul: Human Rights and International Intervention*; Pluto Press: Sterling, VA, 2002; pp. 12, 120, 127, 179.
3. Garrett, S. *Doing Good and Doing Well: An Examination of Humanitarian Intervention*; Praeger Publishers: Westport, CT, 1999; pp. 26–28, 32, 118.
4. Murphy, S. *Humanitarian Intervention: The United Nations in an Evolving World Order*; University of Pennsylvania Press: Philadelphia, 1996; pp. 26, 28, 298, 305–308.
5. Byers, M.; Chesterman, S. Changing the rules about the rules? Unilateral humanitarian intervention and the future of international law. In *Humanitarian Intervention: Ethical, Legal, and Political Dilemmas*; Holzgrefe, J.L., O’Keohane, R., Eds.; Cambridge University Press: Cambridge, UK, 2003; 177–203.
6. Stromseth, J. Rethinking humanitarian intervention. In *Humanitarian Intervention: Ethical, Legal, and Political Dilemmas*; Holzgrefe, J.L., O’Keohane, R., Eds.; Cambridge University Press: Cambridge, UK, 2003; 232–272.
7. Franck, T. Interpretation and change in the law of humanitarian intervention. In *Humanitarian Intervention: Ethical, Legal, and Political Dilemmas*; Holzgrefe, J.L., O’Keohane, R., Eds.; Cambridge University Press: Cambridge, UK, 2003; 204–231.

Impacts of Bureaucratic Reform on State Government Administration

Brendan F. Burke

*Department of Political Science, Bridgewater State College,
Bridgewater, Massachusetts, U.S.A.*

INTRODUCTION

State-level reform has taken several active paths during the past decade. This entry describes three distinct but potentially compatible directions for state reform—the enhancement of internal management subsystems, the movement toward a results orientation in agencies, and the expansion of customer and stakeholder accessibility to the functions of government via on-line technological advancements. For each reform line, this entry offers context and definition, specification of systems or process changes, and assessment of analytic endeavors to contrast state progress. State rankings on these reforms are compared, and a discussion of the measurement of potential outcomes of reform is initiated.

STATE LEVEL EFFORT AT ADMINISTRATIVE REFORM

Recent progress in state administration builds on a solid foundation of institutional advancement during the second half of the twentieth century. Since 1950, the states have acted to streamline organizations, update procedures, enhance the strength of institutions, and reduce the prevalence of independent authorities and commissions.^[1] At the end of the century, the subnational level proved to be the most active in “reinvention”—the reorientation of policy and administrative systems toward concern for customer satisfaction and enhanced results.^[2] The federal system facilitates a healthy competition across like units of government, wherein the 50 states view at least a comparable subset as competitors. The states strive not only to enhance conditions for their own citizens, but also to demonstrate their comparative advantage with regard to other states, or to leave their inferior status behind and climb up the ranks.

Table 1 displays a comparison in rankings for the most and least successful states on three different reform trends. Each trend, as well as a current measurement process and details of implementation, is discussed in turn. The table shows that certain states are

successful under multiple definitions of reform, and that others rank poorly on multiple measures. However, certain states focus their reform efforts more narrowly, with leadership concentrating only on certain techniques (and for those at the lowest rankings, neglecting only certain narrower bands of reform option).

THE REFORM OF INTERNAL MANAGEMENT SUBSYSTEMS

Light^[3] contends that reforms are like the tides—varying in force over time and potentially cyclical in nature. The historic focus of Light’s “scientific management” tide is the enhanced efficiency of administrative systems as a precursor to the best direct service delivery processes. Early advocates, such as Luther Gulick, recognized the public administration specialist as the appropriate actor to ensure public management effectiveness and efficiency. In 1997, the Pew Charitable Foundation granted funds to Syracuse University’s Maxwell School of Citizenship and Public Affairs to pursue the Government Performance Project (GPP), an effort to measure managerial performance at the federal, state, and local levels. The study results are displayed prominently on the project’s web page^[4] and through special issues of *Governing* magazine.

The management philosophy underlying the GPP is as follows: Government performance is a function of the cohesiveness and coherence of policy implementation efforts. But an important precursor of implementation is the strength of staff or administrative processes. Public management is seen as a kind of “black box” involved in the translation of government resources into varied direct results. Successful management includes an effective oversight of personnel, financial, planning, and information technology “subsystem” capabilities. In addition to these separate internal service functions, the integration of all four comes through the contemporary emphasis on “managing for results”.^[5] The unifying strand of the GPP model is a focus on enhancing the capacity of executive leadership and other top management to

Table 1 Comparison of state rankings on alternative reform orientations

Rank	Internal management subsystems ^a	Results or “reinvention” ^b	Accessibility via e-government
1	Michigan	Michigan	Indiana
2	<i>Utah</i>	<i>Florida</i>	Michigan
3	<i>Washington</i>	Georgia	Texas
4	Delaware	Oklahoma	Tennessee
5	<i>Iowa</i>	<i>South Carolina</i>	<i>Washington</i>
6	Kentucky	<i>Ohio</i>	California
7	Maryland	Oregon	New York
8	Missouri	<i>Virginia</i>	Pennsylvania
9	Pennsylvania	<i>Iowa</i>	<i>Florida</i>
10	<i>South Carolina</i>	<i>Utah</i>	<i>Ohio</i>
11	<i>Virginia</i>	<i>North Carolina</i>	<i>North Carolina</i>
40	Alaska	Alaska	Alaska
41	<i>Arkansas</i>	<i>Mississippi</i>	<i>West Virginia</i>
42	<i>Connecticut</i>	New Hampshire	<i>Mississippi</i>
43	<i>Hawaii</i>	<i>Hawaii</i>	Vermont
44	<i>Nevada</i>	<i>Arkansas</i>	Rhode Island
45	New Hampshire	Rhode Island	<i>Oklahoma</i>
46	Oklahoma	New Jersey	Arizona
47	Rhode Island	Tennessee	New Mexico
48	<i>West Virginia</i>	<i>Nevada</i>	New Hampshire
49	<i>Wyoming</i>	<i>Connecticut</i>	Alabama
50	Alabama	Alabama	<i>Wyoming</i>

Source: Internal Management Subsystems: Government Performance Project, <http://maxwell.syr.edu/gpp/> (accessed October 2003); Results or “Reinvention,” Deil S. Wright, ASAP Data Files, Odum Institute for Research in the Social Sciences, University of North Carolina at Chapel Hill, Accessibility Via E-Government, <http://www.insidepolitics.org/egovt01us.html> (accessed November 2003).

Bold print indicates states in all top or bottom groupings for all three reform orientations; *italics* indicates states in two of three top or bottom groupings.

^aUnder Internal Management Subsystems, states ranked 1–3, 4–11, and 40–49 receive tie scores.

^bThe results or “reinvention” ranking does not include missing data from California, and thus ranks from 1 to 49.

make well-informed choices and to disseminate them promptly throughout the organization.

The GPP assigned letter grades on an “A”–“F” scale for achievement on each of the five management subsystems, and a summary grade. The grades represent an accretion of effort mostly during the 1980s and 1990s—not an evaluation for a given time unit. But because the state survey was reported in 1999 and 2001, there is some evidence of rising, steady, and falling performance at the turn of the century. The following are examples of “top-grade” activity within each of the categories:

1. Human resource management: Leading states consolidate job categories, where possible, in the civil service system; use prompt hiring and vacancy review processes; provide pay-for-performance systems; and practice long-term human resource planning. This can include an assessment of future human resource

needs—both the potential types of jobs and responses to hard-to-serve markets within the state. Human resource management remains a relatively weak internal function overall for the states—only three received an “A” or “A–” grade in 2001 (South Carolina, Washington, and Wisconsin), and the average state grade was “B–”.

2. Financial management: Successful states use broad budget authority so that appropriations can be flexibly matched with agency-spending needs; emphasize cost accounting and enterprise fund use; provide effective policies for investment and maintenance of fund balances; perform long-term revenue and expenditure forecasting; and stipulate performance-based contracting rules. Ten states received an “A” or “A–” grade in 2001, providing evidence that this is a relatively strong internal focus (average grade: B).

3. Information technology management: Powerful chief information officers are present to coordinate technology needs; to bridge statewide and agency plans; to optimize the use of the Internet for state web sites; and to ensure quick turnaround in technology replacement plans so that systems are current in a rapidly changing field. This is an improving subsystem, but still with some room for enhancement. The 1999 average grade was “C+” and the 2001 grade was “B–”.
4. Capital management: The capital plan is not “one year of reality and four years of fantasy;”^[6] the future needs are well prioritized and will be implemented to schedule; and the management of maintenance and operations funding is an adjunct to the construction of projects. The states were relatively successful at this management function in the initial GPP study, but there has been some decline since then. Nine states received “A” or A–” grades in 1999, but the number fell to five in 2001. The average grade remained a “B–” across the two studies.
5. Managing for results: The leading states pursue strategic planning between legislatures and agencies; effective performance measures are used in most agencies; performance is compared across agencies and across comparable states in benchmarking programs; and state auditors ensure the effectiveness and accuracy of performance information. Only five states received “A” or “A–” grades on this integrating reform in 2001, with an average grade of “C+”.

THE RESULTS ORIENTATION

Although internal subsystems reform has brought greater efficiency and an enhancement of managerial capacity, it neglects an important determining point for the quality of governmental effort—the transaction between the government and the service user or recipient. “Street-level” bureaucratic effort may benefit from better managerial guidance, information, and control of resources, but it serves also to benefit from enhanced flexibility in decision making and allocation of resources. In Light’s vocabulary, this is the “liberation management” tide, but the most common moniker for this movement is “reinventing government.” The Winter Commission Report, *Hard Truths/Tough Choices*, and Vice President Gore’s National Performance Review (NPR) documented the problems and the possibilities for reinvention at the subnational and national levels, respectively.^[7,8] During the past decade, the most salient theme of reinvention has been the movement from government dialog away from processes, and toward service results.

The American State Administrators Project (ASAP) survey includes questions that track the prevalence of specific “reinvention features” underlying the results-oriented reform wave. Based at the University of North Carolina at Chapel Hill, with funding from the Earhart Foundation, the survey has been performed twice each decade since the 1960s, including inquiries about reinvention in 1994 and 1998. State agency heads respond to the mail survey, making ASAP a valid and expert-based source of knowledge about various trends across American state governments. For a full description of the ASAP survey and its methodology, see Ref.^[9]

Eleven features of reinvention were included in the survey, with the request that the respondent indicate a degree of implementation from “no consideration” to “full implementation” of the different reinvention possibilities (Table 2). A primary difference between the ASAP and GPP studies involves the unit of analysis. The GPP developed state grades on the internal management subsystems, whereas the ASAP results tracked agency activity and breadth of implementation across all states. Here, the findings describe the prevalence of the reform technique, without a detailed analysis of a state-specific nature. The state ranking in Table 1 was developed from the summation of results gained from the agency heads within each state, deleting California from the ranking because of a low response rate.

The ASAP inquiry gained information about the scope of reinvention across three of the same four categories as the NPR: reducing regulation by cutting red tape (reinvention features 8, 10, and 11); empowering employees (3, 6, and 7), and refocusing on the customer (1, 2, 4, and 5). The NPR focus on “cutting back to basics” is probably better served at the state level through the requirements for balanced budgets. Contracting out may be downplayed in this summary of reinvention techniques, but elsewhere in the ASAP survey, it is reported that 72% of state agencies contract out some services to private companies, non-profit agencies, or other governmental units.

ACCESSIBILITY REFORMS VIA E-GOVERNMENT

A third reform trend has quickly gained stature in the twenty-first century. Light contends that the tides of reform can overlap one another, and this is certainly the case with state-level e-government initiatives embracing the Internet. (For an assessment of governmental techniques underlying E-government, see Ref.^[10]. For a more theory-based assessment, see Ref.^[11].) States, through the use of the World Wide Web, have enhanced customer access to governmental services

Table 2 Progress on reinvention in the American States—Agency implementation levels for specific reinvention features, 1998

Reinvention feature	Percent of agencies	
	Partially implemented	Fully implemented
1. Strategic planning to produce clear mission statements	37	50
2. Training programs to improve customer service	54	29
3. Quality improvement programs to empower employees	56	23
4. Benchmarks for measuring outcomes	49	24
5. Systems for measuring customer satisfaction	45	19
6. Decentralization of decision making	42	18
7. Reduction in hierarchical levels	23	21
8. Greater discretion in procurement	33	13
9. Simplification of human resource needs	27	7
10. Privatization of major programs	17	7
11. Greater discretion to carry over funds	14	8

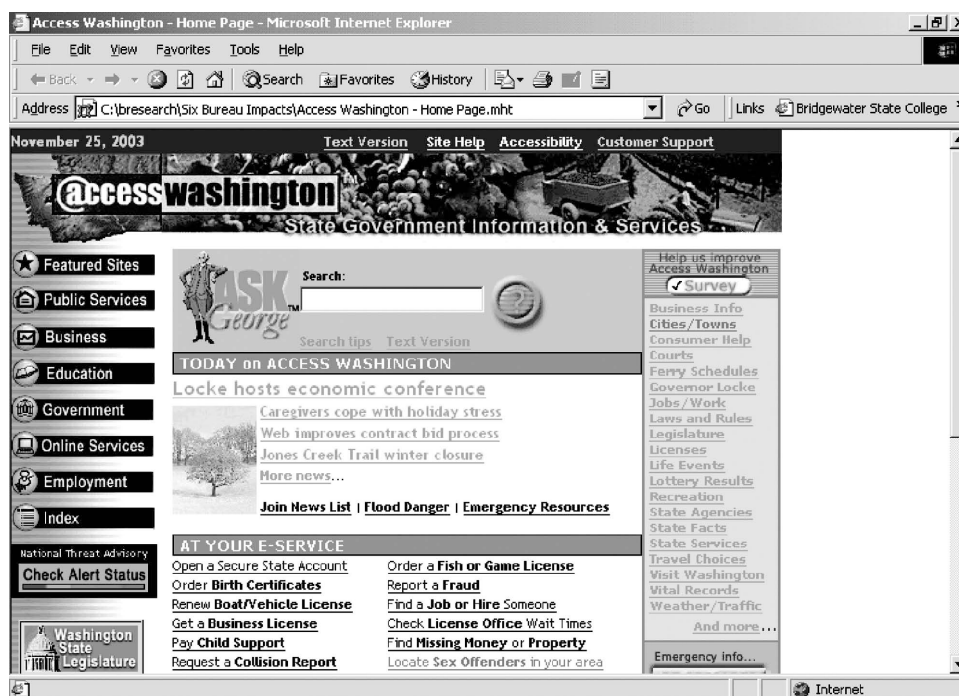
Source: Deil S. Wright, ASAP Data Files, Odum Institute for Research in the Social Sciences, University of North Carolina at Chapel Hill.

Respondents are agency heads asked: "From time to time, state agencies undertake to change the way they do things. Please indicate the extent to which your agency has implemented each of the following changes in the 1990s." Responses range from "no changes considered" to "fully implemented" ($n = 1075$).

and have enhanced the citizens' ability to understand and affect governmental processes. As a reform tide, accessibility enhancement unites the results orientation with a portion of what Light calls the "watchful eye." Earlier legislation under this tide included the Administrative Procedures Act of 1946 and various "sunshine" laws around processes such as public hearings. Citizen access, to be heard by a large number of

governmental actors via e-mail and to read primary documents and databases, has been made much easier via the e-government focus. The Internet is rapidly becoming the facilitator of informed participation.

The accessibility reforms did not arise because of a specific problem; instead, some potential problems could now be treated because of new technology. The Washington State home page on the Internet (Fig. 1)

**Fig. 1** The Washington State home page.

immediately shows the enhanced value for the government's customers and interested citizens. Citizens have immediate and easy access to a number of services, including the renewal of vehicle licenses and payment of child support. The accesswashington site also shows how it may serve "watchful eye" interests, with information links to the state legislature and contract bid processes. Within the state legislature link, one can track legislative actions, their history, and their sponsors.

Darrell West, director of the Taubman Center for Public Policy at Brown University, studied a variety of characteristics of state effort underlying their Internet presence through surveys in 2000 and 2001. At that time, all states except Wyoming provided a variety of services on-line, including the filing of taxes, purchase of vehicle registration and hunting licenses, and the purchase of governmental documents. Forty-eight of 50 states enabled secure access for on-line transactions. All states had some web sites with disability access features, but in some cases, the percentage of accessible sites was low. Between the surveys in 2000 and 2001, West found that most states improved their Internet presence in these service areas and their on-line administration. West ranked the states on their implementation of these citizen services based on the following criteria: user accessibility, interactivity, and ease of use in both years; the 2001 rankings are reported in Table 1.^[12]

MEASURING THE IMPACTS OF REFORM

Many of the intermediate impacts of bureaucratic reform have been specified to this point—public administration techniques bring about improvement and change to administrative systems. But what of the ultimate impacts of reform? More specifically, in what ways have state government reforms improved the status of the citizens? The transaction-enhancing nature of the results-oriented and accessibility-based reforms is too situational or contextual to aggregate into summary measures. One common goal of both the results-oriented and accessibility-oriented reforms is the enhancement of citizen trust in the government. There is situational evidence of the contrasting possibilities regarding this goal and the reform effort. Michigan and Florida are two states with high rankings on both their "reinvention" and e-government initiatives. In Florida, overall trust in the government has risen along with reform implementation; in Michigan, reform implementation does not correlate with higher trust in the government.

The impact of the internal management subsystem reforms is more readily analyzed because of their nature as an intervening variable on the path toward better governmental results. If the GPP grades can be

considered a correlate of the "scientific management" movement, then they should illuminate efficiency gains within government processes. One way to test this is to link the GPP grades and a measure of financial strength—the state bond rating from Standard & Poor's (S&P). Bond ratings constitute a relatively stable factor that affects the cost of governmental debt issues. Small reductions in the interest rate for debt issuance, which correspond to the reduced investor risk of a highly rated governmental entity, can save that government many millions of dollars on a large bond issue. The GPP and S&P measures, when converted to rank order, have a relatively strong 0.50 correlation (0.01 statistical significance). The tie between GPP measures and bond ratings indicates an efficiency impact directly beneficial to citizens within the internal subsystem reforms.

CONCLUSION

During the 1990s and into the twenty-first century, the states have been active implementers of a variety of reform types, designed to enhance efficiency and responsiveness, and to reduce the psychic distance between government agents and varied stakeholders. The intermediate administrative impacts are clear with the types of improvement introduced by state leadership and bureaucracies. The ultimate or substantive impact is clear on internal subsystem reforms, but is harder to track on transaction-based reform tools and techniques. The measurement of service outcomes is a growing advancing area. With effective measures for the results of the myriad of governmental activities, there may prove to be an acceptable method to aggregate these to a higher level, thus giving us a comparative basis across the waves of administrative reform.

REFERENCES

1. Bowman, A.O'M.; Kearney, R.C. *State and Local Government*, 5th Ed.; Houghton Mifflin Company: Boston, Massachusetts, 2002.
2. Osborne, D.; Gaebler, T. *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*; Addison-Wesley: Reading, Massachusetts, 1992.
3. Light, P.C. *The Tides of Reform: Making Government Work, 1945–1995*; Yale University Press: New Haven, Connecticut, 1997.
4. <http://www.maxwell.syr.edu/gpp/> (accessed October 2003).
5. Ingraham, P.W.; Kneeder Donahue, A. Dissecting the black box revisited: characterizing government management and capacity. In *Governance and Performance: New Perspectives*; Heinrich, C.J., Lynn, L.E., Jr., Eds.;

- Georgetown University Press: Washington, DC, 2000; 292–318.
6. Barrett, K.; Greene, R. Grading the states. *Governing* February 1999, 12 (5), 50 pp.
 7. National commission on the state and local public service. *Hard Truths/Tough Choices: An Agenda for State and Local Reform*; Nelson, A., Ed.; Rockefeller Institute of Government: Albany, New York, 1993.
 8. Gore, A. *Creating a Government that Works Better and Costs Less: The Report of the National Performance Review*; Penguin Books: New York, NY, 1993.
 9. Brudney, J.L.; Wright, D.S. Revisiting administrative reform in the American States: the status of reinventing government during the 1990s. *Public Admin. Rev.* 2002, 62 (3), 353–361.
 10. Barrett, K.; Greene, R. *Powering Up: How Public Managers Can Take Control of Information Technology*; CQ Press: Washington, DC, 2001.
 11. Fountain, J.E. *Building the Virtual State: Information Technology and Institutional Change*; Brookings Institution Press: Washington, DC, 2001.
 12. <http://www.insidepolitics.org/egovt01us.html> (accessed November 2003).

Impasse Resolution Procedures

Gary E. Roberts

*School of Global Leadership and Entrepreneurship, Regent University,
Virginia Beach, Virginia, U.S.A.*

INTRODUCTION

The collective bargaining process frequently results in an impasse, an inability to reach an agreement on one or more issues under dispute between labor and management. The objective of this article is to describe the various impasse resolution strategies typically employed to resolve labor disputes as well as their respective strengths and weaknesses (Table 1).

The exact timing of a formal impasse declaration varies according to federal, state, or local labor laws. In some situations, there are well-defined dates to reach an unaided settlement, after which an impasse must be declared. In other settings, the law is silent or ambiguous as to when an impasse must be declared and is left to the discretion of labor and management based upon their bargaining history and prevailing practice (how labor and management customarily resolve disputes). An impasse can also be declared at any time in the bargaining process when the differences between the parties are of such a magnitude that future negotiations are deemed hopeless.

Once an impasse is reached, there are a variety of impasse resolution procedures to aid labor and management in overcoming their differences. They include mediation, fact finding, arbitration, and their variants, such as med-arb. The federal government and 38 states have legislation supporting some type of impasse resolution procedure, with mediation being the most common (36), followed by fact finding (33) and arbitration (30).^[1]

FUNDAMENTAL LABOR RELATIONS PRINCIPLES

A fundamental tenet of the collective bargaining process is that the most equitable settlement is crafted by the parties. As such, impasse resolution techniques are designed to encourage both parties to reach a voluntary agreement before the imposition of a settlement by a strike or lockout or through the actions of a third party, such as the government. Strikes are permitted in the private sector, but only 11 states permit public employees a limited right to strike.^[1] Strikes impose serious costs on labor and management while

disrupting production and service delivery to the public. Conversely, research demonstrates that dispute resolution costs are higher in jurisdictions that fail to provide finality through the right to strike.^[2]

A common misconception regarding labor disputes is that when parties reach an impasse, they disagree over many issues. In most cases, labor and management dissent over a few key bargaining areas. The specific disagreement areas that trigger an impasse vary dramatically from one situation to another, but certain subject areas prompt more disputes than others. The most frequently contested area entails wages and other compensation practices such as longevity pay, merit pay, and employee benefits. Common non-monetary issues include disputes over job security, such as outsourcing, promotion practices, and working conditions, among others.

MEDIATION

When a stalemate is reached, the typical first dispute resolution technique employed is mediation. A mediator is a neutral third party who attempts to restart direct talks between labor and management to facilitate a voluntary settlement. Full-time mediators employed by federal and state governments receive specialized training and are certified by professional organizations. The expertise of ad hoc mediators varies significantly, and their use has uncertain, and sometimes negative, outcomes. There is no formula for success as a mediator given the multitude of variables that influence any single impasse situation. General personality traits associated with success include problem-solving ability, tact, diplomacy, patience empathy, and clear communication skills.^[3]

Mediators employ three strategies based upon their analysis of the situation.^[4] A reflexive approach entails meeting the parties separately to develop an understanding of the disputed issue(s), the parties' positions on each issue, how each side views the other, and identification of the actors with the power to approve or reject a settlement. The non-directive approach attempts to create a climate more conducive to reaching a settlement and to help labor and management become more effective at crafting an agreement.

Table 1 Advantages and disadvantages of impasse resolution procedures

Procedure	Advantages	Disadvantages
Mediation	Flexibility Encourages creativity	Scarcity of skilled mediators Absence of finality
Fact finding	Increases pressure for settlement Formal record maintained	Absence of finality High economic cost
Conventional arbitration	Finality, no strike Decisions are usually binding	High administrative cost Chilling and narcotic effects High settlement costs “Split-the-difference” tendency Delegation of budgeting authority
Final offer arbitration	Finality, no strike Decisions are usually binding Reduces rate of unreasonable offers	High administrative cost Chilling and narcotic effects Chance of unreasonable settlements Illegal delegation of budgeting authority

This frequently entails an educational process in which the mediator outlines the causes and consequences of an extended bargaining impasse. In addition, a mediator may suggest potential solutions or trial balloons that provide novel means for solving the impasse. For example, if the major issue at dispute is a wage increase, the mediator may suggest greater employer pension contributions in lieu of a large salary increase. The final strategy is a directive approach, in which the mediator takes an active role to craft a settlement. This is accomplished by directly proposing a settlement and pressuring the parties to agree, in essence, taking control of the process. Skilled mediators use one or all of these approaches depending on the competency and motivation levels of labor and management negotiators and external factors such as economic or political conditions.

The advantages of mediation include its flexibility, economy, lack of a written record thereby encouraging creativity, ability to educate inexperienced parties, and adherence to the voluntary settlement principle.^[1] The desirable characteristics of mediation can be a serious liability in other situations. For example, the absence of a terminal process makes the outcome of mediation uncertain, because the mediator cannot force a settlement. For mediation to be effective, the mediator needs to be skilled and experienced, qualifications that are in scarce supply.

FACT FINDING

The next level of impasse resolution is fact finding, which is more common in the public than the private sectors. Fact finding is a formal, quasi-judicial process in which a single fact finder or a panel of fact finders hears 1 or 2 days of formal testimony from labor and management on the disputed issues. A transcript of the hearings is sometimes produced. States such as Florida mandate that a collective bargaining impasse

situation culminate in fact finding, while in other states fact finding is voluntary upon the agreement of both parties. There is significant variability in how the fact-finding session is conducted. Some fact finders engage in mediation prior to issuing a decision by informally meeting with the parties and encouraging a voluntary settlement, while others take a more legalistic approach and simply hear the case. There is significant variation in the outcomes of fact finding in terms of the recommendations and the degree of public scrutiny. In some instances, fact finders offer no formal recommendations, while in others, detailed recommendations are issued. Clearly, fact finding absent a formal recommendation differs little from mediation. In terms of public review, in some instances, fact-finding decisions are public record and are widely publicized in the media, while in other settings the recommendations are not released to the public. There is disagreement on the value of public disclosure, with proponents arguing the force of public opinion encourages a settlement, while other observers claim that most labor disputes command little public interest except in a strike situation.^[1]

What criteria does a fact finder use to make decisions? Typical criteria include wage comparability in the local labor market area, the organization's fiscal health and ability to pay, changes in the cost of living, local labor market conditions, and prevailing practice (what most other employers provide or mandate) in terms of working conditions and work rules.

What are the main advantages of fact finding? Fact finding provides an opportunity for parties to reflect on their positions and reduce tensions and anger. It also provides additional negotiating time and can provide “cover” for one or both parties to make concessions that would result in considerable resistance or controversy if a settlement was voluntarily reached. Skilled fact finders can mediate the dispute, resulting in a voluntary settlement during negotiations. The negative aspects of fact finding revolve around its absence of finality given its voluntary nature.

Empirical studies of fact-finding effectiveness have found mixed results with relatively low settlement rates after the issuance of a reward.^[1] Evidence from a large-scale study in New York found that fact finding becomes more successful with a formalized adjudicative (courtroom) structure in contrast to the less successful accommodative (informal) approach.^[5]

ARBITRATION

Binding arbitration is the final common impasse resolution technique. Our focus is on interest arbitration, which entails issuing a decision on the terms of a collective bargaining contract vs. grievance arbitration, which addresses disputes over the interpretation of an existing contract. However, grievance arbitration disputes can enter interest arbitration decisions if they become part of an effort to change or reinterpret the contract. For example, if there are numerous grievances on forced overtime provisions, one or both parties may reach an impasse over contractual language changes. The main benefit of binding arbitration is the closure provided and its avoidance of a strike or lockout. In states with binding arbitration, the incidences of strikes are reduced when compared to statutory prohibitions.^[1]

Arbitration is mandated by legislation for certain categories of public safety employees, such as police and firefighters in states such as Pennsylvania, while in other circumstances, the parties may voluntarily agree to the use of binding arbitration as the terminal conflict resolution procedure. Successful arbitrators are typically highly educated professionals, such as lawyers and college professors. There is a professional organization of labor arbitrators, The American Arbitration Association (AAA), that certifies the qualifications of arbitrators and provides lists of its members to those interested in employing the service of an arbitrator. The most experienced arbitrators conduct the majority of arbitration hearings, given labor and management's interest in having a skilled and seasoned neutral with a proven track record hear their case. Arbitration awards are published by commercial publishers, thereby providing an opportunity for labor and management to research an arbitrator's case history. Unions and management strive to select arbitrators that issue fair, balanced, and well-reasoned decisions.

Larger employers and unions may have a panel of arbitrators hear their case, while smaller employers typically select a single representative. In some industries, union and management contract with an arbitrator to be the exclusive third-party neutral to hear interest and grievance arbitration cases. The advantage of having the same arbitrator is the increased knowledge and understanding gained by the arbitrator

regarding the industry, bargaining history, and economic conditions surrounding the collective bargaining environment. There are a variety of approaches in arbitrator selection. A common strategy is to request a list of names from the AAA, typically 10 arbitrators. Management and labor alternately strike off names until the required number of arbitrators is left.

Most arbitration hearings are conducted in a quasi-judicial manner with sworn testimony, transcripts, and cross-examination permitted. Frequently, union and management are represented by attorneys. Each side presents detailed evidence supporting its position along with voluminous documentation from research studies, prior arbitration decisions, and records of the collective bargaining process that demonstrate the intent and meaning of the respective parties. After hearing the details of the case, the arbitrator issues a written and binding decision. As is the case in most collective bargaining situations, the issues at dispute are limited, narrowing the scope of the arbitrator's decision. In most cases, the only appeal of an arbitration ruling is to the courts, but the standard of review is high in many states and requires a violation of the authorizing statute or a procedural or due process error. Courts are reluctant to overturn an arbitration decision absent a compelling and unremediable error that damages one or both parties and clearly violates case or statutory law.

CONVENTIONAL ARBITRATION

There are several variants to the arbitration process. Conventional arbitration permits the arbitrator wide discretion in crafting a settlement. This places a heavier burden on the arbitrator's analytical and legal-reasoning skills. For example, if the union is asking for a 10% wage increase and the management is offering 5%, the arbitrator operating under a conventional arbitration process could split the difference (a 7.5% raise). If the information submitted warrants a different wage increase from a simple compromise, the arbitrator is free to settle on an alternate figure usually supported by a detailed economic or equity justification. Arbitrators vary in their inclination to mediate, as some encourage the parties to settle voluntarily even after the hearing is conducted and before a ruling is issued, while others make no attempt to reconcile the bargaining positions. There is a variant of arbitration that formally incorporates mediation into the arbitration process and is called med-arb. The parties choose a mediator, and if a settlement cannot be reached, the mediator assumes the arbitrator's role.^[1]

As with fact finding, arbitrators typically utilize well-defined decision rules in making awards. These include wage and benefit comparability, fiscal impact

on the employer including the ability to pay, and industry or sector prevailing practice regarding working conditions and work rules. Union and management typically present detailed wage and benefit surveys to support their contention that their proposals are reasonable. Detailed budgetary information on revenue and expenditures supports fiscal capacity and impact arguments. The financial impact of the settlement on staffing and service levels is another common information source.

FINAL OFFER ARBITRATION

Conventional binding arbitration has been criticized for permitting arbitrators to simply split the difference and not consider the broader economic or policy considerations. If the union believes that the arbitrator will compromise, they are more likely to present a large wage increase and hope that the arbitrator will provide half of a much larger “loaf.” This compromise may impose serious financial burdens on the employing organization.

As such, states such as New Jersey have instituted a variant of conventional arbitration, final offer arbitration.^[6] In final offer arbitration, the arbitrator must select the complete package of labor and management with no alternations. The theory behind the final offer approach is that the threat of having the other parties’ package selected in total reduces the incentive to present unreasonable proposals and have the arbitrator select a compromise position.^[7,8] A variant of the final offer by package approach is a final offer arbitration on an issue-by-issue basis. Each side presents their final offers on each disputed issue, and the arbitrator must select the union’s or management’s final proposal without alternation. Hence, the final offer by issue approach provides more flexibility for the arbitrator and reduces the probability of forcing an arbitrator to choose between final offers that contain unreasonable demands from both sides. For example, in the typical final offer approach, management could demand an unreasonable benefits reduction, while the union package could contain a huge wage increase. This enhanced flexibility, however, reduces the deterrent impact of having the arbitrator reject your complete proposal and select the other side’s.

ARBITRATION CRITIQUES

Arbitration is attacked on a number of fronts. The first critique is the cost involved. Direct costs include the fees charged by arbitrators, the research and proposal development costs, the costs of presenting the proposal, and the time spent analyzing and interpreting

the final settlement. Complex arbitration cases can run into the tens of thousands of dollars.

Another arbitration critique is that the process illegally delegates budgetary authority to an unaccountable third party, an arbitrator. Adherence to arbitration awards can require tax increases, service cutbacks, or layoffs that are not approved by the electorate or the legislative branch. On the basis of this illegal delegation of authority argument, several states forbid binding arbitration. Related to this is the argument that arbitrators place excessive weight on wage and benefit comparability without a complete analysis of the municipality’s fiscal health and ability to finance a settlement. Several New Jersey arbitration awards were overturned in court when arbitrators placed excessive weight on comparability, thereby imposing costly settlements on municipalities that were fiscally stressed.^[6] The aggregate research on the wage impact of arbitration indicate only modest gains in mean salary levels.^[1]

A second critique is that arbitration, on average, favors unions, as arbitrators are more concerned about alienating unions that represent multiple bargaining units across a given geographic area than single employers who are less likely to share information. However, most arbitrators possess a clear incentive to issue balanced awards in order to maintain their credibility. The available empirical evidence indicates relative balance between labor and management in terms of wins and losses in arbitration.^[1]

A third critique is that the availability of arbitration radically reduces the incentive of labor or management to settle voluntarily. There are two variants of this dysfunctional effect. The first is the “chilling effect” in which arbitration attenuates the motivation of the parties to settle earlier in the collective bargaining process given the absence of a strike or lockout threat.^[9] The second drawback is the “narcotic effect,” which occurs when one or both parties perceive that they have a greater chance of success in achieving their collective bargaining goals through arbitration than through traditional collective bargaining.^[10] Hence, the rate of arbitration use increases dramatically over time. Empirical evidence on the chilling and narcotic effects are mixed, however, with strong regional differences.^[1]

Other less common impasse resolution strategies include labor–management committees composed of neutral representatives from labor, management, and the community. The committee members are appointed by the governor based upon nominations by the respective stakeholders. Another strategy is a public referendum, or letting the public vote on issues at an impasse. This strategy manifests several disadvantages, including the difficulty in communicating complex bargaining issues, lack of voter interest and education on the issues, the expense and delays

associated with the voting process, and an increase in political posturing.^[1]

CONCLUSION

The primary goal of impasse resolution is to reduce the number of strikes and lockouts in order to preserve labor peace, safeguard service delivery, and protect the public interest. Voluntary settlement of collective bargaining disputes is the desired outcome, but given the many factors that mitigate against voluntary settlements, impasse resolution procedures are a necessary safeguard. The central objective is to assist the parties in clarifying their basic interests so a reasonable compromise can be crafted. Mediation and fact finding are nonbinding impasse resolution procedures, while arbitration and its variants impose finality to the collective bargaining process.

REFERENCES

1. Kearney, R.C. *Labor Relations in the Public Sector*; Marcel Dekker Inc.: New York, 1992.
2. Hebdon, R.; Stern, R. Do public sector strike bans really prevent conflict? *Ind. Relat.* **2003**, *42* (3), 493–512.
3. Simkin, W.E. *Mediation and the Dynamics of Collective Bargaining*; BNA Books: Washington, DC, 1971.
4. Kressel, K. Labor mediation: an exploratory survey. In *Public Sector Labor Relations: An Analysis and Readings*; Feuille, P., Kochan, T.A., Eds.; Thomas Horton and Daughters: Glen Ridge, NJ, 1977; 252–272.
5. Hebdon, R. Fact finding effectiveness: evidence from New York State. *Ind. Relat.* **2001**, *40* (1), 73–82.
6. Roberts, G.E.; McGill, J. New Jersey interest arbitration reform act: a third year assessment. *Rev. Public Pers. Admin.* **2000**, *20* (3), 28–42.
7. Marburger, D.R. Arbitrator compromise in final offer arbitration: evidence from Major League Baseball. *Econ. Inquiry* **2004**, *42* (1), 60–69.
8. Miller, R.L. High risk final offer interest arbitration in Oregon. *J. Col. Neg. Pub. Sec.* **1999**, *28* (4), 265–279.
9. Champlin, F.C.; Boggnano, M.F. Chilling under arbitration and mixed strike-arbitration regimes. *J. Lab. Res.* **1985**, *6* (4), 375–387.
10. Graham, H.; Perry, J. Interest arbitration in Ohio, the narcotic effect revisited. *J. Coll. Neg. Pub. Sec.* **1993**, *22* (4), 323–326.

Incrementalism, Budgetary

Carol W. Lewis

Department of Political Science, University of Connecticut, Storrs, Connecticut, U.S.A.

INTRODUCTION

Incrementalism is an approach to decision making in general and to budgeting in particular. It is used in descriptive and explanatory variants and may refer to the decision-making process or outputs from the process. This perspective relies upon established roles, decision rules, and shared expectations; it broadly divides budgeting into the base and relatively small changes or marginal adjustments; decision making is characterized by a process of “successive limited comparison;” and politics is the dominant explanatory variable.

DEFINITION

Aaron Wildavsky initially described budgeting as incremental, meaning that it is “based on last year’s budget with special attention given to a narrow range of increases or decreases.” The focus is on “relatively small increments to an existing base” and a “small number of items over which the budgetary battle is fought.”^[1] The literature of public administration and public budgeting include numerous definitions.^[2–4] One definition sees incrementalism as:

... a theory of the budgetary process proposing that policy makers give only limited consideration to small parts of the budget and arrive at decisions by making marginal adjustments in last year’s budget... participants [in different roles] make decisions through a process of bargaining and negotiation.... The budget as a whole is not considered.... Instead, participants make marginal changes on an already existing base.^[5]

There are four central ingredients to budgetary incrementalism. The first is that change is marginal and occurs by small steps. The second ingredient is the base, which answers the question: Change from what? Wildavsky defined the base as the “general expectation among the participants that programs will be carried on at close to the ongoing level of expenditures” and operationalized the concept using appropriations.^[1] As a result, “[t]he largest determining factor of the size and content of this year’s budget is last year’s budget. Most of the budget is a product

of previous decisions.”^[1] (His illustrations include standard items, long-range commitments, mandatory programs, and sacred cows.) He stated the following:

The previous year’s budget, the largest part of which is composed of continuing programs and prior commitments, is usually taken as a base needing little justification.... Attention is normally focused on a small number of incremental changes, increases and decreases, calling for significant departures from the established historical base of the agency concerned.^[6]

The idea of *fair share* is the third ingredient and invokes open, pluralist decision making moderated by participants’ shared expectation of short-term proportionality and longevity. All players anticipate sharing in the action (increases or decreases) and returning again to compete for resources in future budget cycles. This concept is operationalized by budget share, the relative size of the budget going to a program or agency.^[1] As a result, incremental budgeting is often depicted as limiting the scope and intensity of competition. Illustratively, *A Citizen’s Guide to the Federal Budget*, a federal resource published annually for the general public, lays out major categories of spending according to the budget share of each.^[7]

The fourth and last ingredient is roles, of which there are two main types: program advocate or promoter and resource guardian or naysayer. Roles are especially important because, with incrementalism, which relies upon politics as an explanatory factor, decisions are seen as outcomes from negotiations and accommodations among a stable group of players, including the chief executive, bureaucrats, professional budgeters, legislators, and others. Long-term relationships and, therefore, strategy develop as players come back to the process in each budget iteration.

Participation, like roles, is believed to hold down competition in the process. “Whatever else they may be, budgets are manifestly political.... Participants in budgeting... drastically simplify their task by concentrating on the relatively small portion of the budget that is politically feasible to change.”^[6] As a political perspective on decision making, incrementalism:

assumes that political pluralism, limited rationality, and time and resources pressures on public administrators are significant constraints on decision making.

Consequently, administrators typically can take only small steps toward some general policy objective.... The test of the desirability of a decision tends to be measured in terms of political support or opposition, rather than cost-effectiveness or scientific analysis.^[4]

INTELLECTUAL FOUNDATIONS

Formative contributions to incremental decision making were made by Herbert A. Simon and Charles E. Lindblom. It is ironic that incrementalism, developed by economists and strongly influenced by psychology as well as economics, emphasizes the political aspects of decision making. One budgeting textbook asserts, "The most important characteristic of the muddling through or incrementalism approach as applied to budgeting is its emphasis on the proposition that budgetary decisions are necessarily political."^[8]

In his *Administrative Behavior* published in 1947, Simon emphasized observable behavioral factors and decision-making processes over the rationalist precepts, structures, and norms associated with the preceding orthodox perspective on administration.^[9] He argued that, because of cognitive limits, external developments, and other constraints, decision makers operate within *bounded rationality*. In 1958, James G. March and Simon introduced a new word to the vocabulary of administration: *satisficing*.^[10] This refers to the "process of finding a decision alternative that meets the decision maker's minimum standard of satisfaction.... Agencies must satisfice by accepting partial solutions rather than pursue illusory, perfect, or optimal solutions."^[2]

In his now classic article, "The Science of 'Muddling Through,'" first published in 1959, Lindblom built upon these ideas about cognitive limits and decision-making strategies.^[11] "The incrementalist paradigm he advanced is uniquely comprehensive, including elements of pluralism, satisficing, bounded rationality, organizational drift, limited cognition, and decentralized authority."^[2] Lindblom's central thesis is that there are two distinct varieties of decision making. One he calls the rational-comprehensive or root method, and the second, the successive limited comparisons or branch method.^[12] The latter "assumes that a historic chain of decisions exists which the administrator can use as a basis for future choices."^[2] Accordingly, "public administrators pragmatically select from among the immediate choices at hand the most suitable compromise that satisfies the groups and individuals concerned" and actual decision making is incremental, "for small steps are always taken to achieve objectives, not broad leaps and bounds."^[12] Lindblom revisited his view of "muddling through" in a 1979 article.^[13]

The foremost scholar of budgetary incrementalism, Aaron Wildavsky applied the concepts associated with

Simon and Lindblom specifically to budgetary decision making with the publication in 1964 of the first edition of his *Politics of the Budgetary Process*.^[1] Much of the literature in public administration and in budgeting for the next two decades was devoted to examining incremental decision making empirically and theoretically.

IMPACT

One reason for incrementalism's profound influence is that it criticized and ultimately outweighed the faith in rationalism and progress that had dominated public administration since the latter part of the nineteenth century. As the predominant model of decision making and particularly budgetary decision making, the incremental model of decision making has been applied to federal, state, and local governments in the United States and to other governments around the world.^[14-16] "Incrementalism has been the favored political approach to public budgeting in the United States."^[4] Indeed, it served as the conceptual core of budgetary decision making; its basic concepts and writings were routinely featured in texts, scholarly collections, and dictionaries.^[2] Lindblom's "The Science of 'Muddling Through'" is familiar to every student and scholar in public administration and Aaron Wildavsky's *Politics of the Budgetary Process*, once the single most widely used text in the field, went through many editions since its initial publication in 1964.

CRITIQUE

Incrementalism quickly came under fire for analytic biases, scientific shortcomings, and normative implications ascribed to it. A wide-ranging and influential critique written by Lance LeLoup concluded that its "self-fulfilling nature renders incrementalism nearly useless for social science theory."^[14] Another critique acknowledged that incrementalism is "intuitively and experientially appealing," but with a "lure" attributable to "the chaos in human behavior" rather than to its meeting epistemological standards.^[17] At least nine major criticisms have been leveled at incrementalism as a decision-making theory or heuristic model.

1. Incrementalism is associated with bias toward stability and conservatism.^[3,14] Wildavsky asserts, "Once enacted, a budget becomes a precedent; the fact that something has been done once vastly increases the chances that it will be done again. Since only substantial departures from the previous year's budget are normally given intensive scrutiny, an item that remains unchanged will probably be

carried along the following year as a matter of course.”^[1] The thrust, then, is this: the best predictor of an agency’s budget request is history. Of course, stability is anticipated only in the short term, because annual reiterations allow for rapid change, albeit in small steps, compounded over a few years. Lindblom’s 1979 response to this criticism was that “incrementalism in politics is not, in principle, slow moving. It is not necessarily, therefore, a tactic of conservatism. A fast moving sequence of small changes can more speedily accomplish a drastic alteration of the status quo than can an only infrequent major policy change.”^[13]

2. In much of the research, aggregation is at high levels that smooth change at the agency or governmental level and mask underlying changes in programs. Variance analysis is a standard tool in budget analysis and is often included in Comprehensive Annual Financial Reports. Variance analysis at as low a programmatic or organizational level as possible yields a more accurate count. Variance is calculated by disregarding positive and negative signs and simply adding together *all* changes from the prior base. An increase of \$150,000 in one program and a cut of \$150,000 in another in the same agency do not cancel each other out, but rather sum to a \$300,000 variance or change.
3. Predictive power in effect is sacrificed in favor of descriptive accuracy, because incrementalism fails to link decision-making dynamics to specific budgetary results. One slant on this issue is that incrementalism addresses only a small part of what budgeting entails. Surely, demand is a factor weighing on the budget in a politically responsive system, where projected changes in the use of a service or good such as public parks may translate into changes in spending. However, there are numerous other factors pressing on budgets that cannot be directly attributable to current decision makers. In fact, a great proportion of governmental budgets are settled before decision makers even get to work. Debt service on long-term debt incurred by previous decision makers is a legal obligation that current budgets must honor; the same is true of accrued liabilities such as pensions and other personnel benefits. Multi-year labor and service contracts and economic conditions such as inflation, unemployment, gross domestic product, and personal income may generate changes in prices and needs as, for example, reflected in the eligibility pools for entitlement programs. (The public is most familiar with the Consumer Price Index at

<http://stats.bls.gov/cpihome.htm>.) Demographic and other social changes (e.g., age structure, household formation) affect services from schools to public safety to elderly services and public works. Changes in programs and services that are required by law (or *mandated*) may mean changes in service levels, beneficiary eligibility, and benefits. (Some changes may be inter-governmental mandates that are addressed at <http://www.cbo.gov/cost.html>.) Courts also may mandate program changes with significant budgetary impact on, for example, prison conditions or services for the mentally ill. Many of these changes substantially reduce decision-making discretion over a given fiscal year’s budget, and budget outputs are pushed by factors other than incremental decision making. This critique suggests that, while incrementalism is an accurate depiction of a small-to-trivial part of the budget, it fails to address the major factors influencing budgetary outputs.

4. The value or size of marginal change is not specified, although Wildavsky used plus or minus 10% of the prior year’s appropriation. Lindblom argued, “Where the line is drawn is not important so long as we understand that size of step in policy making can be arranged on a continuum from small to large.”^[13] Specific variables defined in terms of clearly empirical referents are necessary if research is to be understood, replicated, and useful.
5. The major explanatory variable, the base, shares the problem of specificity. Scholarly work features differing concepts of the base, which may refer to prior year’s appropriation or current services. Current services refers to the resources necessary to continue programs at the prior year’s level and often allows for inflation, changes in law, and other factors.
6. Incrementalism focuses on expenditures to the exclusion of revenue decisions and outcomes, which limits its usefulness for understanding decision making in jurisdictions in which the budget must be balanced.
7. Incrementalism’s application to decision processes or budgetary outputs or both fosters confusion between them.
8. Incrementalism fails to meet the generally accepted criteria of a scientifically valid model as specified in items three, four, and five, above.

Today, incrementalism is in disrepute among some scholars. For example, Irene Rubin argued that the norms of moderation and fairness have diminished, because transparency and long-term relationships have diminished.^[18] Incrementalism is deemed inadequate in

several major respects as a *theory* that explains and predicts outputs from budget processes.

CURRENT USAGE AND SIGNIFICANCE

Incrementalism is still an important influence on the practice and understanding of budgetary decision making. General texts on public administration, texts on public budgeting, and collections of works considered classics in these fields continue to address incrementalism.^[3,4,12,19]

Explanatory incrementalism is no longer “the dominant theory of budgeting” it was a quarter of a century ago.^[14] On the other hand, its descriptive, process-oriented form remains a useful device by which to portray routine budgetary decision making for many—but by no means all—budget decisions in many jurisdictions. While noting that “budget processes are complex and vary widely among states” in its 1999 report on state legislative budget procedures, the *National Conference of State Legislatures* categorizes a majority^[14] of the states as relying predominantly upon a “traditional/incremental” budget approach (and more, if legislatures such as Delaware, Illinois, and Massachusetts that combine this with another approach are included).^[20] As Wildavsky noted, there is “nothing to be gained by ignoring overwhelming evidence or the observation of everyday life.”^[1]

The information provided and its presentation guide attention in particular directions and trigger particular analytic concerns.^[21] It is especially important that an incremental mode of decision making is built into many of the basic budget documents, because the “way in which budgetary questions are framed has a great deal to do with the outcomes of contests for public dollars.”^[4] (Many state and city budget documents are available online via jurisdictions’ home pages and see also the listings under Internet Resources.)

Four signs suggest that descriptive incrementalism as an analytic orientation is practiced at least to some extent. All four point to process, not outputs or outcomes. These signs include the following: 1) historical spending arrayed in columns alongside the current requests, as shown in the *State of Washington’s spreadsheets on expenditures history*, the *New York City Mayor’s proposed budget for fiscal 2002* with financial summaries and fire department data, and the summary schedule of spending in *A Citizen’s Guide to the Federal Budget for Fiscal 2002*; 2) the concept of current services as a point of departure for decision making, as illustrated by Section 1 of the *State of Washington’s operating budget instructions* for fiscal years 1999–2001; 3) an emphasis on change, as in many budget messages and narratives that

highlight new priorities and initiatives, as shown in the budget highlights for the *Illinois State Police* in the executive budget proposed for fiscal 2002; and 4) special detail on and scrutiny of new spending. *Connecticut’s legislative Office of Fiscal Analysis* noted in its description of budget formulation that agencies are “required to present a current services budget plan and a separate list of programmatic options if changes in expenditures or revenues are requested.” And, Sections 5 and 6 of the *State of Washington’s operating budget instructions* illustrate justification requirements.

Conventional formats display figures from prior years, the base, alongside figures for the current cycle. Some jurisdictions specify that the base is defined in terms of current services, meaning that budget requests allow for inflation, changes in law, and other factors. Change, expressed in percentages or dollars, is the decision-making focus in many budget documents; the executive’s budget message and summary narratives and schedules usually lay out proposed changes in detail. Instruction forms sent to the agencies at the beginning of budget preparation commonly require more elaborate justification of proposed changes than of ongoing operations; as a result, under ordinary circumstances, most of the base goes largely unexamined, and new competitors for the always scarce resources are relatively disadvantaged.

INTERNET RESOURCES

1. U.S. federal budget process—*A Citizen’s Guide to the Federal Budget, Budget of the United States Government*: <http://www.access.gpo.gov/usbudget>. Federal offices and agencies:
 - a. Office of Management and Budget: <http://www.whitehouse.gov/OMB>.
 - b. Bureau of Labor Statistics: <http://www.bls.gov/>.
 - c. Bureau of the Census and its *Statistical Abstract*: <http://www.census.gov/>.
 - d. Bureau of Economic Analysis: <http://www.bea.doc.gov>.
 - e. General Accounting Office: <http://www.gao.gov>.
 - f. Congressional Budget Office: <http://www.cbo.gov>.
2. State budgets and finances: <http://www.census.gov/govs/www/state.html>. National Conference of State Legislatures: <http://www.ncsl.org>. Tax Foundation: <http://www.taxfoundation.org/>.
3. State and local links: <http://lcweb.loc.gov/global/state/stategov.html>.

4. Professional associations:

- a. Governmental Accounting Standards Board: <http://www.rutgers.edu/Accounting/raw/gasb/index.html>.
- b. Government Finance Officers Association: <http://www.gfoa.org>.
- c. National Association of State Budget Officers: <http://www.nasbo.org>.
- d. Association for Budgeting and Financial Management of the American Society for Public Administration: <http://www.fpac.fsu.edu/abfm>.
- e. American Association for Budget and Program Analysis: <http://www.aabpa.org>.

REFERENCES

1. Wildavsky, A. *The Politics of the Budgetary Process*, 2nd Ed.; Little Brown: Boston, 1974; pp. xiii, 3, 13, 15, 17.
2. Chandler, R.C.; Plano, J.C. *The Public Administration Dictionary*, 2nd Ed.; ABC-CLIO: Santa Barbara, CA, 1988; pp. 28, 123, 128, 144, 145, 154, 156.
3. Henry, N. *Public Administration and Public Affairs*, 7th Ed.; Prentice Hall: Upper Saddle River, NJ, 1999; 355 pp.
4. Rosenbloom, D.H.; Goldman, D.D. *Public Administration: Understanding Management, Politics, and Law in the Public Sector*, 4th Ed.; McGraw-Hill: New York, 1998; pp. 32, 312, 318, 576.
5. LeLoup, L. *Budgetary Politics*, 4th Ed.; King's Court Communications: Brunswick, OH, 1986; 13 pp.
6. Wildavsky, A.; Hammond, A. Comprehensive versus incremental budgeting in the department of agriculture. *Adm. Sci. Q.* **1965**, *10* (3), pp. 322–323.
7. Office of management and budget. In *A Citizen's Guide to the Federal Budget for Fiscal 2002*; U.S. Government Printing Office: Washington, D.C., 2001 Chapter 2; on-line at *A Citizen's Guide to the Federal Budget*.
8. Lee, R.D., Jr.; Johnson, R.W. *Public Budgeting Systems*, 6th Ed.; Aspen: Gaithersburg, MD, 1998; 19 pp.
9. Simon, H.A. *Administrative behavior*, 2nd Ed.; Free Press: New York, 1957; originally published in 1947.
10. March, J.; Simon, H.A. *Organizations*; John Wiley and Sons: New York, 1958.
11. Lindblom, C.E. The science of “muddling through.” *Public Adm. Rev.* **1958**, *19*, 79–88.
12. Stillman, R.J., II. *Public Administration: Concept and Cases*, 7th Ed.; Houghton Mifflin: Boston, 2000, pp. 223–224.
13. Lindblom, C.E. Still muddling, not yet through. *Public Adm. Rev.* **1979**, *39* (6), 517–525.
14. LeLoup, L. The myth of incrementalism: analytical choices in budgetary theory. *Polity* **1978**, *x* (4), pp. 488, 492, 509.
15. Rickards, R.C. How the spending patterns of cities change: Budgetary incrementalism reexamined. *J. Policy Anal. Manag.* **1984**, *4* (1), 56–74.
16. Ezzamel, M. Organization change and accounting: understanding the budgeting system in the organizational context. In *Organization Studies*; European Group for Organizational Studies, 1994; (#213), on-line at <http://www.ptg.djnr.com/ccroot/asp/publib/story>.
17. Wanat, J. Bases of budgetary incrementalism. *Am. Polit. Sci. Rev.* **1974**, *68* (3), 1221.
18. Rubin, I.S. *The Politics of Public Budgeting*, 3rd Ed.; Chatham House: Chatham, NJ, 1997.
19. Shafritz, M.J.; Russell, E.W. *Introducing Public Administration*, 2nd Ed.; Longman: New York, 2000; pp. 52, 451.
20. National conference of state legislatures. In *Legislative Budget Procedures: A Guide to Appropriations and Budget Processes in the States, Commonwealths and Territories*; 1999; (May) on-line at <http://www.ncsl.org>.
21. Schick, A. The road to PPB, the stages of budget reform. *Public Adm. Rev.* **1966**, *26*, 243–258.

Indonesia

Rainer Rohdewohld

German Agency for Technical Cooperation (Deutsche Gesellschaft fuer Technische Zusammenarbeit—GTZ) (Support for Decentralization Measures Project/SfDM) (GTZ-SfDM), Jakarta, Indonesia

INTRODUCTION

More than 5 years after the downfall of the “Orde Baru” regime of ex-President Suharto, Indonesia continues to face a multitude of challenges. The political transformation into a competitive multiparty system reduced the political influence of the armed forces and realigned power between the state institutions. The emergence of vocal, but yet disorganized and therefore often still ineffective, civil society organizations is supported by a free but not yet fully professional mass media. The public sector continues to be characterized by inefficiencies, corruption, and lack of regard for the needs of the society at large. Last but not least, consolidation of economic recovery to provide employment opportunities for the millions of school graduates entering the labor market every year faces an increasingly competitive Southeast Asian environment.

In this context, the Indonesian public administration system needs thorough and far-reaching structural and procedural reforms. It is under increasing pressure to improve performance and accountability. However, public sector reform (with the exception of privatizing selected state-owned enterprises) has largely been neglected in the reform agendas of both President A. Wahid (November 1999–July 2001) and Megawati Soekarnoputri (since 23 July 2001). Institutional or individual “champions of reform” are lacking.

BACKGROUND AND TRADITION

The present Indonesian system of public administration reflects a variety of influences: Dutch colonial rule, indigenous (especially Javanese) customs and values, and the legacy of the “New Order” period of President Suharto (1965–1998).^[1] Centralization of power and the paternalistic attitude of the bureaucracy toward the public have been reinforced by the Suharto regime’s instrumentalization of the public administration to stimulate and direct economic and social development after 1967. Accelerating economic growth, improving social welfare, and maintaining political stability were major achievements of the administration during that period,^[2] during which the bureaucracy, together

with the military, became the most powerful state institution. Despite frequent initiatives for economic deregulation since the early 1980 which increased the role of the private sector in economic and social development, and despite the widespread absorption of public sector reform concepts from abroad, the notion of “development administration” continues to find much sympathy with civil servants and political leaders.

THE CIVIL SERVICE

Indonesia has a civil service system based on life-long appointments of civil servants. There are four service ranks (*golongan*) according to educational background, which are divided into four to five grades, giving the civil service a total of 17 service levels. Civil service positions are either staff positions with no specific professional or managerial responsibilities, managerial positions (*jabatan struktural*), or professional/technical positions (*jabatan fungsional*). The latter have their own specific recruitment requirements and career paths, based on technical/professional skills and expertise. Managerial positions are divided into four levels (*eselon*), echelon I level being the highest.

The total size of the Indonesian civil service was around 3.9 million in 2002, thus being slightly below the regional average of 2.6% of the population.^[3] In addition, there is a huge number of contractual staff (especially at the regional level) who do not have civil servant status. There are concerns about the sectoral and geographical distribution of staff^[4] and about the quality of available skills and competencies: only around 25% of the civil servants have tertiary education, and nearly half (around 1.8 million) are classified as administrative-auxiliary staff (*tenaga administratif*).^[5] A civil service survey in mid-2003 found more than 100,000 fictitious civil servants on the public sector payroll, indicating the weakness of existing internal management and control mechanisms (Table 1).

The Indonesian civil service has been observed to show the negative features associated with closed career systems: lack of managerial flexibility and accountability, limitation of the pool of leadership talent

Table 1 Distribution of civil servants, 1989–2002

	1989		1994		1999		2002	
		in %		in %		in %		in %
Central Government	3.151.661	86.9	3.471.595	87.5	3.519.959	87.9	930.602	23.7
Regions	475.954	13.1	494.183	12.5	485.902	12.1	3.002.164	76.3
Total	3.627.615	100	3.965.778	100	4.005.861	100	3.932.766	100

Source: Ref.^[9].

to insiders, dominance of seniority and patronage in promotion and appointments, and lack of incentives to perform well.^[3] Low pay has often been cited as a reason for low performance and for the widespread corruption in the administrative system; however, World Bank studies have concluded that wage bill increases will not by themselves achieve improved outcomes from the public sector,^[6] and that low pay cannot be regarded as the main driving force for corrupt civil service behavior.^[7]

In May 2002, the State Ministry for Administrative Reform presented an outline of reform initiatives, which included reviewing the role and functions of government agencies, regulating working mechanisms, increasing human resource capacities, improving performance appraisal systems, curbing corruption, and increasing the quality of public services.^[8] Considering the disappointing track record in public sector reform and the lack of strategic direction from the political leadership, observers have little optimism about the chances of such a reform program succeeding.

Several government agencies compete in the management of the civil service system: formally, the State Ministry for Administrative Reform [*Kementerian Perdayagunaan Administrasi Negara*—(MenPAN)] is the lead agency for formulating and coordinating public administration policies, including civil service policies, policies regarding the procedures and mechanisms of delivering public services, and policies regarding the institutional setup of government agencies. The National Civil Service Agency [*Badan Kepegawaian Nasional* (BKN)] is in charge of personnel administration and sets national norms and standards for recruitment, promotion, and staff rotation. The National Institute for Public Administration [*Lembaga Administrasi Negara* (LAN)] determines national policies for civil service training, implements key training programs, and conducts research on public administration. The Ministry of Finance determines the budget ceilings for the government's personnel expenditures and for the formation of the individual central government agencies.

The revision of the civil service law in 1999 included the stipulation to establish an independent Civil Service Commission; however, this has not been implemented yet because of shifting policy priorities and internal resistance in the government's bureaucracy.

STRUCTURE OF THE CENTRAL GOVERNMENT ADMINISTRATIVE SYSTEM

The central government bureaucracy consists of several types of organizations: coordinating ministries (*Kementerian Koordinasi*) which oversee a range of government institutions, departments (*Departemen*) with policy making and implementing functions, state ministries (*Kementerian Negara*) which develop and coordinate government policies in specific areas such as research and technology, and non-departmental (technical) agencies [*Lembaga Pemerintah Non-Departemen* (LPND)] which implement government policies in their respective fields of jurisdiction. Currently (October 2003), there are 3 coordinating ministries, 17 departments, 10 state ministries, and 25 technical agencies. The state and cabinet secretariat, which includes the secretariat of the vice president, is the administrative institution supporting the work of the president and the cabinet. The president is formally at the apex of the central government administration and determines the government's policies, while the ministers are regarded as “assistants” to the president.

The organizational structure of a department consists of the minister as the politically appointed head of the department. A secretariat general is in charge of policy coordination and of internal services (such as personnel management, finance, and public relations), while directorate generals are the main operational units. All are headed by an echelon I official. Most departments have their own training, as well as research and development units. Departments can have subordinated implementation units [*Unit Pelaksana Teknis* (UPT)] in charge of more technical or administrative matters. Before decentralization, departments used to have deconcentrated offices in the regions. This is now limited to those departments implementing exclusive central government functions.

While the total number of management positions in central government agencies has decreased by around 60% since 1999, the number of echelon I positions nearly doubled during the same period. Some departments have drastically reduced their personnel expenditures, and others report drastic increases.^[9] Contrary to expectations, there has been no comprehensive

reconfiguration of the central government administration as a result of the decentralization policy.

THE CHALLENGES OF DECENTRALIZATION

The decentralization policy included the administrative transfer of more than 2 million civil servants from central government agencies to the regions. Regions had no discretion to select or to reject staff in line with the regions' quantitative or qualitative needs. As a result, all regions are overstaffed, and many have created a vast landscape of local government agencies that does not reflect their actual needs. Nepotism and the need to buy in political support have contributed to this mushrooming of local government agencies. Recently, the central government issued a new and controversial regulation on the organizational setup of local governments (Government Regulation No. 8/2003) in an attempt to limit the overall size of the local government sector.^[10]

The decentralization law grants the regions full authority to manage their regional civil service in accordance with national regulations, norms, and standards. The revised civil service law maintains key features of the Indonesian civil service as a unified system, in which entry requirements, preconditions for promotion, performance appraisal system, and all issues related to remuneration are determined by the national level and apply to all civil servants irrespective of whether they work at central or regional level. The continuing existence of a unified national civil service is regarded as a major factor in the context of maintaining national unity.

In general, the structure of administration at the local level consists of a regional secretariat [*sekretariat daerah* (Setda)], which coordinates the formulation of regional policies and provides administrative support to the head of region, of technical agencies (*dinas*) which implement services in the main sectors (such as infrastructure, health, and education), of other technical institutions (*lembaga teknis daerah*) including a regional development coordination agency [*Badan Perencanaan Pembangunan Daerah* (BAPPEDA)], and subordinated implementation units [*unit pelaksana teknis daerah* (UPTD)]. Local government agencies not only implement decentralized tasks (*tugas desentralisasi*), but can also be tasked to implement central government tasks in their respective geographical areas (*tugas dekonsentrasi*).

Part of the decentralization policy was a radical overhaul of the system of local government financing. The percentage of public funds managed by the regions nearly doubled, and in 2002, regions controlled public funds representing 5.6% of the GDP.^[11] The new system of fiscal transfer consists of untied block grants

[*Dana Alokasi Umum* (DAU)], regional shares on certain taxes and revenues from natural resources, and specified grants [*Dana Alokasi Khusus* (DAK)] in priority policy areas of the government. Regions continue to be financially dependent on central government transfer, with own-source revenue in most cases remaining below 10% of the total regional revenue.^[12,13]

PUBLIC ADMINISTRATION REFORM

The lack of pace in reforming the public administration is deeply linked with the issue of public sector corruption. Indonesia has consistently been rated as one of the most corrupt countries in Asia and the world. Public sector corruption involves petty corruption where citizens have to pay a few thousand rupiah to obtain certificates and official documents, the taking of commissions and kickbacks by officials for the procurement of goods and services, the large-scale embezzlement of public funds by senior officials, and corruption within the civil service where appointments and promotions are being bought.^[14] Any attempt to streamline the public administration and to introduce innovative concepts of public sector management is facing the major hurdle of how to eradicate corruption first. Some steps have been taken after 1998: in 1999, parliament passed the Clean Government Law No. 28/1999 which requires state and public officials to declare their assets before assuming their posts. They also have to agree to have their assets officially audited during and after their terms of office. Law No. 31/1999 on the Eradication of Criminal Acts of Corruption provides the legal basis for establishing charges and procedures for prosecution. Political support for a consistent enforcement of both laws has been lacking, and only recently has the required Anti-Corruption Commission been set up. Because corrupt behavior is so deeply embedded in public sector behavior, only a concerted effort with full political support can hope to achieve significant improvements.

There are some areas where reforms have been initiated to improve public sector performance and to make public administration more accountable. Efforts are being made to reform the public financial management system,^[15] with the recently enacted Law on State Finance (Law No. 17/2003) being a key building block in this reform strategy.^[16] Two other laws (on state treasury and on state audit) are currently being debated in the parliament. Efforts are being made to reform the public procurement system.^[17] The emphasis of Law No. 17/2003 on performance budgeting is being complemented by a Ministerial Decree of the Ministry of Home Affairs (KepMendagri No. 29/2002) which requires all regions to use performance

budgeting as of 2004. While these efforts have their own merits, they lack integration in a wider public sector reform strategy, which is certainly not on the present agenda of the political leadership.

CONCLUSION

In its 2002 Governance Assessment of Indonesia, the Asian Development Bank (ADB) came to the conclusion that “the national system of administration and the civil service system are not sufficiently conducive to good governance and improved performances. Both systems need fundamental reform taking internationally recognized good practices into account.”^[14] Increasing the demand for accountability of public institutions and public officials is one option to press for public sector reform. It is no coincidence that if positive changes of the public administration occur, they occur at the local level,^[18] where decentralization and democratization put increasing pressure on local governments to justify their performance toward their electorate. Improving public finance management by fully enforcing the new law on state finance and by concentrating public finance functions in the Ministry of Finance will also have a positive impact on public sector performance. Incremental reform efforts will be more effective if integrated in wider, medium-term reform strategy which receives the political report needed.

REFERENCES

1. Rohdewohld, R. *Public Administration in Indonesia*; Montech Pty Ltd.: Melbourne, 1995. (For the Graduate School of Government, Monash University.)
2. Legowo, T. The bureaucracy and reform. In *Indonesia: The Challenge of Change*; Baker, R.W., Hadi Soesastro, M., Kristiadi, J., Ramage, D.E., Eds.; Institute of Southeast Asian Studies: Singapore, 1999; 81–98.
3. Research Institute for Asia and the Pacific, University of Sydney. *Public Sector Challenges and Government Reforms in South East Asia (Report of the Building Institutional Capacity in Asia Project Commissioned by the Ministry of Finance, Japan)*; University of Sydney: Sydney, 2001; 56–57.
4. MenPAN (Menteri Pendayagunaan Aparatur Negara), Sambutan, Delivered at the Workshop Pemantapan Kebijakan Rasionalisasi (Penataan) Pegawai Negeri Sipil, Jakarta, 2 July, 2002, 2. (mimeo).
5. Hardijanto. Pokok-pokok Pikiran Tentang rasionalisasi pegawai negeri sipil, Paper Delivered at the Workshop Pemantapan Kebijakan Rasionalisasi (Penataan) Pegawai Negeri Sipil, Jakarta, 2 July, 2002; 2–3. (mimeo).
6. World Bank. *Indonesia Civil Service Review, Summary Report*; 1999. Unpublished paper, Jakarta. (mimeo).
7. Filmer, D.; Lindauer, D.L. *Does Indonesia Have a “Low Pay” Civil Service?* Policy Research Working Paper, Development Research Group, World Bank: Washington, DC, 2001; Vol. 2621.
8. MenPAN (Menteri Pendayagunaan Aparatur Negara), Kebijakan Pendayagunaan Aparatur Negara, Disampaikan pada Sidang Kabinet Terbatas Tanggal, Mei, 23, 2002; Jakarta, 2002; (mimeo).
9. Rohdewohld, R. Decentralisation and the Indonesian bureaucracy: major changes, minor impact? In *Local Power and Politics in Indonesia. Decentralisation and Democratisation*; Aspinall, E., Fealy, G., Eds.; Institute of Southeast Asian Studies: Singapore, 2003; 259–274.
10. GTZ-Support for Decentralization Measures (SfDM). *Comments on Government Regulation No. 8/2003 Regarding the Organisational Structure of Regional Governments*; GTZ-SfDM Discussion Paper, 2003; Vol. 1/2003. SfDM: Jakarta. Download from http://www.gtzsfdm.or.id/documents/about/sfdm_pa_doc/DiscussionPaper_PerangkatDaerah_FinalFebr2003.pdf (October 2003).
11. Kadjatmiko. Perimbangan Keuangan Pusat dan Daerah dalam Kaitannya dengan Kebijakan Perhitungan Belanja Pegawai, Paper delivered at the Workshop Pemantapan Kebijakan Rasionalisasi (Penataan) Pegawai Negeri Sipil, Jakarta, 2 July, 2002; (mimeo).
12. Lewis, B. *Revenue-Sharing and Grant-Making in Indonesia. The First Two Years of Fiscal Decentralization*; 2002. Unpublished paper; (mimeo). Download from http://www.gtzsfdm.or.id/documents/dec_ind/o_pa_doc/BLewis_FiscalDecen2002.pdf (October 2003).
13. Fane, G. Change and continuity in Indonesia’s new fiscal decentralisation arrangements. *Bull. Indones. Econ. Stud.* **2003**, 39 (1), pp. 87, 159–176.
14. Asian Development Bank (ADB). *Country Governance Assessment Report—Republic of Indonesia (draft)*; Asian Development Bank: Jakarta, 2002; (mimeo).
15. Ministry of Finance (Republic of Indonesia). *The White Paper. Reform of the Public Financial Management System in Indonesia: Principles and Strategy*; Ministry of Finance: Jakarta, 2002. (Publication Series: 2002/KPMK/VII/MK/003).
16. MacIntyre, A.; Resosudarmo, B.P. Survey of recent developments. *Bull. Indones. Econ. Stud.* **2003**, 39 (1), 133–156.
17. World Bank, Indonesia. *Country Procurement Assessment Report. Reforming the Public Procurement System*; World Bank Office Jakarta: Jakarta, 2001. (Report No. 21823-IND).
18. Kusumaatmadja, S. Indonesia’s Current Crisis and the Impact on Civil Service Reform, Paper Presented at the Public Policy Challenges for Indonesia Symposium, Jakarta, 18 September, 2002; (mimeo).

Information Sources and State Policy Making

Jill Clark

Department of Political Science, University of Texas at Arlington, Arlington, Texas, U.S.A.

INTRODUCTION

In the last three decades, many American state legislatures have adopted reforms that expand institutional information sources (more professional staff and fewer committees) and encourage legislative independence from information provided by interest groups or the executive department. Studies suggest that individual legislators sometimes prefer information from constituents, other legislators, and legislative staffs. However, interest groups are also key information sources. Committees rely heavily on information from the executive department and interest groups, but information from legislative staffs is important as well.

Outside information (e.g., from other states or from national policy organizations) is transmitted through contacts made by the legislative and executive staffs and legislators, or through interstate interest group networks. This outside information may be most important for scientific issues, economic development policies, and, perhaps, for legislative leaders and minority representatives. Future research might focus on whether different information sources are more or less important at different stages of the policy-making process, for various policy areas or types, in different legislative contexts, or for particular legislators such as leaders.

LEGISLATIVE REFORMS: INFORMATION SOURCES

In the 1960s, “good government” reformers (e.g., The Citizens Conference on State Legislatures) advocated that American state legislatures enhance their policy-making capacity by developing institutional information sources: legislative research bureaus and larger and more professional committee staffs.^[1] Reformers also favored the reduction of committees so that legislators could develop substantive policy expertise with a limited number of committee assignments.^[2] Reformers argued that as a result of professional staffing and “expert” legislators, committees would become islands of policy information and recommend workable, efficient legislation. More importantly, legislators would no longer need to rely so heavily on information from interest groups.

Reformers also encouraged the adoption of an executive budget for the governor, so that professional staff

could provide information for policy making. However, some reform groups maintained that legislative reliance on information from the executive department would compromise the independence of the legislature. Thus, preserving the independence of the legislature became another rationale for developing professional information sources inside the legislature.^[3]

STATE LEGISLATURES: THE RELATIVE IMPORTANCE OF INFORMATION SOURCES

Research on information sources in state legislatures initially focused on the adoption of institutional reforms (fewer committees and larger legislative staffs).^[4] Researchers also surveyed legislators about the relative importance of information sources both for individual voting choices and for committee decision making. Constituents, the media, policy organizations, think tanks, and universities were included as possible information sources along with legislative and executive staffs, interest groups, and expert legislators.^[5–7] One expectation was that constituent information would be the most important source because a legislator’s primary concern was likely to be reelection or career advancement.^[8] In other words, legislator motivations for information were assumed to be quite different from those of reformers. Another expectation was that certain characteristics of information would be salient to legislators. As a result, legislators would prefer information from legislative colleagues or staffs (insiders) because insider information was likely to be reliable, timely, and easily accessible.^[9]

Overall, research findings suggested that there were multiple sources of information for legislative policy making, and that constituents and insiders were sometimes the most important information sources for individual members.^[10,11] A recent survey of legislators in 12 states indicated that grass-roots organizations, legislative staffs, statewide lobby groups, and national ethnic organizations were the four most important information sources.^[12] Furthermore, there were differences in information source preferences among legislators: Minority representatives used more sources of information than other legislators and were more likely to rely on information from conferences, the Internet, ethnic associations, and local branches of

national organizations. Other legislators were likely to rely more heavily on information from the media and statewide lobby groups.

Legislative committees tended to rely most heavily on information from executive agencies and interest groups.^[13] In other words, legislative staffs had not made committees independent of interest groups or executive agencies as primary information sources. However, policy information from legislative staffs was also important particularly on technical policy issues.^[14] Specialist committee members made the most contacts with outside sources.^[15,16] However, in general, policy organizations, the media, and universities did not rank highly as legislative information sources for committee decision making.

Most of the research on information sources has relied on surveys of legislators' reported preferences for information sources, and only a few studies analyzed the legislators' information use.^[17] Furthermore, some findings are based on a single state, and most are based on a small group of states. As a result, study results are tentative, and it is difficult to generalize across states or over time. There has been some interest in determining whether states with professional staffs and fewer committees choose different policies from those with fewer institutional resources. In general, the conclusion is that they do not; instead, policy differences are tied to differences in state environments. The most affluent states are those most likely to adopt legislative reforms and to choose generous expenditures for programs such as education.^[18] The same is true for policy innovation: Professional legislatures had little independent effect on whether states are leaders in adopting new policies.^[19]

However, in 1971, The Citizens Conference on State Legislatures found that reformed legislative institutions were more efficient,^[20] and the Advisory Commission on Intergovernmental Relations, in 1982, suggested that reformed state legislatures were more professional, open, responsive, and representative than unreformed legislatures.^[21] However, some observers argued that legislative staffs had developed too much power for influencing elected officials. Later, legislative reforms came under scrutiny by citizens who believed that state legislatures have become remote, careerist, and elitist. These arguments have resonated in some states: Legislative term limits have been enacted in 21 states, and both annual legislative meetings and staff size are targets for change in several other states.

INFORMATION SOURCES: INTERSTATE POLICY DIFFUSION

Although information sources research underscored the importance of information sources within a state,

cross-state diffusion research suggested that states quite frequently borrow policies from other states, especially neighbors. States often have similar policies, suggesting a flow of policy information among them and the use of that information in legislative policy making.^[22,23] Moreover, the tendency of states to adopt similar policies has increased in recent decades,^[24] perhaps in part because information technology facilitates both easier and timely access to information about other states' policies. Interstate tax base competition has stimulated similar business incentives policies,^[25] tax types or rates,^[26] state-run lotteries,^[27] and educational reforms.^[28] The notion of "welfare magnets" may encourage some states with the most generous welfare benefits to reduce those benefits in comparison with other states.^[29]

One possibility for reconciling the findings of the information sources and policy diffusion literatures is a two-step communication process where staffs, expert legislators, or interest groups acquire information about other states' policies and then provide that information to legislators. Surveys of executive department staffs suggest that bureaucrats are actively involved in professional networks and are aware of policy developments in other states.^[30-32] Specialist legislators typically report more contacts with information sources outside the state than other members. Certain individual characteristics such as ethnicity, educational level, or years of experience^[33] may differentiate specialists from other members. Some interest groups rely on interstate advocacy coalitions to diffuse policy innovations to state legislatures.^[34,35]

Although interstate staff and interest group networks have received some attention, there has been no research on the role of governors in acquiring information from outside the state.^[36] The National Governor's Association is typically described as an organization that presents governors' views to the national government as a means of enhancing the reputation of governors with presidential aspirations. The assumption seems to be that governors rely primarily on political parties or executive department staffs to provide specific policy solutions. Legislators, like governors, have access to several national policy organizations (e.g., the National Conference of State Legislators, the Council of State Governments, the National Legislative Exchange Council, and the State Government Affairs Council), and there is some evidence that legislators acquire information independently of staffs.

In fact, legislators now have more incentives for outside searches for policy information because legislators are more "careerist." Policy accomplishments or campaign issues can be part of a strategy to stay in office.^[37] In other words, policy information from outsiders may have political uses for legislators: job security and/or career advancement. Moreover, pressures from constituents, citizen groups, the media, public

opinion polls, and, in some cases, direct democracy^[38] could also motivate legislators to initiate their own policy proposals. Independent policy information could make the legislature an initiator of policy, rather than a more passive institution that responds to proposals from either the executive department or from the citizens.

National policy organizations (such as the Council of State Governments) offer a number of advantages as information sources for legislators. Professional staffs prepare research, and organizational meetings offer legislators contacts with peers from other states. Information is available on political feasibility, and there is policy evaluation on programs that are in place in other states.^[39] Internet communications make it possible to gain information quickly, and legislators can interact with peers. Thus, the same characteristics of information that appealed to legislators inside the institution are now available from national organizations: peer information and information that is reliable, timely, and easily accessible.

Policy solutions borrowed from national organizations can be adapted to the political context of a particular state or to a legislator's political agenda.^[40] One possibility is that constituents or interest groups inside the state define problems, but solutions are borrowed from national policy organizations. Alternatively, policy information may be more relevant to the initial formulation of legislation and constituent or insider information, to agenda setting, or to the adoption of legislation. Policy information may also be useful in persuading other legislators or the public, or in justifying existing political beliefs.^[41]

INFORMATION SOURCES: LEGISLATIVE LEADERS

Given their positions, legislative leaders may be most attuned to legislative independence, election agendas, and career advancement. As a result, leaders may engage in more outside information searches than members. Recent research from a national survey has focused on legislative leaders' (speakers, senate presidents, and minority and majority leaders) contacts with information, and their use of information from national policy organizations.^[42,43] Results suggested that leaders from "springboard legislatures" and/or from states without legislative term limits relied more on national organization information than leaders from dead-end or term-limited legislatures. Perhaps the most active leaders are motivated by the value of outside information in the context of legislatures that provide long-term career opportunities.

Additionally, both majority and minority leaders from states with one-party control of the legislative and executive institutions are more active in joining

national organizations and in attending meetings than those from states with divided-party control. It could be that for these leaders, national organizations provide possible policy solutions independent of instate actors, and leaders can claim credit for policy accomplishments.

Compared with leaders in states with larger staffs, leaders in states with small legislative staffs reported that national organization information was more important. Leaders from small states were more active in joining professional associations and in attending meetings than those from large states, possibly because small states have fewer instate information sources. Perhaps these leaders seek more independence from the executive department, or they see political advantages associated with outside information.

CONCLUSION

Future research might investigate the information sources of policy makers other than legislators. There has been no attention to information sources for governors, and there has been little interest in executive department staffs. Future research on legislative policy making might focus more on identifying information sources utilized at different stages of the policy-making process (agenda setting, formulation, and adoption) and in further distinguishing the information sources of committees from those of individual legislators. Information uses should include not only policy making, but also political advantage (career advancement). Perhaps, ironically, the reformers' goal of legislative independence may be embraced by legislators who see career opportunities in acquiring policy information. Furthermore, interstate policy diffusion may be at least partially a product of politicians' political ambitions.

Additionally, different information sources may be more or less important for certain policy areas. For example, outside information may be more salient for policies that involve interstate competition for economic development, or for scientific or technical issues. Information use may also be influenced by differences in legislative context. Perhaps outside information is more important in legislatures with limited staff capabilities and/or in legislatures with the most career opportunities. Legislative leaders may be more attuned to the political advantages of outside information than to its members. Women and minority legislators may be more likely to seek information from national organizations that emphasize their concerns.

REFERENCES

1. Bowman, A.O.M.; Kearney, R.C. *The Resurgence of the States*; Prentice-Hall: Englewood Cliffs, NJ, 1986; 84-88.

2. Morehouse, S.M. *State Politics, Parties and Policy*; Holt, Rinehart, and Winston: New York, 1981; 287–290.
3. Bingham, R.D.; Hedge, D. *State and Local Government in a Changing Society*, 2nd Ed.; McGraw-Hill, Inc.: New York, 1991; 287–290.
4. Wissel, P.; O'Connor, R.; King, M. The hunting of the legislative snark: information searches and reforms in U.S. legislatures. *Legis. Stud. Q.* **1976**, *1* (2), 251–267.
5. Sabatier, P.; Whiteman, D. Legislative decision making and substantive policy information: models of information flow. *Legis. Stud. Q.* **1985**, *10* (3), 395–421.
6. Shaw, G.M. The role of public input in state welfare policymaking. *Policy Stud. J.* **2000**, *28* (4), 707–720.
7. Bradley, R.B. Motivations in information use. *Legis. Stud. Q.* **1980**, *5* (3), 393–406.
8. Webber, D.J. State legislators' use of policy information: the importance of legislative goals. *State Local Gov. Rev.* **1985**, *17* (2), 213–218.
9. Mooney, C.Z. Peddling information in the state legislature: closeness counts. *West. Polit. Q.* **1991**, *44* (2), 433–444.
10. Mooney, C.Z. Information sources in state legislative decision making. *Legis. Stud. Q.* **1991**, *16* (3), 445–455.
11. Bradley, R.B. Motivations in information use. *Legis. Stud. Q.* **1980**, *10* (3), 393–406.
12. Jackson-Elmoore, C.; Knott, J.H.; Verkuilen, J.V. *An Overview of the State Legislators' Survey: Sources of Information and Term Limits Impacts*; W.K. Kellogg Foundation: Battle Creek, MI, 1998. <http://www.wkkf.org/Programming/ResourceOverview> (accessed September 2003).
13. Hamm, K.E. Patterns of influence among committees, agencies and interest groups. *Legis. Stud. Q.* **1983**, *8* (3), 379–426.
14. Guston, D.H.; Jones, M.; Branscomb, L.M. The demand for and the supply of technical information and analysis in state legislatures. *Policy Stud. J.* **1997**, *25* (3), 451–469.
15. Porter, H.O. Legislative experts and outsiders: the two-step flow of communication. *J. Polit.* **1974**, *36* (3), 703–730.
16. Sabatier, P.; Whiteman, D. Legislative decision making and substantive policy information: models of information flow. *Legis. Stud. Q.* **1985**, *10* (3), 395–421.
17. Webber, D.J. State legislators' use of policy information: the importance of legislative goals. *State Local Gov. Rev.* **1985**, *17* (2), 213–218.
18. Dye, T.R.; McManus, S.A. *Politics in States and Communities*, 11th Ed.; Prentice Hall: Upper Saddle River, NJ, 2003; 200 pp.
19. Nice, D.C. *Policy Innovation in State Government*; Iowa State University Press: Ames, IA, 1994.
20. The Citizens Conference on State Legislatures. *State Legislatures: An Evaluation of Their Effectiveness*. Praeger: New York, 1971.
21. Advisory Commission on Intergovernmental Relations. *State and Local Roles in the Federal System*. U.S. Government Printing Office: Washington, DC, 1982; 75–88.
22. Mooney, C.Z. Citizens, structures, and sister states: influences on legislative professionalism. *Legis. Stud. Q.* **1995**, *20* (1), 47–67.
23. Walker, J.L. Diffusion of innovations among the American states. *Am. Polit. Sci. Rev.* **1969**, *63* (3), 880–899.
24. Savage, R.L. Diffusion research traditions and the spread of policy innovations in a federal system. *Publius* **1985**, *15* (4), 1–28.
25. Eisenger, P.K. *The Rise of the Entrepreneurial State: State and Local Development Policy in the United States*; University of Wisconsin Press: Madison, WI, 1988.
26. Berry, F.S.; Berry, W.D. Tax innovation in the states: capitalizing on political opportunity. *Am. J. Polit. Sci.* **1992**, *36* (3), 715–742.
27. Barry, F.S.; Berry, W.D. State lottery adoptions as policy innovations: an event history analysis. *Am. Polit. Sci. Rev.* **1990**, *84* (2), 395–415.
28. Mintrom, M.; Vergari, S. Policy networks and innovation diffusion: the case of state education reform. *J. Polit.* **1998**, *60* (1), 126–148.
29. Peterson, P.; Rom, M.C. *Welfare Magnets: A New Case for a National Standard*; Brookings Institution: Washington, DC, 1990.
30. Chi, K.S.; Grady, D.O. Innovators in state government: their organizational and professional environment. In *The Book of the States*; Council of State Governments: Lexington, KY, 1995; 382–404.
31. Freeman, P.K. Values and policy attitudes among state legislators and administrators. *Public Adm. Q.* **1984**, *7* (4), 483–497.
32. Gray, V. Innovation in the states: a diffusion study. *Am. Polit. Sci. Rev.* **1973**, *67* (4), 1174–1185.
33. Freeman, P.K. Interstate communication among state legislators. *Publius* **1985**, *15* (4), 99–112.
34. Glick, H.R.; Hays, S.P. Innovation and reinvention in state policymaking: theory and the evolution of living will laws. *J. Polit.* **1991**, *53* (3), 835–850.
35. Mintrom, M. Policy entrepreneurs and the diffusion of innovation. *Am. J. Polit. Sci.* **1997**, *41* (3), 738–770.
36. Dresang, D.L.; Gosling, J.J. Governors. In *Politics and Policy in American States and Communities*; Allyn and Bacon: Boston, 1999; 216–218.
37. Loomis, B.A. *Time, Politics and Policies: A Legislative Year*; University Press of Kansas: Lawrence, KS, 1994.
38. Rosenthal, A. *The Decline of Representative Democracy: Process, Participation, and Power in State Legislatures*; CQ Press: Washington, DC, 1998.
39. Robertson, D.B. Political conflict and lesson-drawing. *J. Public Policy* **1991**, *11* (1), 55–78.
40. Hays, S.P. Patterns of reinvention: the nature of evolution during policy diffusion. *Policy Stud. J.* **1996**, *24* (4), 551–566.
41. Mooney, C.Z. Information sources in state legislative decision making. *Legis. Stud. Q.* **1991**, *16* (3), 445–455.
42. Clark, J.; Little, T.H. National organizations as sources of information for state legislative leaders. *State Local Gov. Rev.* **2002**, *34* (1), 38–44.
43. Clark, J.; Little, T.H. State legislative leaders' contacts with national legislative organizations: differences in cosmopolitanism, resources, and political environments. *Polit. Policy* **2003**, *31* (2), 282–295.

Information Systems

Mary Maureen Brown

Piper S. Charles

*Department of Political Science, University of North Carolina at Charlotte,
Charlotte, North Carolina, U.S.A.*

INTRODUCTION

An information system (IS) can be defined as a set of interrelated components, such as hardware, software, policies, communications, users, and data, that collect, manipulates, stores, and disseminates data and information that support the business activities of the users. As the public continues to place pressure on government to provide services cheaper, better, and faster, public managers have sought ways to streamline their operations to achieve those objectives and improve services. Information systems are the support mechanism to enhance and improve an organization's ability to carry out its primary mission and achieve its organizational goals. Through the use of information systems, government operations can be conducted more quickly, accurately, and efficiently. In addition, information systems permit government agencies to make a wide range of information available to the public, as well as conduct business electronically. By the end of 2000, nearly 40 million Americans were doing business with the government electronically.^[1]

Strategic information management is the process by which top agency officials and line managers plan, direct, and evaluate the use of information and information systems to help accomplish their programmatic objectives. Organizations need accurate, reliable, and timely information for strategic planning, identifying objectives, improving productivity, and facilitating service delivery. Technology is becoming the vehicle from which accurate, reliable, and timely information is produced to achieve organizational goals.

Government agencies can leverage the strengths of information systems to achieve efficiencies and operational gains, such as eliminating duplication and process bottlenecks, streamlining processes, increasing productivity, broadening the range and scope of service delivery, improving analysis and decision-making capabilities, enabling coordination of processes across distances and functions, reducing complexity, and adding value for the customer.

TYPES OF INFORMATION SYSTEMS

There are several types of information systems: office information systems, transaction processing systems, management information systems, decision support systems, expert systems, executive information systems, and geographic information systems.

Office Information System

An office information system improves workflow and communications among employees in an organization. The software supports activities such as word processing, spreadsheets, databases, presentations, e-mail, scheduling, and web browsing. It may even include voicemail, fax, scanner, and videoconferencing from the desktop computer.

Transaction Processing System

A transaction processing system collects and processes data on a routine basis, generally performing the same transactions on large amounts of data. The strength of these systems is their ability to process large amounts of data quickly and repetitively. Examples of transaction processing systems are a payroll system, utility billing system, or a computer-aided dispatch system for fire and police.

Management Information System

A management information system (MIS) is a system designed to support the information needs of managers to make decisions and solve problems within the organization. A management information system has the ability to analyze data and provide timely, accurate, and meaningful information in the form of reports. For example, a utility billing MIS system might produce reports to indicate high water usage in the summer months that managers may need to react

to, or identify water leakage based on unusual water usage. Water usage information can be useful for planning water line expansions or additions. Likewise, the police department can use MIS systems to examine call for service data to assist in resource deployment and to identify problem areas with repeat calls for service.

Decision Support System

A decision support system (DSS) is used more in business than in public organizations. A decision support system is used to provide a greater level of detailed information to support a particular decision. A DSS is able to handle large amounts of data from internal and external sources to support the problem-solving process. For example, some coastal jurisdictions are looking to decision support systems to determine the best time and method for evacuating coastal residents in the event of a hurricane. The DSS can calculate a number of factors, such as number of residents affected; characteristics of the hurricane, such as its rate of approach, wind speed, size, and potential for damage; available roads; and other factors to assist emergency management authorities in determining the best plan for evacuation.

Executive Information System

An executive information system (EIS) is similar to a decision support system designed to assist the decision-making process of the executive. An executive information system will often present the data in the form of charts and graphs to indicate pattern and trends and other statistics to assist the executive in developing strategic plans for the future activities and direction of the organization.

Expert System

An expert system is an information system that collects and stores the knowledge of human experts, draws conclusions from complex relationships, and provides the reasoning used to arrive at a suggested decision. The expert system can assist less-experienced users through the decision-making process. The expert system has two components: the knowledge base and inference rules. The knowledge base is the collection of human knowledge, and the inference rules are a set of business rules applied to the knowledge base to generate a suggested solution or decision. Expert systems are seen in the medical field to assist doctors in diagnosing diseases based on a patient's particular set of symptoms. A unique feature of the expert system

is its ability to capture human expertise that can be shared with other less-experienced users.

Geographic Information System

A geographic information system (GIS) can take data in tabular form that is associated with a location and present that data in a graphical format on a predrawn map. An example of the use of a GIS is in local area planning, where a map can indicate various zoning classifications over the county map. This geographic representation of data can aid managers and planners in developing zoning policy and future area development plans. Police departments often utilize GIS to identify areas of concentrated crimes, called "hot spots," or to identify patterns of crimes occurring in particular locations. Geographic identification aids police in developing tactics to address specific crime problem locations.

COMPUTER LITERACY VERSUS INFORMATION LITERACY

In order for the public manager to be able to develop appropriate information systems, the public manager must be computer literate as well as information literate. Computer literacy is the knowledge and understanding of hardware components, communications and networks, programming languages, and software applications, such as databases. Information systems literacy includes not only knowledge of computer technology but also the broader scope of what data and information are required by the organization's personnel and the customers in order to achieve organizational mission and goals. This includes how to find information, how to analyze the information, as well as how to use the information effectively. Understanding how to apply the various types of information systems within the organization to obtain maximum use of available data is the essence of information literacy. Knowledge management is the ability to derive knowledge from data and information to be applied to provide solutions to future business problems.

SYSTEMS DEVELOPMENT

Systems development is the process of creating, developing, or modifying information systems to support agencies' missions and goals. The systems development life cycle (SDLC) is a set of activities developed to assist and guide the development of information systems. The SDLC provides several steps, such as

planning, analysis, design, implementation, maintenance, and review.

The Clinger–Cohen Act of 1996 set requirements for performance-based and results-oriented decision making for all major developments of information technology at the federal level. Investments in information systems and technology can have a significant impact on an organization's performance but only when they are well managed and focused on supporting the organization's mission. A key goal of the Clinger–Cohen Act is for agencies to establish mechanisms to ensure that information system projects contribute in a measureable way to improved mission performance and are implemented in a reasonable time frame and at a reasonable cost.

These mechanisms include the first stage of selection: developing a portfolio of information system projects; identifying the costs, benefits, and risks associated with the projects; determining priorities; and selecting those projects that should be funded for the near future, based on how well the projects support the goals of the organization. The second stage involves controlling of the project and ensuring that the projects deliver the expected benefits on time and within budget. The third stage is the evaluation of projects and the use of the lessons learned in the next selection stage.^[2]

CHALLENGES AND OBSTACLES

The implementation of information systems in the public organization brings several challenges for the public executive. The public manager must ensure that strategic goals, information, and technology are integrated into common organizational procedures, ensure that effective management strategies are applied to all information technology adoption efforts, and ensure that everyone on the project is involved in risk management.

However, public managers are faced with obstacles often unique to public organizations. Although IT implementation is considered a high-risk endeavor in public and private sectors, the obstacles to successful implementation in the public sector tend to be greater for several reasons:

1. Public managers often are not motivated to implement large technology projects. Rarely does a public manager receive any incentive, monetary or otherwise, for a successful technology implementation that results in significant productivity gains. However, the public manager is fully exposed to public criticism if the system fails in any number of categories. As a

result, public managers tend to be conservative with technology implementation, relying on only proven technology.

2. Agency executives often do not have a clear line of authority for information system implementation. Often, their decisions must comply with existing laws, budget appropriations, legislative mandates, and a host of other political influences. Further, many agency executives have tenure that is short-lived, compared to the relatively long life cycle of system implementation.
3. Government agencies and programs are influenced by a number of stakeholders who often have competing interests and goals. The act of balancing and mediating these competing interests often stifles the creativity and implementation of information systems.
4. Most government agencies budgets are developed on an annual basis. The uncertainty of the future availability of funds to continue an information system project implementation as well as support and maintain the project after implementation hinders the ability of public managers to plan and implement information systems successfully. In addition, budgets and policy or program mandates are often legislated separately and may not correlate with each other.
5. Public agencies have more intangible and diverse goals and objectives with efficiency that is harder to measure. They provide services on the basis of public interest rather than services exchanged in the economic market.
6. The procurement mechanisms for government are cumbersome and time-consuming. While designed to maintain integrity and fairness in government acquisitions, the competitive procurement processes are often a source of problems and delays. These delays create additional complications for the procurement of technologies that are constantly changing.^[3,4]

SECURITY, PRIVACY, AND ETHICS

Security, privacy, and ethics are additional issues that must be considered during the development and management of public agency information systems. Security pertains to the protection of the components of information systems from loss, damage, or unauthorized access. Computer viruses and hackers are two examples of security breaches that can damage information systems or allow unauthorized access to and exposure of confidential information about citizens. System failure, electrical power loss or power surges, as well as the physical security of equipment from

tampering, theft, or damage from natural disasters can place information systems at risk.

Public agencies collect a wealth of information on individuals. Privacy refers to the right of these individuals to control or restrict the collection, storage, or dissemination of personal data without their knowledge or consent. This issue surfaced recently when several state departments of motor vehicles were said to be selling drivers' license information to third parties in order to earn revenue to ease strained budgets. Many citizens have a strong interest in who gains access to personal information contained in records retained by public agencies.

Ethical issues are of concern in not only how the information collected and retained by public agencies is used but also in the accuracy of the information. One example involves a man whose identity was "stolen" by another man who lives in another state. The other man has an extensive arrest record, and the man whose identity was stolen nearly lost a job opportunity because of his "arrest record." The problem is in placing responsibility on a particular agency for data accuracy and developing procedures for ensuring data accuracy. Unfortunately, in this case, the "victim" was unable to correct this inaccuracy, and the national crime computer still shows an arrest history for him. Other ethical issues involve the appropriateness of supervisors to monitor employee use of

the Internet, or track their whereabouts during work hours via the use of electronic key cards.

CONCLUSION

Information systems can provide accurate, reliable, and timely information to public agency managers to support the achievement of the agency missions and goals. In addition, information systems can be used to facilitate the delivery of services to the public (see the article *Electronic Government*).

REFERENCES

1. National Partnership for Reinventing Government. *Access America: E-Gov*. 2001, Available at: <http://govinfo.library.unt.edu/npr/initiati/it/index.html>.
2. United States General Accounting Office. *Assessing Risks and Returns: A Guide for Evaluating Federal Agencies' IT Investment Decision-making*. 1997.
3. Rainey, H.G. *Understanding and Managing Public Organizations*, 2nd Ed.; Jossey-Bass: San Francisco, 1997.
4. Dawes, S.; Kelly, K.; Anderson, D.; Bloniarz, P.; Cresswell, A. *Making Smart IT Choices: A Handbook*; Center for Technology in Government: Albany, 1996.

Innovation and Public Policy

Stuart S. Nagel

Political Science Department, University of Illinois, Urbana, Illinois, U.S.A.

INTRODUCTION

The purpose of this article is to discuss some aspects of the relations between public policy and creativity. There are basically two kinds of relations in this context. One set deals with how public policy can help stimulate creativity. The other set of relations deals with how creativity can be helpful in improving public policy.

On public policy as a stimulant to creativity, this article briefly discusses policies that relate to politics, economics, socialization, and psychology. On creativity in improving public policy, this article briefly describes pushing, facilitating, and pulling factors. It also discusses sources of policy goals and policy alternatives. It further discusses concepts that are useful in arriving at win-win public policies.

Win-win policies are alternatives that can enable conservatives, liberals, and other major viewpoints to all come out ahead of their best initial expectations simultaneously. Win-win is also called superoptimizing or doing better than the previous best of all major groups.

There are basically five steps to win-win policy analysis:

- What are the major goals of conservatives, liberals, or other major groups who are disputing what policy should be adopted for a given policy problem?
- What are the major alternatives of those groups for dealing with the policy problem?
- What are the relations between each major alternative and each major goal? In their simplest form, these relations can be expressed in terms of a minus sign (relatively adverse relation) and a plus sign (relatively conducive relation) and a zero (neither adverse nor conducive relation).
- What new alternative is there that might be capable of:
 - Achieving the conservative goals even better than the conservative alternative.
 - Simultaneously capable of achieving the liberal goals even more than the liberal alternative. Whatever new alternative meets these two criteria is a win-win alternative or a superoptimum solution (SOS).

- Is the proposed win-win alternative capable of getting over various hurdles that frequently exist? These hurdles may be political, administrative, technological, legal, psychological, and economic in random order. Win-win solutions should also consider how to upgrade workers and firms that may be displaced by downsizing due to increased productivity, free trade, defense conversion, immigration, merit treatment, labor utilization, creativity, and related factors.

PART ONE: PUBLIC POLICY IN STIMULATING CREATIVITY

The following political, economic, sociological, and psychological institutions or ways of doing things in a society are conducive to innovative and effective public policy making. This includes public policies that can enable conservatives, liberals, and other major viewpoints to come out ahead of their best initial expectations simultaneously, i.e., superoptimum or win-win solutions.

Political Methods

Competitive political parties

This is a key facilitator, because the out-party is constantly trying to develop policies (including possibly SOS policies) in order to become the in-party. The in-party is also busy developing new policies in order to remain the in-party. New policies are developed largely as a result of changing domestic and international conditions, not just for the sake of newness. Without the stimulus of an out-party, the in-party would have substantially less incentive to be innovative. More important, without the possibility of becoming the in-party, the out-party would lose its incentive to be innovative. More innovation generally comes from the out-party than the in-party (all other factors held constant), including the possibility of SOS innovations.

Better policy analysis methods and institutions

SOS solutions are likely to be facilitated by policy analysis methods that deal with multiple goals,

multiple alternatives, missing information, spreadsheet-based decision-aiding software, and a concern for successful adoption and implementation. Better policy analysis institutions refer to training, research, funding, publishing, and networking associations. These institutions can be part of the activities of universities, government agencies, and independent institutes in the private sector. The extent to which these policy institutions deal with superoptimizing analysis will make them even more relevant to facilitating SOS solutions.

Economic Policies

Competitive business firms

Competition among political parties may be essential for facilitating SOS public policy. Competition among business firms may be essential for facilitating a prosperous economy and a prosperous world through international business competition. Numerous examples can be given of nations that failed to advance and collapsed due largely to a one-party system, such as the former Soviet Union. Likewise, numerous examples can be given of business firms that failed to advance and virtually collapsed due largely to lack of substantial competition, such as the American steel industry. The American automobile industry has not collapsed, but it failed to develop small cars, cars that resist style changes, safer cars, less expensive cars, and more durable cars in comparison to the international competition that was not taken seriously until almost too late.

Well-targeted subsidies and tax breaks

In the context of superoptimum solutions, this tends to mean subsidies and tax breaks that increase national productivity and international competitiveness. Such subsidies and tax breaks are the opposite of handouts that provide a disincentive to increased productivity on the part of welfare recipients or big business. Good targeting in this regard especially refers to upgrading skills and stimulating technological innovation and diffusion. A dollar invested in those kinds of subsidies is likely to pay off many times over without necessarily having to wait very long for the results.

Increased national productivity

All these facilitators are important. Economists might rightfully consider increased national productivity to be especially important. It leads to an increased gross national product or national income, which means an increased tax base to which the tax rate is applied.

If increased productivity increases the tax base, then tax rates can be lowered and still produce more tax money for well-targeted subsidies that produce further increases in national productivity. These increases, however, are not an end in themselves. The increased national income can facilitate finding and implementing SOS that relate to employment, inflation, agriculture, labor, business, poverty, discrimination, education, families, the environment, housing, transportation, energy, health, technological innovation, government structures, government processes, world peace, international trade, and every other public policy-field. In other words, with more money and resources available, SOS solutions are facilitated, but SOS solutions often draw upon creativity that is associated with doing much better on relevant goals with constant or decreasing resources.

Sociology: Childhood Socialization

Risk-takers get generated from about age 0 to 5 in little children, depending on whether they are allowed to take chances or are treated in such a way that they never come in contact with anything that might hurt them. There is certainly a need for encouraging more experimentation on the part of children, within reason. More rebelliousness, more of the kind of trying out to see what will happen if you push your food off the highchair onto the floor without being punished for doing so, to see if the bowl will break or not. That does not necessarily mean that you jump off the third-story porch to see if your head will break.

Liberals have a lot of trouble talking about socialization, because it sounds like brainwashing people. It can be done in a brainwashing way, or it can be done in a way that encourages children to think things out for themselves to some extent. An example might be telling children not to discriminate on the basis of race or gender, as contrasted to setting up a situation where they more creatively reason that discrimination is undesirable. Such a situation might involve the teacher calling for volunteers to erase the blackboard and virtually everyone volunteers. The teacher says we cannot have so many people erasing the blackboard, and we might therefore just pick the African-American girls. She then asks for reactions and alternative suggestions. She thereby stimulates creativity and possibly an implicit understanding of such concepts as merit treatment, sharing benefits, sharing costs, having a minimum benefit threshold, having a maximum cost threshold, and other such ideas without using those words.

In the SOS context, this refers to creating a frame of mind that causes adults to do what is socially desired, because the alternative is virtually unthinkable. This

can be contrasted with a less effective emphasis on deterrence, whereby socially desired behavior is achieved through threats and bribes. Examples include childhood socialization to reduce adult behavior that is violent, alcoholic, drug addictive, and hostile toward constitutional rights.

Psychology of SOS Solutions

Innovative risk taking

This is an important SOS facilitator, because many SOS solutions involve technological fixes. In order to develop new technologies, many people usually had to risk substantial amounts of money, time, effort, and other resources. There may have been a strong possibility that it would have all been wasted. An SOS society needs more people who are willing to take such chances. Classic examples include Marie and Pierre Curie who sacrificed about 30 years of work plus their health to develop radium and thus radioactivity, which is part of the basis for nuclear energy. Thomas Edison frequently not only risked his resources but also his reputation by announcing inventions before he had developed them in order to give himself an ego risk as a stimulus to quickly inventing what he falsely said he had already done.

Sensitivity to opportunity costs

This means trying through socialization or an appropriate incentive structure to get decision makers to be more sensitive to the mistake of failing to try out a new idea that might work wonders, as contrasted to being so sensitive to sins of commission rather than omission. Both wrongs are undesirable. One can, however, say that a police officer who wrongly beats a suspect is doing less harm to society than a president who wrongly fails to adopt a new health care program that could save numerous lives or a new education program that could greatly improve productivity and the quality of life. A person who is sensitive to opportunity costs tends to say “nothing ventured, nothing gained,” whereas an insensitive person tends to say “nothing ventured, nothing lost.” We need more of the former in order to facilitate the generating, adopting, and implementing of SOS.

SOS combination of pessimism and optimism

This does not mean a balance or a compromise between being pessimistic and being optimistic. It means being 100% pessimistic or close to it regarding how bad things are and how much worse they are going to get unless we actively do something about

them, including developing SOS solutions. It simultaneously means being 100% optimistic or close to it regarding how good things can get in the future if we vigorously work at them, including developing SOS. This is in contrast to those who say the present is wonderful and needs little improvement. It is also in contrast to those who say the present may be wonderful or not so wonderful, but some invisible hands or automatic forces of Adam Smith, Karl Marx, or God will automatically improve the future.

Constantly seeking higher goals

This list of societal facilitators is in random order. Some of the items overlap or interact, but it is better to overlap than leave gaps in this context. It is appropriate perhaps to have the last facilitator relate to constantly seeking higher goals. Traditional goal seeking leads to compromises. Worse, it can lead to one side trying to win 100% and the other side losing 100%, but the war, strike, litigation, or other negative dispute resolution leads to both sides losing close to 100%. Obviously, seeking higher goals is more likely to result in higher goal achievement than seeking lower goals, including SOS goal achievement. The counterargument sometimes made is that higher goals lead to frustration because of the gap between goals and achievement. There may be more frustration in fully achieving low goals that provide a low quality of life when others are doing better. High societal goal seeking (including SOS) is facilitated by all of the above factors, but it is a factor in itself, because high goal seeking tends to become a self-fulfilling prophecy.^a

PART TWO: CREATIVITY FOR IMPROVING PUBLIC POLICY

Pushing, Facilitating, and Pulling Toward Innovative Public Policy

A useful way of organizing ideas for stimulating creativity is in terms of pushing factors, facilitators, and pulling factors. That three-part organization comes from Frederick Jackson Turner's analysis of the causes of people moving west in the 1800s. The pushing factors included undesirable aspects of the East, such as overcrowding, lack of jobs, and debts. Facilitating factors included wagon trails, railroads, river systems, and other means of transportation. Pulling factors

^aOn public policy in stimulating creativity, see Ref. ^[1]. Also see Refs. ^[2,3].

included attractions in the West, such as free land and business opportunities.

In this context, the pushing factors include other people and commitments. The facilitators include relevant literature, working style, and multicriteria decision making. The pulling factors include the rewards that go to successful imagination. The rewards here emphasize intellectual rewards, partly because the article is based on experience in academic and government activities, where monetary rewards are not as great as they are in business. The reader can adjust the ideas, however, to fit other contexts besides the academic and governmental contexts.

Pushing factors

Other People as Pushing Factors. Talk with someone else about generating alternatives. Trying to explain alternative ways of achieving something with an audience listening stimulates more ideas than talking or thinking to one's self. Put one's head together with someone else who is trying to come up with ideas. The interaction of two or more people trying to generate ideas tends to work better than one person alone. Have contact with stimulating colleagues via correspondence, conventions, informal campus relations, or other on-the-job relations. Work with graduate students and undergraduates to develop dissertations, seminar papers, and term papers.

Work with different people to provide a variety of interaction. Arrange to be asked questions by people with a variety of orientations, including sincere inquiry, skepticism, cynicism, and even a touch of malice. Try to operate in an interdisciplinary environment for a great variety of perspectives. Apply one's creative ideas to see what happens in practice.

Commitments as Pushing Factors. Accept a commitment to write an article, a book chapter, or a conference paper on how to deal with a policy problem. That is likely to generate new alternatives. Teach in those fields in which one wants to generate policy alternatives. Take on obligations to coauthor articles, chapters, or papers.

Take on obligations to do consulting work that involves generating alternatives. Prepare grant proposals. Arrange for competitive situations as a stimulus to developing new ideas.

Facilitators

Literature. Consult the literature in the field. There may be lots of alternatives already suggested. There are some software checklists that might be worth trying, such as "Trigger" published by Thoughtware

and the "Idea Generator" published by Experience in Software, Inc., 2039 Shattuck Avenue, Suite 401, Berkeley, CA 94701. Keep up with the newest ideas in various policy fields. Read provocative literature. Know the general literature in the fields in which one is interested.

Read some of the literature on creativity, including the list of references attached to this article. Have theoretical frameworks that can serve as checklists and prods for developing alternatives. Be familiar with the methods of knowing, including how to inductively generalize, how to deduce conclusions, how to determine what authorities hold, and how to do sensitivity analysis. Think about ways of generating ideas like this article or adding to this article.

Working Style. Talk out loud about the possible alternatives. Dictating is better than thinking in generating ideas. Delegate work to others in order to have more time to think. Have a pencil and paper handy at all times or dictating equipment to write or dictate ideas that come to one's mind before they are lost. Schedule time periods for creative development and idea implementation. The more time periods the better. Occasionally travel in order to provide a variety of environments.

Multicriteria Decision Making. Try listing some alternative, even if one only has in mind one or two alternatives to begin with. Merely trying to generate a list tends to result in more items being listed than one originally had in mind, or thought one had in mind. After generating some alternatives, then list some criteria for evaluating them. That will lead to more alternatives. After generating alternatives and criteria, then generate some relations between the alternatives and criteria. That will lead to more alternatives. After generating alternatives, criteria, relations and initial conclusions, then do various forms of sensitivity analysis designed to determine what it would take to bring a second-place or other-place alternative up to first place. That may generate still more alternatives.

If there is a situation in which there are two conflicting sides, each one favoring a different alternative, look to see what kind of alternative could maybe satisfy the goals of both sides. Also look to the possibility of a compromise alternative that will partially satisfy each side if it is not possible to find an alternative that will fully satisfy both sides. Then, observing how the alternatives score on the criteria, ask how each alternative can be improved. Try to convert the alternatives, criteria, relations, tentative conclusions, and sensitivity analysis into a publishable table with notes. That may generate new alternatives.

Pulling factors: rewards

Be motivated to want to generate alternatives. Arrange to be in situations where one is rewarded for generating alternatives, such as recognition, grants, publishing opportunities, graduate students, consulting opportunities, etc. Non-intellectual rewards can also be arranged for. These might include money, power, love, food, sleep, pure recreation, etc. Operate in a permissive environment that encourages experimentation and new ideas. The earlier one can get into such an environment the better, preferably starting at birth.

Some people use heredity as an excuse for not being creative. In both areas, there is a substantial range in which each person can operate. If one is more determined, then one can operate closer to the top (rather than the bottom) of one's inherited range. Creativity is probably less a matter of heredity than intelligence is. It is more susceptible to the kind of pushing, facilitating, and pulling factors mentioned above. Thus, one can more easily arrange to be a more creative person than one can arrange to be a brighter person by seeking more favorable occurrences of those factors. Doing so can be rewarding, as well as produce the kinds of rewards mentioned above. The broader rewards accrue not only to the individual, but also to the many potential beneficiaries of individual creativity. It is an ability well worth stimulating by society and by one's self.^b

Sources of Goals and Alternatives

Public-policy evaluation can be defined as the determination of which various governmental policies or decisions are best for achieving a given set of goals in light of the relations between the alternative policies and the goals and various constraints and conditions.

This definition emphasizes four key elements in public-policy evaluation:

1. A set of goals to be achieved within various normative constraints.
2. A set of alternative policies or combinations of policies that could be relevant to achieving the goals.
3. A set of relations between the policies and the goals.
4. The drawing of a conclusion from those goals, policies, and relations as to which policy or combination is best.

^bOn pushing, facilitating and pulling toward innovative public policy, see Ref. [4].

Where do these goals, policies, and relations come from? The answer includes four main possibilities:

1. Authority: one or more persons, books, articles, or other reliable sources of information regarding the relevant goals, policies, or relations.
2. Statistical or observational analysis: the analyzing of specific instances in order to generalize what the goals, policies, or relations might be.
3. Deduction: the drawing of a conclusion from premises that have been established from authority, observation, and intuition.
4. Sensitivity analysis: the guessing of the goals, policies, or relations and the determination of what effect, if any, the guessed values have on the final decision regarding which policy is best.

The four basic sources can be subclassified in various ways. For example, authority can be meaningfully discussed in terms of expert authority and general public opinion. Authority could also be contemporary or historical. Observation can be impressionistic or systematic, including statistical. Deductive approaches can be based on intuitively accepted premises or on empirically validated premises. Sensitivity analysis is threshold analysis in which we want to know the break-even point, above which we should take one course of action, and below which we should take another.

Authority

Consulting authorities, rather than establishing the goals, feasible policies, or relations in a policy evaluation with original data or reasoning, can be a big time saver. Traditional social science tends to downplay introspective information-gathering methods, in contrast to non-obtrusive methods. In policy evaluation, however, perhaps more consultation with insiders is needed in order to obtain more meaningful information about relationships than can be obtained from the limited and questionable data records that are available.

Who constitutes an authority on goals, policies, or relations? The answer depends on the subject matter. The Supreme Court is an authority, for example, on what goals are legitimate in satisfying the right-to-counsel clause of the Sixth Amendment to the Constitution. The Court has said that saving money is not an appropriate goal, but that saving innocent persons from being convicted is. If, however, the issue is not where right to counsel should be provided but rather how it should be provided, then saving money is an appropriate goal. For this issue, the goals of a county board would be relevant, because it generally appropriates money to pay court-appointed lawyers to

represent the poor. Such goals might include satisfying local lawyers while minimizing expenditures. The board might, therefore, decide on a salaried public defender system, rather than on a less expensive but less politically feasible assigned counsel system or a less legally feasible volunteer system. For other policy problems, the key authorities might be legislative opinion, public opinion, the head of an administrative agency, or the like.

Statistical observation

Statistical analysis is the most systematic form of inducing generalizations from many instances or observations. It is generally used for establishing relations, rather than for establishing goals or feasible policies. Statistical analysis can, however, be useful in establishing goals or weights for the goals, whenever the goals, rather than being ultimate, are instrumental for achieving higher objectives.

Accounting is a variation on statistical analysis. Like statistical analysis, it involves aggregating data, but accounting data is generally more precise than statistical analysis that is based on averages or the fitting of curves to scattered data points. A public opinion survey is not a variation on statistical analysis in the context of the typology of sources used in this article. Rather, it is a form of consulting authority in which the authority is the general public or a special segment of it. A statistical analysis (as a distinct source of information on goals, policies, or relations) involves a cross-tabulation, an analysis of the variation between averages, or a regression-equation analysis. These forms of statistical analysis involve determining a relation that is relevant to weighting goals, deciding which policies are feasible to choose among, or relating a policy to a goal.

Deduction

Deduction involves arriving at a conclusion from premises that have been established by way of authority, empirical validation, prior deduction, or intuition. The more acceptable the premises are, the more acceptable the conclusions should be, assuming the conclusions have been validly deduced from the premises. Deduction is especially helpful where there is no authority and no empirical data for determining the information desired.

It is important to note that deductively analyzing premises may lead to an alternative policy that might be missed if one only relies on authority or statistical analysis. Authority is often not very creative in foreseeing problems, and statistical analysis is incapable of dealing with policies that have never been adopted.

Sensitivity analysis

In policy evaluation, sensitivity analysis is a useful source of information about goals, policies, and relations when authority, statistics, and deduction do not provide clear answers regarding them. Sensitivity or threshold analysis enables one to determine how much room for error there is in weighting the goals, listing the policies, or measuring the relations. Often, the controversy over precision in these matters is wasted, because within the range in which the controversy occurs, the overall conclusion as to which policy or combination is best is still the same. Sensitivity analysis also enables the policy evaluator to convert difficult questions about goals, policies, and relations into relatively easy questions, such as, "Is a given weight, policy, or relation above or below some threshold?" rather than, "What is the exact weight, policy score, or relation?"

In using sensitivity analysis to determine a set of feasible policies, we have to distinguish between a method that will provide a set of policies from which we can choose, rather than a method designed to arrive at an optimum policy. All four sources of information can be used to arrive at feasible policies or an optimum policy.

Intuition

Intuition is closely related to sensitivity analysis as a source of goals, policies, and relations. Sensitivity analysis frequently involves determining how different guessed values affect the optimizing conclusions. Intuition is also a form of guessing or basing estimates on strong feelings. Goals are sometimes accepted intuitively rather than being justified in terms of authority, statistics, or higher premises. This is especially so if the goals are general or near-ultimate goals, rather than instrumental. Policies may often be suggested as a result of a flash of insight, which is the case with hypotheses in traditional social science research. Although it is not generally respectable in social or policy science to arrive at relations through intuition, one can repeatedly guess at a relation until the reasonable possibilities have been exhausted and then see how these guesses affect the optimizing conclusions. One may find that it is unnecessary to be any more scientific than that, because all the reasonable guessed values may yield the same conclusion as to which policy is best.

Ultimately, all goals and relations depend on intuition. Goals can be justified by appeal to authority, statistics, or deduction. However, how does one justify the authority, the dependent or goal variable in a statistical analysis, and the basic premises in a deductive analysis? One can likewise ask for a justification of

these justifications. In policy evaluation, one usually has an overall goal that is accepted intuitively, such as promoting the greatest happiness for the greatest number or satisfying the decision makers. Likewise, one can ask, why does policy *X* cause goal *Y*? The answer might be that there is a *Z* variable between *X* and *Y* which is caused or increased by *X*, and which in turn, causes or increases *Y*. One can then ask, why does *X* cause *Z*, and why does *Z* cause *Y*? At each stage of the causal regress, one tends to move further away from substantive policy and social science toward natural science and metaphysics. Ultimately, the question becomes, how do we know there is an *X* or a *Y*? That is, how do we know there is such a thing as a congressional statute or an American population that has social-indicator characteristics? In other words, on a philosophical level, we have to accept some empirical reality, such as the existence of the world. Fortunately for most policy evaluation, the goals in dispute are seldom ultimate goals, but rather instrumental goals that can be justified in terms of authority, statistics, or deduction. Similarly, the relations are seldom, if ever, metaphysical; rather, they can also be explained in a satisfactory, non-philosophical way in terms of authority, statistics, and deduction.

We can conclude from this analysis of the sources of goals, policies, and relations in policy evaluation that there are a variety of sources that can be systematically classified. We can also conclude that perhaps policy evaluation should be making more use of the variety of sources available. Unfortunately, certain disciplines tend to overlook some sources at the expense of others. Law and political science seem to rely heavily on authority as a source, especially legal authority. Psychology and sociology may rely too heavily on statistical analysis, which tends to overemphasize variables that are easily measurable and policies that need to be adopted before they can be evaluated. Economics and engineering often rely too heavily on deduction, especially mathematical modeling, which sometimes involves unrealistic or incomplete premises. By working with a combination of authority, statistics, and deduction, one provides a form of triangulation that increases the likelihood of arriving at more meaningful goals—weights, policies, and relations.

There is no need to argue over which source between authority, statistics, and deduction is the most desirable. Authority is clearly a big time saver if an accessible and respected authority is involved. Deduction enables one to draw conclusions about goals, policies, and relations without having to gather original data, but instead by synthesizing already known information. Statistical analysis constitutes a more ultimate, but more difficult, form of proof. In any concrete policy evaluation situation, the best source depends on the subject matter and what is to be done with it.

If the policy evaluation involves constitutional policy, an appeal to Supreme Court authority may be most relevant. If it involves the effects of a strike in the coal industry on another segment of the economy, a deductive input–output model may be the preferable type of analysis. If it concerns the trade-off problem of inflation and unemployment, a time-series statistical analysis may be especially appropriate in relating inflation and unemployment to suicide rates, to the percentage of the two-party vote that goes to the incumbent party, or other social indicators.

We can also conclude that sensitivity of threshold analysis is a useful tool in policy evaluation, because even with authority, statistics, and deduction, it may still not be possible to arrive at precision in weighting goals, measuring policies, or determining relations. Sensitivity analysis enables one to determine whether increased precision is needed. It is only needed if the range of unclearness on a goal–weight, a policy, or a relation happens to encompass a threshold value. Thus, if the range of unclearness on a goal–weight or a relation is between 20 and 30, but the threshold value of the goal–weight or the relation is 10, then one can forget about clarifying the unclearness if one is mainly concerned with determining which policy is best. If, however, the threshold value is 26, then one should seek additional information from authority, statistics, and deduction to determine whether the actual value is above or below 26.

The purpose of this article has been to discuss the sources of goals, policies, and relations in policy evaluation. The article represents a synthesis of reasonable common sense, at least as a matter of hindsight. That is what good policy evaluation should be, namely, codified common sense. For thousands of years, many human beings have been making effective and efficient decisions. What decision science and policy science should now try to do is to capture the essence of what these good decision makers have done implicitly. Less naturally competent decision makers can then improve their decision-making or policy-evaluating skills.^c

A Typology Toward Win–Win Public Policy

There are about 14 different ways of arriving at win–win superoptimum solutions, whereby conservatives, liberals, and major viewpoints can all come out ahead of their best initial expectations simultaneously. The list could be used as a checklist to prod one's mind into thinking of solutions to specific problems.

^cOn sources of goals and alternatives toward innovative public policy, see Ref. [5].

More resources to satisfy all sides

Expanding Resources. An example might include well-placed subsidies and tax breaks that would increase national productivity and thus increase the gross national product and income. Doing so would enable the tax revenue to the government to increase, even if the tax rate decreases. This would provide for a lowering of taxes, instead of trying to choose between the liberal and conservative ways of raising them. It would also provide for increasing domestic and defense expenditures, instead of having to choose between the two.

Third-Party Benefactor. Some situations involve a third-party benefactor that is usually a government agency. An example is government food stamps, which allow the poor to obtain food at low prices, while farmers receive high prices when they submit the food stamps they have for reimbursement. Another example is rent supplements, which allow the poor to pay low rents, but landlords receive even higher rents than they would otherwise expect.

More efficiency in achieving goals

Setting Higher Goals. An example of setting higher goals than what was previously considered the best while still preserving realism might include the Hong Kong labor shortage with unemployment at only 1%. Hong Kong is faced with the seeming dilemma of having to choose between foregoing profits (by not being able to fill orders due to lack of labor) and opening the floodgates to mainland Chinese and Vietnamese (in order to obtain more labor). An SOS might involve adding to the labor force by way of the elderly, the disabled, and mothers of preschool children. It also would provide more and better jobs for those who are seasonally employed, part-time employed, full-time employed but looking for a second job, and full-time employed but not working up to their productive capacity.

Decreasing Causes of Conflict. An example of removing or decreasing the source of the conflict between liberals and conservatives, rather than trying to synthesize their separate proposals, would be concentrating on having a highly effective and acceptable birth control program to satisfy proponents and opponents of abortion, because abortions would then seldom be needed. Another example would be concentrating on a highly effective murder-reduction program to satisfy proponents and opponents of capital punishment. Such a murder-reduction program might emphasize gun control, drug medicalization, and reduction of violence socialization.

Redefining the Problem. Quite often, a highly emotional controversy between liberals and conservatives may be capable of being resolved beyond the best expectations of each side through the approach of redefining the problem. They may be arguing over how to deal with a problem that is relatively unimportant in terms of achieving their goals, as contrasted to a more important problem on which they might be likely to get mutually satisfying agreement. This involves seeing beyond a relatively superficial argument to the higher-level goals that are endorsed by liberals and conservatives, although possibly not to the same relative degree.

Increasing Benefits and Decreasing Costs. There are situations where one side can receive big benefits but the other side incurs only small costs. An example is in litigation, where the defendant gives products that it makes. The products may have high market value to the plaintiff, but low variable or incremental cost to the defendant, because the defendant has already manufactured the products or can quickly do so.

Early Socialization. The socialization matter could be discussed across every field of public policy. If one is going to have a superoptimum society, then it is important what kinds of attitudes children have with regard to discrimination, poverty, world peace, crime, education, consumer-merchant relations, labor-management relations, free speech, and fair procedure. One could even say that the key purpose, or a key purpose of public policy, is to provide for a socialization environment in which children have socially desired attitudes on every field of public policy. If that is done properly, then a good deal of the problems of what policies to adopt will take care of themselves, because the need for public policy will be lessened. If children, for instance, are imbued with more of the idea of judging each other in terms of their individual characteristics rather than in terms of ethnic characteristics, then we have less need for public policies dealing with racism, because there is likely to be a lot less racism.

Technological Fix. The second level of insight is to communicate a recognition that such SOS are realistically possible and not just conceptually possible. A good example relates to the ozone problem and the use of fluorocarbons in hair sprays and other aerosol containers. As of about 1985, such devices represented a serious threat to depleting the ozone layer, thereby causing a substantial increase in skin cancer throughout the world. The solution was not to rely on an unregulated marketplace, which normally provides almost no incentives to manufacturers to reduce their pollution. The solution was not regulation or prohibition, which tends to be evaded, is expensive to enforce, and

is enforced with little enthusiasm given disruptions that might occur to the economy. The most exciting aspect of the solution (although the problem is not completely solved) was the development of new forms of spray propellant that are less expensive for manufacturers to use and simultaneously not harmful to the ozone layer.

This kind of solution tends to be self-adopting, because manufacturing firms, farmers, and others who might otherwise be polluting the environment now have an important economic incentive to adopt the new low-polluting methods, because they reduce the expenses of the business firm. This approach requires substantial research and substantial government subsidies for research and development as contrasted to paying the polluters not to pollute, which is even more expensive and often not so effective, because they may take the money and pollute anyhow. The business firms generally do not have capital for that kind of research and development, or the foresight or forbearance which public policy and governmental decision making may be more capable of exercising. This includes international governmental decision makers, as well as those in developing nations.

Contracting Out. As for how the SOS operates, it involves government ownership, but all the factories and farms are rented to private entrepreneurs to develop productive and profitable manufacturing and farming. Each lease is renewable every year, or longer if necessary to get productive tenants. A renewal can be refused if the factory or farm is not being productively developed, or if the entrepreneur is not showing adequate sensitivity to workers, the environment, and consumers.

As for some of the advantages of such an SOS system, it is easier not to renew a lease than it is to issue injunctions, fines, jail sentences, or other negative sanctions. It is also much less expensive than subsidies. The money received for rent can be an important source of tax revenue for the government to provide productive subsidies elsewhere in the economy. Those subsidies can be used especially for encouraging technological innovation–diffusion, the upgrading of skills, and stimulating competition for market share, which can be so much more beneficial to society than socialistic or capitalistic monopolies. The government can more easily demand sensitivity to workers, the environment, and consumers from its renters of factories and farms than it can from itself. There is a conflict of interest in regulating oneself.

International Economic Communities. An exciting new development with regard to international interaction to deal with shared policy problems is the international economic community (EC). It involves

a group of countries agreeing to remove tariff barriers to the buying and selling of goods among the countries as a minimum agreement to constitute an EC. The agreement may also provide for removal of immigration barriers to the free flow of labor, and removal of whatever barriers might exist to the free flow of communication and ideas. The European Economic Community is a good example, but other examples are developing in North America, Africa, Asia, and East Europe.

The alternative of having an economic community does well on the conservative goal of preserving national identity, because no sovereignty is lost in an IEC, as contrasted to the sovereignty that is lost in a world government or a regional government. The IEC may also add to the national stature of the component parts by giving them the increased strength that comes from being part of an important group. Thus, France may have more national stature as a leader in the European Economic Community than it has alone.

Likewise, the alternative of having an economic community does well on the liberal goal of promoting quality of life in terms of jobs and consumer goods. Jobs are facilitated by the increased exporting that the IEC countries are able to do. Jobs may also be facilitated by free movement to countries in the EC that have a need for additional labor. Consumer goods are facilitated by the increased importing that the EC countries are able to do without expensive tariffs.

More combinations of alternatives

Big Benefits on One Side, Small Costs on the Other. An example of this kind of SOS is the case of growers versus farmworkers in Illinois. The essence of the solution is that the growers agree to deposit \$100,000 to begin an employee credit union. Depositing \$100,000 costs nothing to the growers, because it is insured by the federal government and can be withdrawn after an agreed-upon time period, possibly even with interest. The \$100,000, however, serves as the basis for the beginning of an economic development fund that enables the workers through real estate leveraging to obtain a mortgage for building over \$500,000 worth of housing as a big improvement over their current housing. The existence of the credit union also enables them to avoid having to get advances from the growers, which generates a lot of friction as a result of alleged favoritism in giving and collecting the advances. There are other elements involved, too, such as new grievance procedures and reports regarding compliance with other rules governing the working conditions of migratory labor. The essence of the solution, though, is that both sides come out ahead of their original best expectations.

Combining Alternatives. An example of combining alternatives that are not mutually exclusive is combining government-salaried legal-aid attorneys with volunteer attorneys. Doing so could give the best of public-sector and private-sector approaches to legal services for the poor. Another example is combining tax-supported higher education plus democratic admission standards with contributions from alumni and tuition plus merit standards. Doing so results in universities that are better than pure government ownership or pure private enterprise.

Developing Multifaceted Packages. One can develop a package of alternatives that would satisfy liberal and conservative goals. An example is pretrial release, where liberals want more arrested defendants released prior to trial, and conservatives want a higher rate of appearances in court without having committed a crime while released. The package that increases the achievement of both goals includes better screening, reporting in, notification, and prosecution of no-shows, as well as reduction of delay between arrest and trial.

Sequential SOS. We can put the land reform example in with sequential SOS. The current verbalization does not say anything about encouraging the landless peasants to subsequently upgrade their skills to be able to take on non-agricultural work or to upgrade the skills of their children. We could change the SOS definition to say simultaneously or sequentially. One drawback is that there is subjectivity and favoritism as to which alternative goes first. Simultaneity has an air of equality and equity; doing it sequentially may be essential in terms of developing feasibility. It is not so feasible to do various alternatives or goals simultaneously.^d

CONCLUSIONS

Public policies that tend to stimulate creativity are those that relate to the following:

1. Competitive political parties.
2. Better policy analysis methods and institutions.
3. Competitive business firms.
4. Well-targeted subsidies and tax breaks.
5. Increased national productivity.
6. Childhood socialization.
7. Innovative risk-taking.
8. Sensitivity to opportunity costs.
9. Combination of pessimism and optimism.
10. Seeking higher goals.

Some of these stimulants are part of the culture, not just official or unofficial public policy.

Factors that stimulate innovative improvements in public policy include:

1. Pushing factors, including other people and commitments.
2. Facilitators, including literature, working style, and multicriteria decision making.
3. Pulling factors or rewards.
4. Consulting authorities as a source of policy goals and alternatives.
5. Statistical observation.
6. Deduction.
7. Sensitivity analysis or experimenting.
8. Intuition.
9. More resources to satisfy all sides in win-win policy.
10. More efficiency for achieving goals in win-win policy.
11. More combinations of alternatives in win-win policy.

It is hoped that this article will help build some bridges between people interested in creativity innovation and people interested in improving public policy.^e

REFERENCES

1. Nagel, S., Ed. Win-win societal facilitators. In *Super-Optimum Solutions and Win-Win Policy: Basic Concepts and Principles*; Greenwood-Quorum, 1997.
2. Nagel, S., Ed. The economic process of SOS solutions. In *The Policy Process and Super-Optimum Solutions*; Nova Science, 1994.
3. Nagel, S., Ed. The psychology-sociology of SOS solutions. In *The Policy Process and Super-Optimum Solutions*; Nova Science, 1994.
4. Nagel, S., Ed. Multiple alternatives and criteria with discrete choices. In *Evaluation Analysis with Microcomputers*; JAI Press, 1989.
5. Nagel, S., Ed. Sources of goals, policies, and relations. In *Public Policy: Goals, Means, and Methods*; St. Martins, 1984.
6. Nagel, S. *Creativity and Public Policy: Generating Super-Optimum Solutions*; Ashgate, 1999.
7. Nagel, S., Ed. *Creativity: Being Usefully Innovative in Solving Diverse Problems*; Nova Science Publishers, 2000.

^eOn creativity and public policy in general, see the quarterly journal called *Creativity Plus*. It is a journal of the Policy Studies Organization and the Creativity Plus Association. Both are headquartered at the Everett-Dirksen-Adlai Stevenson Institute for International Policy Studies, 711 Ashton Lane South, Champaign, Illinois 61820. Also see Ref. [7].

^dOn a typology toward win-win public policy, see Ref. [6].

BIBLIOGRAPHY

1. Austin, J. *Chase Chance and Creativity: The Lucky Art of Novelty*; Columbia University Press, 1978.
2. Baker, S. *Your Key to Creative Thinking: How to Get More and Better Ideas*; Harper & Row, 1962.
3. Brookfield, S.D. *Developing Critical Thinkers: Challenging Adults to Explore Alternative Ways of Thinking and Acting*; Jossey-Bass, 1991.
4. Buffington, P.W. Strokes of genius. *Sky* February **1987**, 121–125.
5. Campbell, D. *Take the Road to Creativity and Get Off Your Dead End*; Center for Creative Leadership, 1985.
6. Coogan, W.H.; Woshinsky, O.H. *The Science of Politics: An Introduction to Hypothesis Formation and Testing*; University Press of America, 1982.
7. De Bono, E. *Lateral Thinking: Creativity Step by Step*; Harper Colophon Books, 1973.
8. De Bono, E. Thinking in America: the lost art. *Crit. Intell.* October **1994**, 3–9.
9. Dogan, M.; Pahre, R. *Creative Marginality: Innovation at the Intersections of Social Sciences*; Westview Press, 1990.
10. Harman, W.; Rheingold, H. *Higher Creativity: Liberating the Unconscious for Breakthrough Insights*; G.P. Putman's Sons, 1984.
11. Harriman, R. Creativity: moving beyond linear logic. *Futurist*, August 1984, 17–20.
12. Nierenberg, G.I. *The Art of Creative Thinking*; Barnes & Noble, 1996.
13. Nuernberger, P. Mastering the creative process. *Futurist*, August **1984**, 33–36.
14. Osborn, F. *Applied Imagination: Principles and Procedures of Creative Problem-Solving*; Charles Scribner's Sons, 1963.
15. Parnes, S.J. Learning creative behavior: making the future happen. *Futurist*, August **1984**, 30–32.
16. Quester, G.H. Creativity and bureaucracy: the search for success. *Futurist*, August **1984**, 27–29.
17. Rosenfeld, R.; Servo, J.C. Business and creativity: making ideas connect. *Futurist*, August **1984**, 21–26.
18. Rothenberg, A. *The Emerging Goddess: The Creative Process in Art, Science, and Other Fields*; University of Chicago Press, 1990.
19. Stein, M.I.; Heinze, S.J. *Creativity and the Individual: Summaries of Selected Literature in Psychology and Psychiatry*; Free Press of Glencoe, 1960.
20. Tatsuno, S.M. Japan's move toward creativity. *GAO J.* Summer **1990**, 13–18.
21. Technology is changing basic structure of education. *News Rep.* Summer 1993, 2–5.
22. Waitley, D.E.; Tucker, R.B. How to think like an Innovator. *Futurist*, May–June **1987**, 9–15.
23. Weinstein, B. *20 Ways to be More Creative in Your Job*; Simon & Schuster, 1983.

Integrated Healthcare Systems: New Trends, Emerging Models, and Future Shocks

Yvonne J. Kochanowski

SteelEdge Business Consulting, Placerville, California, U.S.A.

INTRODUCTION

Integrating the continuum of healthcare providers both vertically across different levels of care and horizontally within a level has been an ongoing trend during the last two decades. The degrees of success in these ventures, however, have varied greatly. Some emerging models are incorporating more than traditional health organizations and are embracing new methods of governance, structure, and financing to reinvent the integrated holistic healthcare system of the future. This e-paper provides an overview of these new trends, emerging models, and future shocks.

HISTORICAL AND THEORETICAL BACKGROUND

To understand why healthcare organizations are evolving into integrated systems, it is important to study the unique development of healthcare in the United States. The history of this development can be identified in three specific categories:

- Individualism and autonomy transitioning to formal medical study: When the country was being settled, experienced lay healers were called upon to use herbs and folk medicines to cure ills and ease pain. As towns developed, fledgling physicians would apprentice with a more experienced doctor and would then move into a solo practice. Knowledge was shared by way of professional journals. The advent of medical schools after the War of 1812, credentialing and standards of practice with a medical license in the mid-1800s, and consultations between physicians in the mid- to late 1800s contributed to the reorganization of medicine.
- Facilities transitioning from dying wards to places of healing: Early facilities could do little to heal those who were unfortunate enough to enter their doors. Until capital was invested by the wealthy or by the physicians, hospitals did not have the resources to perform healing. The move from social welfare to medical science began when the profession of nursing was formalized and when antiseptic surgery became available well after the

Civil War. By 1900, hospitals had changed from voluntary structures to religious and ethnic-sponsored organizations and eventually evolved into a for-profit business run by physicians and corporations.

- Payors transitioning from social programs to risk insurance: Originally intended to guarantee wage replacement, to cover medical fees, and to pay the indirect costs of illness to society, insurance became a right to health benefits and shifted the focus to expanding the access to medical care in the early 1930s. By the 1950s, health insurance became a standard employment benefit, extended to the family as well as to the wage earner. Sophisticated systems of cost containment and utilization management provided the final shift in the past two decades.

According to Starr,^[1] five trends shifted the individualized pieces of the health system toward the integrated structures existing today:

- Changes in type of ownership and control from non-profit and government organizations to for-profit companies.
- Horizontal integration of similar providers into multi-institution systems.
- Diversification and corporate restructuring to corporations that strongly influence a geographic market.
- Vertical integration across the continuum of care at all levels of services.
- Industry concentration covering multiple regional and national geographies.

Four recent significant trends can be added to this list:

- Shifts from facility-based to outpatient, ambulatory, and home-care services, reducing the hospital power base.
- Changes in staffing patterns of care providers, reducing the impact of specialist physicians and increasing the emphasis on alternative staffing models.
- Increased pressures to reduce insurance premium costs, placing greater burdens on insurance

companies to manage limited financial resources without being accused of withholding quality patient care.

- Recognition that social services play an integral part of a successful system, providing the infrastructure of supports that meets the functional needs of those receiving care.

It is within this new framework that the need for collaboration and alliances between physicians, providers, payors, and social services organizations to serve the individual has arisen. This framework requires integrating all levels of patient care and stakeholders interested in the efficiency and the effectiveness of American healthcare.

INTEGRATED SYSTEMS DEFINED

Integrated healthcare delivery models are patterned on horizontal and vertical combinations of different levels of the continuum of care, types of providers, categories of payors, and needs of patients. This organizational realignment is traditionally performed to dominate the market—provide services to a greater portion of the patient population in a given geographic area. Theoretically, this realignment reduces the costs of services, improves service delivery for patients, and increases satisfaction of the various clients of the system.

A fully integrated delivery system is most typically defined as a system that ties the financing for services to the provision of all services, with all the parties sharing risk. Care is provided at the most appropriate level of care—clinically and from a cost-effective perspective. Accountability for community health status—wellness as well as treatment of illness—is also a cornerstone. Conrad and Dowling^[2] expanded this definition:

A vertically integrated health care system is an arrangement whereby a health care organization (or closely related group of organizations) offers, either directly or through others, a broad range of patient care and support services operated in a functionally unified manner . . . Full functional integration requires both administrative and clinical integration.

While this definition identifies the clinical continuum and support services that are necessary to the design of an integrated system, little is mentioned about the cost-effective provision of care. In responding to economic realities, the financial component must be built into the integration paradigm.

Coddington et al.^[3] define an integrated system as:

... a system that combines physicians, hospitals, and other medical services, along with a health plan

(or the ability of the system to enter into risk contracts), in order to provide the complete spectrum of medical care for its customers. In a fully integrated system, the three key elements—physicians, hospital(s), and health plan(s)—are kept in balance by common management and financial incentives so they can match medical resources with the needs of payers and patients.

The key characteristics of this definition include providers and payors, shared risk for the success of the venture, and cost effectiveness and information management systems which can maintain the efficacy of the system overall.

Shortell et al.^[4] add the following key characteristics to the definition:

We define an organized delivery system as a network of organizations that provides or arranges to provide a coordinated continuum of service to a defined population and is willing to be held clinically and fiscally accountable for the outcomes and the health status of the population served.

These researchers further point out that common ownership is not mandatory, and that a variety of strategic alliances and contractual arrangements is appropriate. Expanding the organization further, this structure takes the perspective from the individual to the organization and from the organization to the industry overall. What emerging trend and models can we expect based on this theoretical background?

EMERGING TRENDS AND MODELS

In a 1994 study, 71% of hospital and health network leaders said that their organizations are trending toward integrated delivery systems, with 81% saying that they would not exist as stand-alone institutions within 5 years.^[5] Four indicators—strategic direction, economic processes, organizational culture, and unique clinical capabilities—are normally used to gauge an organization's readiness for integration. With this framework in mind, more recent trends have emerged in the integrated healthcare marketplace. A sample of three—legal structure issues, incentive alignment, and organizational evolution—is discussed here.

Legal Structure

With the formation of a new or reorganized legal relationship, single ownership gives way to shared ownership management contracts, external party contracts, affiliation agreements, informal or formal memorandums of understanding, and formal acquisitions

and mergers. However, many individual providers, poised on the brink of integration with an established or fledgling system, face difficult issues in determining and then gaining approval for their legal structure. These approvals are not only internal to the organizations (governance, decision-making authority, responsibility for services) but also external (antitrust, market power, tax status).

According to Teevans,^[6] "... Antitrust laws seek to protect health care consumers from persons or entities that seek to, among other things, increase prices of, lower the quality of, or limit access to health care services." In both urban and rural settings, health care organizations are coping with the realities of the external controls by demonstrating that the creation or expansion of an integrated network not only improves access, but also allows the organizations to continue to operate, thereby providing vital services to a community. For example, if individual physicians were each to recreate information systems, contract negotiations, and risk coverage, they would have little time left for patient care. Furthermore, many may be forced to leave a market simply based on the economics of investing in all of these support systems.

Incentives Alignment

Clinical, financial, and patient satisfaction incentives are most commonly used to measure the improvement in health status of the community. Clinical integration occurs through the development of critical paths, care tracks, and collaborative case management plans, resulting in an interactive patient care process. Financial incentive alignments come about through capitation, contracting, and more closely managed cost and reimbursement. Improved patient satisfaction typically results from better communications, data sharing, and consumer inclusion in decisions about their care and services.

It is difficult, however, to develop these alignments quickly. Some, such as the integrated care tracks across different levels of care, can take months of education and negotiation between providers who had previously worked independently. In the financial incentives arena, Linenkugel^[7] asks, "How do providers keep going financially when the incentives are misaligned—when a lack of rewards for keeping people healthy is coupled with diminishing reimbursement from payers (government and commercial insurances) to care for persons who don't keep themselves out of healthcare trouble?" Finally, how can organizations consistently engage consumers in developing and implementing processes and systems that improve patient satisfaction?

Many organizations are addressing this through systems that:

- Include non-medical determinants of health, considered within a framework of health behaviors, living and working conditions, personal resources of patients/consumers, and environmental factors.
- Place greater emphasis on population health status, with indicators such as mortality and morbidity, balanced scorecards of changes in population health status, and self-reports of patient changes.
- Examine health status throughout the life cycle, considering both ends of the spectrum and all points in between, from babies to the senior population.
- Assess the coordination at all levels, both between levels of care and within care levels.^[8]

Organizational Evolution

While it is generally recognized that the theories behind the organization development, evolution, and utilization of system tools such as information systems have generally lagged in the health care practice field behind other industries, health systems have, in more recent years, begun to catch up. Barber et al.^[9] have applied a standard organizational evolution model to the healthcare field with the following findings:

The evolving IHS (integrated health system) will need to address the following infrastructure issues: a systemwide mission statement; a corporate culture that welcomes and integrates previously independent entities; an information system that will clinically and financially integrate the system; a clinical support system that emphasizes clinical efficiency and patient-focused care; and an actuarial approach to healthcare planning. The IHS's human resource challenges include recruiting and retaining administrators to initiate and manage complex arrangements among acute, subacute, long-term, and ambulatory care services; requiring governance to be proactive, structured, and focused on the resources and services of the entire system versus individual entities; and preparing physicians for the transition from independent fee-for-service ambulatory and hospital practices to integrated, clinically efficient managed care.

These researchers further state that external factors will influence and alter the life cycle of the organization. The keys are to continue to reexamine and reinvent the integrated system to better meet the needs of its community and to plan for contingencies that may arise from the external factors.

FUTURE EXPECTATIONS

How then can a forward-thinking healthcare organization continue to reinvent itself and remain a vibrant, contributing member of its community? According to some thinking, integrated healthcare systems are failing at an alarming rate, at a time when other industries that have long attempted integration (such as manufacturing, telecommunications, and services) are refocusing on core competencies.

According to Friedman and Goes,^[10] the reasons for failure are both structural and functional:

Structural problems:

- Information system conflicts.
- Focus on hospital/physician/payer rather than on the patient.
- Unwillingness of network partners to surrender autonomy.
- Inefficient supply chain processes.
- Supremacy of financial performance measures.
- Difficulty in determining and measuring meaningful outcomes.

Functional and process problems:

- Misalignment of cultures and incentives.
- Problems associated with building trust between and among key stakeholders.
- Inadequate time and attention paid to managing employee responses to change.
- Problematic leadership.
- Uncertain vision of the desired outcome.
- Lack of overall organizational commitment and understanding.
- Poor or inadequate communication.

Focusing on community as well as organizational needs is critical to integration effectiveness. If the community needs a large network of primary care providers and the organization is attracting specialists,

it will not be competitive. Similarly, reengineering the cost of healthcare delivery at the expense of community programs will disconnect the healthcare network from its greater purpose. Improved communications and requisite data collection will similarly meet community needs. As community report cards become more widely used, providing and maintaining measurable and superior healthcare quality which responds to community needs will become the true measure of integrated network strength.

REFERENCES

1. Starr, P. *The Social Transformation of American Medicine*; Basic Books, 1982; 429 pp.
2. Conrad, D.A.; Dowling, W.L. Vertical integration in health services: theory and managerial implications. *Health Care Manage. Rev.* **1990**, *15* (4), 9–22.
3. Coddington, D.C.; Moore, K.D.; Fischer, E.A. In pursuit of integration. *Healthc. Forum J.* **1994**, *37* (2), 53–59.
4. Shortell, S.M.; Gillies, R.R.; Anderson, D.A. The new world of managed care: creating organized delivery systems. *Health Aff.* **1994**, *13* (5), 46–64.
5. Bartling, A.C. Integrated delivery systems: fact or fiction. *Healthc. Exec.* **1995**, *10* (3), 6–11.
6. Teevans, J. Applying legal analysis to integrated rural networks. *Calif. State Rural Health Assoc. E-Newsl.* Jan–Feb **2001**, 1–14.
7. Linenkugel, N. Integrated health networks are not created equal. *Front. Health Serv. Manag.* **2001**, *17* (4), 41–44.
8. Green, C.J.; Moehr, J.R. Performance evaluation frameworks for vertically integrated health care systems: shifting paradigms in Canada. *J. Am. Med. Inform. Assoc.* **2000**, *7*, 315–319. Supplement.
9. Barber, J.B.; Koch, K.E.; Parente, D.; Mark, J.; Davis, K.M. Evolution of an integrated health system: a life cycle framework. *J. Healthc. Manag.* **1998**, *43* (4), 350–377.
10. Friedman, L.; Goes, J. Why integrated health networks have failed. *Front. Health Serv. Manag.* **2001**, *17* (4), 3–28.

Internal Controls

Robert W. Smith

Department of Political Science, Clemson University, Clemson, South Carolina, U.S.A.

INTRODUCTION

Internal controls are typically treated as a narrow specialization within or under broader fields such as organizational theory, management, accounting, auditing, or budgeting and financial management. However, internal controls are better understood as a thread running through each of these fields rather than as a separate concentration. Organizations as diverse as the U.S. General Accountability Office (GAO),^[1] the New York State Office of the State Comptroller (1999), and the University of California^[2] define internal controls as “An integral component of an organization’s management that provides reasonable assurance that the following objectives are being achieved: effectiveness and efficiency of operations, reliability of financial reporting, and compliance with applicable laws and regulations.” (See p. 3 of Ref.^[1].) Hence, internal controls constitute both the tools and processes that help public managers achieve effectiveness and efficiency while avoiding serious problems such as overspending, operational failures, and violations of law.

EVOLUTION OF INTERNAL CONTROLS

Internal controls have been around throughout the history of large organizations. In the private sector, the original purpose of internal controls was to protect against misuse of assets, overspending, or even theft by employees.^[3] As commerce flourished, internal controls were viewed as an effective means of corporate command and control. In their 21st century application, internal controls are as much a part of the public and nonprofit sectors as they are of the private sector. In the United States, as systems of commerce evolved, so too did the federal government’s regulatory function and oversight responsibility for commerce and other societal activity. Concomitantly, the level of effort as well as the intricacies of internal control systems also advanced. As large public and private organizations were created in the 20th century, concerns over how best to manage, ensure accountability, and achieve efficiency in expanding and complex agencies and companies were paramount. To understand

why internal controls are an established feature of contemporary public administration, it is necessary to briefly examine the evolution of internal controls in federal, state, and local government as well as the private sector.

The first real advance in modern governmental financial management came with the creation of the Budget and Accounting Act of 1920 (BAA). This act created the National Budget, the Bureau of the Budget, and the General Accounting Office.^[4] As of July 7, 2004, the GAO’s legal name became the Government Accountability Office. Through the BAA, the federal government established that internal controls were necessary to combat misuse, waste, and fraud. “In 1929, the Federal Reserve Bulletin (first issued in 1917 to provide detailed instructions for preparing balance sheets) stressed reliance on internal controls and the use of testing, instead of detailed verification, when such internal control was deemed reliable.”^[3, p. 35] The American Institute for Certified Public Accountants (AICPA) defined internal control “as those measures and methods adopted within the organization to safeguard the cash and other assets of the company as well as to check the clerical accuracy of the bookkeeping” (*Examination of Financial Statements by Public Accountants*, 1936). The AICPA later revised its definition of internal controls in 1949 to include “the plan of organization and all of the coordinated methods and measures adopted within business to safeguard its assets, check the accuracy and reliability of its accounting data, promote operational efficiency, and encourage adherence to prescribed management policies.”^[3]

The federal government continued on its path of crafting legislation and regulations that mandated the implementation of internal controls with the passage of the Budget and Accounting Procedures Act of 1950. This act assigned responsibility for establishing and maintaining adequate systems of internal control to the heads of federal agencies. The act also required that: 1) agencies’ accounting systems and systems of internal control conform to standards set by the Comptroller General; and 2) all such systems be approved by him.^[4] The 1950 act had little success. Elmer B. Staats, former U.S. Comptroller General (1966–1981), explained the reasons for the failure of the act:

The reason internal control systems are in a state of disrepair is that top management has devoted most of its concern and emphasis to delivering funds and services and that effective controls over tasks and functions which lead to the delivery of these funds and services had a low priority. Because of top management's insufficient concern for internal controls, middle management reflects this same indifference.^[4]

Concerned with financial and accounting turmoil in the 1970s, the federal government once again tried to step in to establish some uniformity of internal controls within the various federal agencies. The Inspector General Act of 1978 created the post of Inspector General in 12 federal agencies that combined the auditing and the investigation forces under one authority, which would report to the agency head as well as to Congress.^[4] Later, in 1982, the Federal Managers' Financial Integrity Act provided a strong stimulus for government managers to insist upon sound management systems, including both internal administrative and internal accounting controls.^[4] This Act also mandated that the GAO issue standards for internal controls in government.^[1]

Although the federal government had taken the more visible lead in the internal control movement up to the mid-1980s, the private sector continued to utilize internal controls, but in a more decentralized fashion. This all changed with the creation of the Committee of Sponsoring Organizations (COSO) of the National Commission on Fraudulent Financial Reporting, commonly known as the Treadway Commission, a private-sector initiative that was launched to understand the factors leading to the rising incidences of fraudulent financial reporting in the 1970s and 1980s. The committee issued *Internal Control—Integrated Framework*, which recognized the seriousness of the issue, of not having a common agreement on how to define internal control and not having a set standard that entities could consult to determine whether their own internal control system was effective. "The report presented a common definition of internal control to meet the needs of diverse users and provides a framework against which entities can assess and improve their internal control systems."^[5]

The federal government embraced this definition and related developments. Additional federal legislation has assisted in the continued evolution of internal controls as well as an increased awareness and compliance on the part of government managers. The Chief Financial Officers Act of 1990 called for financial management systems to comply with internal control standards, the Government Performance and Results Act of 1993 required agencies to clarify their missions, set strategic and annual performance goals, and measure and report on performance toward those goals,

and the Federal Financial Management Improvement Act of 1996 identified internal control as an integral part of improving financial management systems.^[1]

"Since the passage of the Single Audit Act (SAA) of 1984; the 1988 revision of the GAO *Standards for Audit of Governmental Organizations, Programs, Activities, and Functions* (The 'yellow book'); and the adoption of Statements on Auditing Standards (SAS) Nos. 53, 54, 55, 60, and 63, a greater emphasis has been placed on the quality of internal controls and the related reporting process."^[6] These mandates have the single purpose of assuring the public that government resources are being used efficiently, economically, and effectively, and are free from waste, abuse, or fraud.

The Single Audit Act of 1984 (SAA) and the Single Audit Act Amendments of 1996 required each audit of a governmental body to encompass the entire financial operation of the entity. As expanded in OMB Circular A-128, auditors must conduct a study and evaluation of the internal control systems used in administering federal assistance programs. The review and evaluation require auditors to test whether the internal control systems are functioning in accordance with prescribed procedures. The auditor's report on the internal control systems must identify the organization's significant internal accounting controls and those controls designed to provide reasonable assurance that federal programs are being properly managed. The report must also identify the controls that were not evaluated and the material internal control weaknesses identified as a result of the evaluation (OMB Circular A-128). Statement of Auditing Standard 55 and OMB Circular A-128 also require a compliance report, which identifies any noncompliance with laws and regulations by the governmental entity that may have affected the use of federal funds.

There is no better evidence of the imperative of internal controls in modern organizations than the passage of the Sarbanes-Oxley Act in 2002 (PL 107-204; see <http://corporate.findlaw.com/industry/corporate/docs/publ107.204.pdf> for a full text of the law). Following in the wake of corporate scandals, Sarbanes-Oxley does many things to address abuses, but section 404 of the Act specifically addresses internal controls. Subsequent regulatory actions by the Securities and Exchange Commission corporations are required to file reports assessing internal controls over financial reporting (<http://www.sec.gov/news/press/2003-66.htm>). In addition, the International Organization for Standardization (ISO), which is a network of national standards institutes for 148 countries, issued ISO 9000 in 1987. Although ISO 9000 focuses on quality management systems, it has the effect of providing standards for an integrated internal control system (see <http://www.iso.ch/iso/en/iso9000-14000/iso9000/>

iso9000index.html). Updated in 2000, these international standards utilize eight quality management principles that organizations can certify to that in large measure parallel requirements of an internal control system.

DEFINING INTERNAL CONTROLS

The federal government, as well as many state governments and other organizations concerned about internal controls, have embraced the COSO-based definition. A host of other professional bodies have built upon this definition to guide their efforts. For example, internal control organizations such as the Financial Reporting Council of the London Stock Exchange (the Cadbury Report), the Criteria of Control Board of the Canadian Institute of Chartered Accountants (COCO), the International Organization of Supreme Audit Organizations (INTOSAI), and the Information Systems Audit and Control Association (Control Objectives for Information and Related Technology—COBIT) embrace the COSO definition of internal control:

Internal control is a process, effected by an entity's board of directors, management and other personnel, designed to provide reasonable assurance regarding the achievement of objectives in the following categories: effectiveness and efficiency of operation; reliability of financial reporting; and compliance with applicable laws and regulations. (From <http://www.coso.org/key.htm>.)

While there are both similarities and differences among these documents (COSO, COCO, INTOSAI, Cadbury Report, and COBIT), they are consistent in contending that internal control has taken on a new and expanded meaning . . . that affects virtually every aspect of an organization's operation (New York, 1999).

The standardized COSO-based definition is built on several fundamental concepts. These concepts are that internal control is a process, a means to an end, not an end in itself. Also, internal controls are affected by people; therefore, they are reliant upon the actions and decisions made by all levels of an organization.^[7] Internal control is a basic element that permeates an organization, but is not just an addition to it; it incorporates the qualities of good management; and internal control provides a level of comfort to an organization but does not guarantee success (State of New York, 1999).

Most government organizations use the five major components of internal control outlined by COSO that apply to all aspects of an organization's operations: programmatic, financial, and compliance. These components

are necessary in that by accomplishing them, the organization is one step closer to having a truly effective internal control system. The components outlined are: control environment, communication, assessing and managing risk, control activities, and monitoring (Fig. 1) (State of New York, 1999; Commonwealth of Massachusetts, 2000).

The *control environment* is defined as “the attitude toward internal control and control consciousness established and maintained by the management and the employees of the organization” (State of New York, 2000). The control environment is essentially composed of the organization's management philosophy, style, supportive attitude, competence, ethical values, integrity, and morale. *Communication* is defined as “the exchange of useful information between and among people and organizations to support decisions and coordinate activities.” A key component of communication is that information must be transferred to management and the employees within a reasonable time frame to allow them to achieve the organization's goals. *Control activities* are defined as “the tools—both manual and automated—that help prevent or reduce the risks that can impede accomplishment of the organization's objectives and mission.” Examples of such control activities are documentation, approval and authorization, verification, supervision, and separation of duties. The fourth component to internal control is that of monitoring. *Monitoring* is defined as “the review of an organization's activities and transactions to assess the quality of performance over time and to determine whether controls are effective.” A key factor to note within the monitoring component is that everyone within



Fig. 1 COSO model. (Reprinted with permission from COSO.)

an organization has some responsibility for monitoring, and the focus of that individual depends upon the position held. The final component is that of *assessing and managing risk*. Risk is defined by the State of New York as “events that threaten the accomplishment of objectives.” Therefore, risk assessment and management is “the process of identifying, evaluating, and determining how to manage these events.” Management should identify and evaluate each risk in terms of significance, likelihood, and cause, and they must therefore provide safeguards against the most qualifying of these assessments.

INTERNAL CONTROLS IN THE FIELD: SELECTED STATE EXAMPLES

The State of North Carolina offers its agencies and smaller government entities an outline of expected policies to be adopted in order to maintain “reliability of financial reporting, efficiency and effectiveness in operations, compliance with applicable laws and regulations.” (See p. 3, 4 of Ref.^[8].)

The State Auditor’s Office (SAO) in the State of Texas builds upon the COSO concept by including a self-assessment program. “The effect of the action would be to expand management’s understanding of the controls, provide more knowledgeable people to assist in controls resources and expand the utility of the internal audit function.”^[9] The state suggests that agencies and local governments use a risk assessment methodology that includes all levels of management. This would result “in a more accurate and comprehensive risk assessment; fewer special projects to interrupt the audit plan; and more effective use of internal audit resources.”^[9]

The Commonwealth of Massachusetts’ Office of the Comptroller’s *Internal Control Guide for Managers* and also *An Act Relative to Improving the Internal Controls Within State Agencies*^[10] outline six standards for internal control for departments. Those six standards are: documentation, transaction management, authority, segregation of duties, supervision, and access privileges. These six standards overlap greatly in the areas of financial management, overall management, and auditing. The guide then briefly discusses both the department’s and the state auditor’s responsibility in the department’s internal control system.

Finally, the State of New York, through the Office of the State Comptroller and the Division of the Budget, identifies and issues its departments certain internal control standards. Although the Comptroller’s guide relates the importance of internal control systems in budgeting, financial management, and auditing, its clear focus is on that of management. “To some degree, everyone in an organization has the responsibility

for ensuring the internal control system is effective. However, the greatest amount of responsibility rests with the managers of the organization.” (See p. 6 of Ref.^[11].) Simply put, the organization’s top executive has the ultimate responsibility for the effectiveness in operations of the internal control system, since he/she has been entrusted with achieving the organizations goals. The *Standards for Internal Control in New York State Government* state, “the strength of the system of internal control is dependent on people’s attitude toward internal control and their attention to it. Executive management needs to set ‘the tone at the top’ in this regard.”

Further refinement of New York State’s internal control system led to the creation of Internal Control Officers (ICO), similar to positions in the Commonwealth of Massachusetts (see <http://www.nysica.com/>). The main responsibilities of these officers are to establish a system of internal control, review process and management policies, and implement education and training concerning internal controls. Even with a clearly identified internal control position, the guidelines reiterate that “it is management’s responsibility to establish the internal control system and for preventing breakdowns from occurring.”

INTERNAL CONTROLS IN THE FIELD: SURVEY RESULTS

A recent survey of state governments shows some interesting trends in how internal controls are utilized in the responding states.^[12] The 12-question survey was designed to assess the current “state of the state” of internal controls in the 50 U.S. states. Senior central budget office staffs in all 50 states were sent the survey. Twenty-two states responded—a 44% return rate—with one state responding from three different offices (budget, auditing, and accounting offices). First, 44% of the responses stated that the controller, comptroller, or treasurer had primary responsibility over the internal control system, and another 32% responded that internal control was decentralized to individual agencies. Only 12% responded that the central budget office had primary responsibility over the state’s internal control system.

Another finding was that only 56% of respondents believed that each state agency had one individual primarily responsible for internal control implementation and oversight. This finding shifts attention to the issue of dispersion of internal control implementation and oversight, and the potential consequences that arise when a multitude of individuals are responsible for the internal control system.

Survey respondents believed that their internal control environment was strong; yet, only 60% agreed or

strongly agreed. At the same time, 84% of respondents believed that their internal control system was used to make improvements in the state or agency operations. However, 40% of respondents agreed or strongly agreed that internal controls are viewed primarily as a financial control rather than a part of daily management operations. Moreover, 32% of the respondents agreed that internal controls and related reporting requirements are viewed as extraneous or supplemental to budgeting or managerial responsibilities.

This survey suggests that internal controls have been widely implemented by state governments, but to varying degrees. A preliminary observation about the states included in the sample was that the responsibility for internal control was largely decentralized to individual agencies. This would seem to indicate that states embrace the notion of internal controls being the responsibility of front line managers, but most states still retained a level of centralized oversight and control over the process. This opens a question about the diffusion of responsibility for internal controls in state government and who is ultimately responsible for internal control failures. With respect to the utility of internal controls for management improvement in state government, the survey suggests the linkage to such improvements is problematic at best.

CONCLUSIONS

Although a single definition is possible and identification of key elements can be presented to describe internal controls, it is nonetheless more accurate to say that internal controls are not one concept or even one process—they are the application of multiple concepts and processes that seek to enhance an organization's efficiency and effectiveness. In many respects, internal controls represent the best of what a management command and control structure has to offer, but at the same time, internal controls embrace empowerment of all organizational actors as responsible for

efficiency, effectiveness, and accountability. Hence, internal controls represent a powerful, underutilized, and little understood tool in the repertoire of managers across all sectors and types of organizations.

REFERENCES

1. United States General Accounting Office. *Standards for Internal Control in the Federal Government*; November 1999.
2. University of California, Office of Controls and Accountability. *Understanding Internal Controls: A Reference Guide for Managing University Business Practices*; 1999 (www.ucop.edu/ctlacct/under-ic.pdf).
3. Tipgos, M.A. Why management fraud is unstoppable. *CPA J.* **2002**, December, 34–41.
4. McHugh, J.A. State management systems: the case for internal controls. *Assoc. Gov. Accountants J.* **1988**, Summer, 65–72.
5. Tanki, F.J.; Steinberg, R.M. Internal control—integrated framework: a landmark study. *CPA J.* **1993**, June, 16–20.
6. Cox, C.T.; Wichmann, H., Jr. Reporting on internal controls for governmental units. *Intern. Audit.* **1992**, Summer, 37–47.
7. Commonwealth of Massachusetts, Office of the Comptroller. *Internal Control Guide for Managers*; 2000.
8. State of North Carolina, Department of State Treasurer. *Internal Controls Policy*; Section 80; 2000.
9. State of Texas, State Auditor's Office. Best Practices List (<http://www.sao.state.tx.us/resources/intaud/bestprac.html>).
10. Commonwealth of Massachusetts. *An Act Relative to Improving the Internal Controls Within State Agencies*; 1989; Chapter 647.
11. State of New York, Office of the State Comptroller. *Standards for Internal Controls in New York State Government*; 1999.
12. Smith, R.W. *Results of a Web-Based Survey of Internal Controls in all 50 States*, Unpublished Research Report, 2003.

International Public Sector Accounting Standards

James L. Chan

*Department of Accounting, University of Illinois at Chicago,
Chicago, Illinois, U.S.A.*

INTRODUCTION

International Public Sector Accounting Standards (IPSAS) are financial measurement and reporting rules recommended for adoption by governments around the world and are promulgated by the Public Sector Committee (PSC) of the International Federation of Accountants (IFAC).

Government officials should care about IPSAS for practical reasons. Their governments are directly or indirectly urged to consider adopting IPSAS. If a country already has government accounting standards, convergence or harmonization with IPSAS is recommended. Scholars in public administration and public policy, especially those specializing in budgeting and financial management, would find IPSAS intriguing. International Public Sector Accounting Standards represent a new kind of government accounting that challenges the traditional superiority of budgeting over accounting in government, and the conventional notion that budgeting and financial management practices are government's own business.

After identifying the political and economic forces that have led to the emergence of IPSAS, the article describes the institutional structure and process for setting IPSAS. The standards are explained in non-technical terms and commented upon. Implications for public administration are stated in the Conclusion.

POLITICAL AND ECONOMIC FORCES

The winds of democracy and market economy sweeping the globe during the last two decades have encouraged government accountability and transparency.^[1] These two values are the foundation for the practice of accounting and auditing in the public sector. Even when it is narrowly and condescendingly regarded as glorified bookkeeping, accounting can still contribute to government fiscal accountability by requiring accurate financial recordkeeping of public money received and spent by government. The creation and preservation of such audit trails could deter corruption and graft—violations of elementary fiscal accountability.

The modern practice of government accounting is much more than financial recordkeeping. It can facilitate sound financial management by making sure that

the budget is faithfully and efficiently executed through authorized transactions and effective government actions. In particular, financial accounting in government is the systematic process of identifying, recognizing, and measuring the financial consequences of these government transactions and actions. Up to this point, accounting, even at its best, is only an adjunct to administration: the information lies inside government. As such, the information helps the government manage its financial affairs, but it does not help stakeholders hold government accountable. Stakeholder groups include legislators and regulators; voters, taxpayers, and the general public; creditors and investors in government securities; employees and vendors; and grantors and donors. All these stakeholders provide economic and political resources to government, thereby acquiring the right or need to know its finances.

The contemporary interpretation of the scope of accounting extends to financial reporting, especially to those who have the political right to know and economic need to know and yet lack the authority and resources to demand information. Accounting thus has become both a product and an enforcement mechanism of democratic accountability in the political system and of economic reciprocity in a market economy. These are the values championed by Western democracies, and by the international institutions they support and dominate especially in the post-Cold War era. No matter how imperfect democracy is and how poorly the market economy work in a particular country, as normative ideals they have inspired what has been called “a global revolution in government accounting.”^[2]

The revolutionary thinking in government accounting has several key elements. 1) Governments should issue audited financial reports shortly after the end of a fiscal year. Most governments prepare budgets, but not all prepare and make public financial reports derived from their accounting systems to serve as a factual check on the budget. 2) Financial reports should not be limited to cash receipts and disbursements, but should report the government's economic resources and short- and long-term liabilities. This would be a major departure from the traditional cash budget.^[1] Government accounting and reporting rules should be set by independent bodies and not by governments themselves. Year-end financial reporting on

the organization as a whole on the accrual basis in accordance with generally accepted accounting principles (GAAP) is required of public and private institutions in the United States and developed countries in the British Commonwealth. The “global” dimension of the revolution reflects the influence of these countries on the international and financial institutions that provide financial and technical assistance to developing countries and those in transition to a market economy.

It is also the case that Anglo-American private-sector accountants and auditors have the tradition of involvement in government. The involvement takes several interrelated forms. Their professional bodies might be directly or indirectly involved in setting government accounting standards. They might appoint a technical committee, or cosponsor a standard-setting body. Or they might designate the authoritative bodies for promulgating GAAP. These principles are used by their members to evaluate the quality of financial statements in order to express audit opinions. When governments engage private-sector auditors to audit their financial statements, accounting professional groups have considerable leverage over governments. This relationship between the accounting/auditing profession and government is elevated to the international level in creating the mechanism for setting IPSAS.

STANDARD-SETTING INSTITUTION AND PROCESS

The IFAC is worldwide alliance of 156 accountancy bodies in nearly 120 countries.^[3] The IFAC’s PSC, with financial support of several international organizations, began a standards program in 1996 to promote government financial accountability and performance through better accounting and financial reporting.^[4]

The committee currently (in the year 2004) has part-time members from Australia, Argentina, Canada, France, Germany, Israel, Japan, Malaysia, Mexico, the Netherlands, New Zealand, Norway, South Africa, the United Kingdom, and the United States. Before its current French chairman took office, the committee had been chaired by a Canadian, a New Zealander and an Australian. The committee has a small full-time staff and is assisted by the volunteer efforts of a large number of consultants and participants on its Consultative Group and Steering Committees for various technical projects.^[5]

International Public Sector Accounting Standards are set through a due process. Research is undertaken and deliberations are held before tentative positions are adopted. Exposure drafts of proposed standards

are disseminated to solicit the views of interested parties. These views are considered in revising and finalizing a standard. While the primary purpose of this process is to improve the quality of the standards, a side benefit is a greater understanding and support for the final standards and the institution for setting IPSAS in general. All the committee’s documents are available online for free downloading.^[6]

Financial support for the committee’s standard-setting project has come from the International Monetary Fund, the World Bank, the United Nations Development Program, and the Asian Development Bank. Institutional collaboration, but not funding, is provided by a group of observers: the International Organization of Supreme Audit Institutions, Organization for Economic Cooperation and Development, International Accounting Standards Board, and the European Commission.^[7]

The issue of IFAC/PSC’s future was recently addressed by a review panel named by the IFAC and chaired by an external chairman. The panel has recommended that the committee be renamed the IPSAS Board and that it continue to develop additional standards.^[7] Such a structure would require continued funding from organizations interested in increased quality of external reporting by governments. Since IPSAS have attributes of public goods, their production and dissemination will have to rely on the donors identified above. These organizations stand to benefit from reliable government financial data, and, if they choose, could require IPSAS be used in accounting for their loans and other forms of assistance.

INTERNATIONAL PUBLIC SECTOR ACCOUNTING STANDARDS

As of August 2004, IPSAS consists of 20 standards: two on accounting measurement issues; 8½ on financial reporting issues; and 9½ dealing with specific transactions or items in financial statements.^[5] These standards are highly technical and are almost impossible to understand without a working knowledge of financial accounting. As a whole, IPSAS are impossible to assess without examining their basic assumptions. Consequently, this section will state the essence of IPSAS without attempting to summarize individual standards (see Table 1 for key terms); it will also try to discern the premises underlying the standards.

The PSC believes that governments should prepare government-wide (consolidated) financial statements. These general-purpose statements are intended for users who lack the ability to request specific information. A complete set of financial statements would describe the government’s financial position,

Table 1 Key terms in financial accounting and reporting*Financial reporting*

Reporting entity: consists of one or more organizations whose finances are of interest to intended users of a financial report.

Financial statements: report the recognized effects of transactions and events, consisting of statements of financial position and of financial performance, and of cash flows.

Government-wide (or consolidated) financial statements: financial overview of the government as a whole, after eliminating the effects of internal transactions.

General-purpose external financial statements: intended to be the primary source of historical financial information to users outside of an entity, prepared with generally accepted accounting principles.

Financial report: financial statements and other financial disclosures.

Financial position: as of a particular date, an entity's net assets are its assets offset by its liabilities.

Financial performance: changes during a period in an entity's amount of net assets, as measured by revenues net of expenses.

Cash flows: sources and uses of cash from operating, financing, and investing activities.

Financial accounting measurement

Transactions: the primary source of financial accounting data, exchanges of economic considerations between two or more parties.

Event: a significant occurrence that has direct or indirect consequences to an entity.

Financial consequences: impacts on an entity's financial position and performance or cash flows.

Elements of financial statements: indicators of financial consequences and therefore major components of financial statements: assets, liabilities, revenues, and expenses.

Assets: range from cash to any economic resources capable of producing future benefits.

Liabilities: obligations to pay in cash or to provide services due to having received benefits.

Revenues: increases in assets (or decreases in liabilities) attributable to service efforts and accomplishments or government's power to tax.

Expenses: assets used or liabilities incurred in the course of generating revenues or providing services.

Realization: financial consequences having been confirmed by an actual transaction or event.

Recognition: the decision and action to formally acknowledge certain financial consequences in the accounts and financial statements.

Cash basis: measures financial performance in terms of cash receipts and disbursements.

Accrual basis: in contrast to cash basis, broadens revenue to include receivables (claims against customers or taxpayers) and expenses to include non-cash resources used and incurrence of liabilities.

Matching: offsets revenues by related expenses to arrive at a net periodic measure of financial performance (i.e., income or loss).

financial performance, and cash flows. Comparisons with the budget are encouraged but not required. The reporting entity should encompass the entities owned and controlled by the government. Appropriate disclosures should be made about distinguishable segments within the government, as well as the government's financial relationships with other entities.

The information in financial reports is produced by the financial accounting system of an entity. The primary function of financial accounting is to recognize and measure the financial consequences of past transactions of an entity and of the events that directly affect it. The consequences are measured by a number of financial variables. The accounting equation "assets minus liabilities equals net assets" describes financial position. Financial performance is measured by revenues and expenses, and therefore income or loss. The inclusion of non-current assets and liabilities is necessary to fairly portray a government's

financial position at the end of the period. The accrual basis of accounting provides a more accurate measures of service effort and accomplishments than does the cash basis of accounting. The use of the accrual basis requires a complementary statement of cash flows from operations and investing and financing activities to show changes in the government's liquidity position.

The key terms in Table 1 and the above description are intended to help readers put the individual IPSAS in a better organized framework. The first three IPSAS describe the form and contents of the three financial statements that constitute general-purpose financial statements. After that, the sequence of the individual statements does not reveal any discernible pattern. As a remedy, Table 2 differentiates general and specific standards. The very small number of standards on the important topics of recognition and measurement is noticeable.

Table 2 Structure of current IPSAS (number and issue date)*General recognition and measurement standards*

- Revenue from exchange transactions (no. 9; 2001)
- The effects of changes in foreign current exchange rates (no. 4; 2000)

General reporting standards

- Presentation of financial statements (no. 1; 2000)
- Cash flow statements (no. 2; 2000)
- Consolidated financial statements and accounting for controlled entities (no. 6; 2000)
- Segment reporting (no. 18; 2002)
- Financial reporting of interests in joint ventures (no. 8; 2000)
- Related party disclosures (no. 20; 2002)
- Financial reporting in hyperinflationary economies (no. 10; 2001)
- Events after the reporting date (no. 14; 2001)
- Fundamental errors and changes in accounting policies (no. 3; 2000)

Standards on specific transactions and items

- Financial instruments: disclosure and presentation (no. 15; 2001)
- Accounting for investments in associates (no. 7; 2000)
- Inventories (no. 12; 2001)
- Investment property (no. 16; 2001)
- Property, plant, and equipment (no. 17; 2001)
- Lease (no. 13; 2001)
- Provisions, contingent liabilities and contingent assets (no. 19; 2002)
- Construction contrasts (no. 11; 2001)
- Borrowing costs (no. 5; 2000)
- Net surplus or deficits for the period (no. 3; 2000)

In addition to the standards in Table 2, in response to some countries' inability or reluctance to adopt the preferred accrual basis, a standard on how to account and reporting under the cash basis was also issued. This is, however, regarded as an interim measure since all countries are urged to move to the accrual basis, even though many conceptual issues and implementation obstacles remain.

The standards in Table 2 are the products of Phase 1 of the PSC's standards program from 1996 to 2002. They are said to constitute a "core" set of IPSAS for public sector entities. This is an overstatement because those standards hardly address the core issues in government accounting. The core issues in government accounting have to do with the unique characteristics of government: the power to tax, the production of public goods, virtually limitless responsibility, unfathomable public property (at least in financial terms), and the disclosure of budgets prior to the fiscal year (a norm in Western countries).

If the standards are to be useful for developing countries and transitional economies, the standards would have to deal with issues that connect business

and government. These concern the shifting boundaries between public and private sectors, the conversion of some public property to private property, government subsidies for state-owned enterprises, government assumption of debts of failing or failed enterprises, etc. The list can go on. These are not easy and technical accounting issues, because the institutional mechanisms for defining and enforcing legal and contractual rights and obligations are either weak or non-existent. Precisely for this reason, the research and development of accounting standards could be instrumental in reinforcing the political, economic, and institutional infrastructure of accountability.

Phase 2 of the PSC's standards program from 2002 to 2005 is addressing some of the issues unique to the public sector. These include: 1) whether and how budgetary information should be included in financial statements; 2) non-reciprocal revenues, such as taxes, grants, and transfers; 3) social policy commitments, such as social insurance; 4) assistance to developing countries; 5) heritage assets; and 6) public/private-sector construction/financing arrangements.^[4,5] Information about the current status of these projects as well as the committee's activities is available from the Internet (www.ifac.org).

The structure and contents of the first 20 IPSAS are based on several assertions. First and foremost is the belief that it is possible and desirable to have a single set of financial accounting and financial reporting standards for both business and government. Even though terms like convergence and harmonization are used, in practice, the current IPSAS are modest adaptations of business accounting and financial reporting standards promulgated by the International Accounting Standards Committee (now Board). Second, the PSC, being an international professional group, should focus on external financial reporting by government. By implication, compliance with legal, budgetary, and contractual requirements and support for financial management are outside of the PSC's concern. Third, it is appropriate to take the incremental approach by setting IPSAS one by one; a conceptual framework is in effect regarded as a luxury not as a necessity. These are debatable points. One might argue on the contrary that: there are numerous and critical differences between government and business; a government accounting systems have both internal and external missions; and a conceptual framework is essential for ensuring coherent and consistent standards.^[1,8]

CONCLUSIONS

International Public Sector Accounting Standards are in essence the worldwide accounting profession's advice to governments on how to report to the public and other

external parties. Based on the premise that government accounting should be as similar to business accounting as possible, they reflect Anglo-American thinking and practice. As such, the United States and developed nations in the British Commonwealth already comply with the spirit, if not the letters, of IPSAS. Other developed nations may, but do not have to, adopt IPSAS unless they are persuaded as to the merits of these standards. However, developing countries now do not have such discretion. According to the IFAC, the World Bank has accepted IPSAS for preparing financial reports on activities financed by the bank.^[5] If other international donors and lending institutions follow the World Bank precedence and begin to require IPSAS, developing countries will have to consider whether to embrace IPSAS as government-wide accounting rules.

Such a decision will not be easy, in part because IPSAS are still a work in progress: standards addressing unique government issues are still being formulated. If the PSC continues to use the incremental approach, individual standards will be issued over a long period of time. Changing government accounting systems to comply with IPSAS will predictably be a prolonged and costly process. Budget directors and chief administrators responsible for approving and justifying the costs to legislatures may soon be asking: what price is more, even if better, accounting? While it is laudable to demonstrate external accountability to external users with financial statements, should not more resources be devoted to forward-looking

planning and real-time financial control system? Answers to these questions may be too important to be left to accountants. It behooves public administrators and legislators to monitor and participate in the development of national and international government accounting standards.

REFERENCES

1. Chan, J.L. Government accounting: an assessment of theory, purpose and standards. *Public Money Manage* **2003**, 23 (1), 13–20.
2. Heald, D. The global revolution in government accounting: introduction. *Public Money Manage* **2003**, 23 (1), 11–12.
3. International Federation of Accountants (IFAC). 2003 Annual Report; IFAC: New York, 2004.
4. Sutcliffe, P. The standards programme of IFAC's Public Sector Committee. *Public Money Manage* **2003**, 23 (1), 29–36.
5. IFAC. Public Sector Committee Update 12, July 2004.
6. IFAC Public Sector Committee. In *Handbook on International Public Sector Accounting*; IFAC: New York, 2004.
7. IFAC. Report of the Externally Chaired Review Panel on the Governance, Role and Organization of the International Federation of Accountants Public Sector Committee, June 2004.
8. Chan, J.L. Une revolution mondiale dans la comptabilite publique? Une analyse des IPSAS et quelques recommandations. *Rev. Fr. Comptabilite* **2004**, 362, 27–31.

Inventory Management

Ronald John Hy

*College of Arts and Sciences, Texas A&M University-Kingsville,
Kingsville, Texas, U.S.A.*

INTRODUCTION

Sound inventory management is designed to save money by minimizing inventory costs. Inventories are items—such as building materials, office supplies, and chemicals, to name a few—that are used to produce products and/or services. Managing inventories efficiently is important because inventories are among the largest non-wage expense of most public and non-profit agencies.^[1]

Inventory management saves money by purchasing as inexpensively as possible the optimal number of needed items so that they are always readily available, but in an amount that minimizes storage and holding costs. Having a sufficient number of items readily available, therefore, must be balanced against the cost of storing and holding items.

Inventory management requires:

- Preparing specifications, obtaining competitive bids, negotiating, and receiving items
- Buying specific items
- Monitoring the deterioration, obsolescence, storing, issuing, theft, handling, interest, and insurance of items
- Overseeing the storage and holding costs of items.

The most common mistakes made by inventory management are:^[2]

- Stocking too many items
- Stocking items which are seldom, if ever, needed
- Failing to balance purchasing cost and storage costs
- Failing to consider the full cost of holding inventories

Sound inventory management, then, controls costs and thereby saves money by determining the number or amount needed to be readily available, while keeping storage cost to a minimum. Inventory management also contributes to the availability of working capital, which refers to the organization's short-term or current assets (e.g., cash, short-term marketable securities, accounts receivable, and inventories). Spending as few dollars as needed on inventory allows an agency to spend more money on products and/or services,

to invest more money for future use, or to pay off current debts.

VALUING INVENTORIES

In terms of accounting, inventory usage either is first-in, first-out (FIFO) or last-in, first-out (LIFO). With FIFO, the oldest items in the inventory are used first. With LIFO, the newest items in the inventory are used first. Using one of these two systems is necessary because each item in an inventory is not labeled individually. Thus, when similar items have been purchased at different prices, which frequently is the case, the price of each particular item may not be readily known, just the price of a group of items. For example, if the purchasing department buys a chemical at \$5 an ounce at one time and \$6 an ounce at another, it probably does not know which chemicals were purchased at \$5 and which were bought at \$6 because in storage the two shipments are ordinarily mixed together. Purchasing, however, surely knows how many chemicals were bought at \$5 and how many were bought at \$6. Therefore, to keep track of costs the agency has to use a consistent policy, either FIFO or LIFO. A consistent, systematic policy is the only way to keep an accurate record of actual inventory costs.

Under FIFO, the first item leaving the inventory is costed at the newest price in the inventory. Consequently, the value of the inventory at any point in time is determined by the most recent cost of the item. Under LIFO, the first item leaving the inventory is priced at the most oldest price in the inventory. In inflationary periods, using the higher-priced item has a decided advantage over using the lower-priced item because the most expensive items are used first and the least expensive items are held in inventory. Last-in, first-out pricing lowers the value of the inventory over time. Consequently, not as much money is tied up in inventory, and the inventory cost is less. Last-in, first-out pricing also allows the agency to increase reimbursable costs.^[3]

Besides lowering costs and saving money, inventory management keeps inventories amply stocked. Because the time lag between ordering and receiving some items is sometimes substantial in terms of money, time, and

even life, it is important that these items be available for immediate use. An agency cannot always project with certainty the number of items that will be needed at any given time. Therefore, to meet the demand and ensure the continuation of services, inventories must be kept suitably stocked. If a sufficient number of cast iron pipes and fittings are not readily available, then water lines cannot be fixed even when personnel are available to work on the lines. A sufficiently stocked inventory will prevent such discontinuities from occurring.

COSTING INVENTORIES

Given the need to minimize inventory cost, an agency must decide on the optimal number of items to keep in inventory. Inventory management relies on various formulas that are based on certain assumptions of future behavior, as well as common sense and experience.^[4] Inventory cost consists of purchasing cost, order cost, carrying cost, overstock cost, and stock-out cost. The inventory formulas discussed in the following sections are taken from Breman, Kukla, and Weeks.^[5]

Purchasing Cost

Purchasing cost is the price paid for each item. This expense is essential. Even though an agency has little control over the price it pays for items (the price is set by the seller), it can reduce the purchase cost by buying in large enough quantities to qualify for a discount. The timing of the order, as well as its specifications and bidding procedures (when applicable), also influence the price of items. As a result, the price of any item should always be considered to some extent variable and negotiable.

The purchasing cost formula is:

$$\text{Purchasing cost} = PN$$

where:

P = price of each unit.

N = number of units purchased annually.

Order Cost

Order cost is direct expenses incurred when purchasing items, expenses such as writing specifications, letting and analyzing bids, preparing order forms, and receiving items. This cost normally is minimal, especially when items are standard and purchasing processes

routine. However, first-time, unstandardized orders do increase the order cost. The order cost, unlike the purchasing cost, varies with the number of orders placed, not the size of the order. Therefore, the order cost component of the inventory cost encourages large and fewer orders.

The order cost formula is:

$$\text{Order cost} = (D/Q)O$$

where:

D = number of units purchased annually.

Q = size of the order.

O = average cost of placing an order.

Carrying Cost

Carrying cost is the expense involved in holding items in the inventory. This cost actually consists of two elements: opportunity costs and storage costs. Opportunity costs are incurred when the organization spends money purchasing and holding items for inventory, instead of spending the money for something else that will reduce its operating cost (e.g., new equipment). Opportunity costs occur when an agency foregoes the opportunity to buy something other than inventory in favor of stocking the inventory. Generally, the greater the inventory cost, the greater the opportunity costs.

Besides opportunity costs, carrying cost includes storage costs, which are incurred when organizations, as they must, properly store, insure, secure, and otherwise protect their inventory. Storage costs tend to be positively related to the size of the inventory. As with opportunity costs, the larger the inventory, the larger the storage costs.

The carrying cost formula is:

$$\text{Carrying cost} = HQ + IP(Q/2)$$

where:

H = storage cost per unit, determined in much the same manner as rates are fixed.

Q = size of the order, usually the maximum number of items stored at any given time.

$IP(Q/2)$ = opportunity cost,

where:

I = interest.

P = price per unit.

$Q/2$ = average inventory holdings.

Overstock Cost

Overstock cost is the expense of holding unused or unneeded items in the inventory. This cost consists of surplus carrying costs and perishable costs. Surplus carrying costs exist when items are purchased for use in a particular time period, but are not used during that period. Thus, an agency is forced to pay an unexpected carrying cost for the surplus items until they can be used. Overstocking, therefore, contributes to an increase in the carrying cost.

The second element of overstock cost is perishable costs. Although perishable costs do not apply to most public and non-profit agency inventories, they do to some, especially foodstuffs and chemicals. If an agency overstocks and is unable to use a chemical before the expiration date, it incurs a cost similar to that of a tomato that must be discarded.

The overstock formula is:

$$\text{Overstock cost} = [HQ + IP(Q/2)] + S$$

where:

$HQ + IP(Q)/2$ = carrying cost.

S = perishable costs.

Stock-Out Cost

Stock-out cost is the expense incurred when there is an insufficient supply of a needed item. Keeping the stock-out cost minimal increases the inventory cost. This trade-off, however, is almost mandatory when avoiding stock-out is critical. Generally, public and

non-profit agencies seldom, if ever, experience this cost. They generally compensate by keeping substantial inventories on hand.

In summary, inventory cost (IC) is based on the following components:

$$IC = PN + (D/Q)O + [HQ + IP(Q/2) + [[HQ + IP(Q/2)] + S]$$

When using this formula, one should keep in mind that, according to Berman, Kukla, and Weeks, about 15% of inventory items represents 70% of the inventory value.^[5] Another 15% represent approximately 25% of the inventory value. The remaining 70% of the items account for only 5% of the inventory value. By concentrating on the 30% of the items representing 95% of the inventory and ignoring the rest, one can more readily and systematically estimate cost factors.

REFERENCES

1. Hy, R. *Financial Management for Health Care Administrator*; Quorum Books: New York, 1989; 119 pp.
2. Lynch, T. *Public Budgeting in America*; Prentice-Hall: Englewood Cliffs, NJ, 1979; 295 pp.
3. Lusk, E.; Lusk, J. *Financial and Managerial Control: A Health Care Perspective*; Aspen: Rockville, MD, 1979; pp. 102–107, 125.
4. Lynch, T. *Public Budgeting in America*; Prentice-Hall: Englewood Cliffs, NJ, 1979; 295 pp.
5. Berman, H.; Kukla, S.; Weeks, L. *The Financial Management of Hospitals*; Health Administration Press: Ann Arbor, MI, 1974; pp. 261, 268.

Israel Civil Society: Characteristics

Nitza Nachmias

School of Political Science, University of Haifa, Haifa, Israel

Amiram Bogot

Former Registrar, Non-Profit Organizations, Jerusalem, Israel

INTRODUCTION

In the early 1980s the term “civil society” (CS) turned into a “user friendly” theoretical tool, gaining great attention from scholars and practitioners alike. Known variously as “Non-profit,” “Voluntary,” “Third Sector,” and “Non-government Organizations (NGO),” civil society organizations (CSO) are operating outside the confines of the market and the state. This set of institutions provide a middle way between sole reliance on the market (the “second sector”) and sole reliance on the state (the “first sector”). CSOs, also known as NGOs, include a sometimes bewildering array of entities—hospitals, universities, social clubs, professional organizations, day-care centers, environmental groups, sports clubs, job training centers, human rights organizations, and many more.

The new CS approach has replaced both the Schumpeterian paradigm of democracy and the paradigm of the “welfare state.” A comparative study of CSOs in 22 developed countries shows that “even excluding religious congregations, the non-profit sector in 22 countries we examined is a \$1.1 trillion industry that employs close to 19 million full-time equivalent paid workers.”^[1] In Israel, as in most liberal democratic countries, the government pays third party service provider organizations to deliver social services to the citizens. These CSOs replace or complement public services in such areas as education, health, immigrant absorption, welfare, and culture. The practice of subcontracting and outsourcing of social services to CSOs follows the New Public Management (NPM) approach that maintains an intense sharing relationship between the government and a myriad of service providing CSOs. While the use of NPM in Israel has widened in the past two decades, the practice has gained both praise and criticism. New Public Management was hailed, mainly by politicians, for promoting democracy, effectiveness, and efficiency in the delivery of public services. But important regulatory and oversight institutions, namely, the state comptroller’s office (SCO), and the courts, voiced strong criticism regarding rampant fraud in the subcontracting procedures, and the unwarranted distribution of public funds to NGOs whose activities were neither necessary nor relevant to the public interest.

Notwithstanding the criticism, the CS in Israel is among the largest in the world relative to the economy as a whole. It takes fourth place among the 22 developed nations after Holland, Ireland, and Belgium, and its economic weight amounts to nearly one-tenth of the Israeli economic activity. However, research on the Israel CS has been limited and late in developing. Serious research began only in the 1990s, with the establishment of the Center for Non-Profit Organizations at Ben Gurion University, headed by Gidron. The rapid growth of the Israeli Third Sector has caught the government unprepared and unequipped to regulate, control, or oversee the burgeoning “civil society.” In fact, a tactical conflict has emerged between the Israel CSOs aspiring to maximize their public support while minimizing oversight, and the government aspiring to maximize its oversight while minimizing its cost. Our research shows that even though the government had introduced new and unique regulatory mechanisms, these bureaucratic and legal instruments need to be further strengthened to ensure the CS accountability.

THE ISRAEL CIVIL SOCIETY

The Israel CS includes over 30,000 registered non-profit associations of individuals who claim to share a common purpose with the rest of the members of the group. Many groups receive direct and indirect public support. The main characteristics of the diverse network of Israel CS are that: 1) they are required to incorporate by law for specific purposes detailed in their bylaws; 2) they must exhibit a firm bureaucratic structure and permanent governing institutions; and 3) they have to strive to achieve specific, detailed economic, social, or strategic goals. Finally, their activities have to be congruent with Israel national interest. Most of the Israeli NGOs’ activities concentrate on three areas: 1) education and research; 2) culture and recreation; and 3) religious activities. The Ministry of Finance (MOF) data show that the highest number of NGOs is found in the religion section. Indeed, 44% (10,358) of all the registered CSOs are Orthodox/Haredi, but only a quarter of them (2635) are

active. In contrast, in the late 1990s, only 1067, or 4.5% of all registered NGOs were Arab Israeli groups, and only 28% of them have been active. Between 1991 and 1997, the religious-orthodox education system received 37.2% of the funds transferred to all NGOs. In comparison, Israel's higher education CSOs received 3.6% of the total funds (Table 1).

Historically, the Jewish NGOs in the Diaspora provided all the social services for the community. This tradition continued in Palestine where between 1882—the year considered the beginning of the Jewish settlement in Palestine—and until 1948, a highly sophisticated network of NGOs functioned as an effective and efficient *bone fide* government, providing all the social services for the Jewish community in Palestine. The newly created state maintained the long tradition of support (financial and other) by Diaspora Jews for the Jewish community living in the Holy Land. During the first Israel coalition government headed by David Ben Gurion, special relationships have been established between the state and the orthodox NGOs. The government approved the religious parties' demand to grant their sectarian orthodox education institutions a national status, eligible for all the support and benefits granted to the Israel national-secular education system. Consequently, over 41% of Israel NGOs are orthodox educational institutions fully supported by the state, while the percentage of the orthodox people in the population is less than 20%. This thorny political issue has not been resolved either by the NGO Law of 1980, or the various government regulations that followed in the 1990s.

While the Israel CSOs have always played an important socio-economic role in the Israeli society, their importance has grown in the past two decades. This is clearly exhibited in the following data: In 2000, NGOs employed 10.7% of the Israeli work force

(non-agricultural work force), and produced 12.8% of the Israeli gross domestic product (GDP). The Third Sector employs 30,000 paid employees and an unknown number of volunteers. Between 1991 and 1995, the total NGOs expenditures grew by 13.5% reaching 33 billion new Israeli Shekel (NIS) (over US \$8 billion), and the number of their employees grew by 19,000. On the other hand, the number of government employees decreased by 22%, from 66,631 to 51,995.^[2]

The Israel government initiated intense partnership and shared responsibilities with the CSOs, creating an environment of “mutual dependence.” The CS has become greatly dependent on public funding for providing most of the education, health, and welfare services, while the state became greatly dependent on the social services provided by the CS. The research shows that “A high rate of 78% (4705 organizations) of the 6030 active CSOs engage in service provision. Only 5.7% engage in advocacy, and 5.8% are foundations funding NGOs.^[2] The government has increased its support of the Israel CS, and in 2004, transferred 36 billion NIS to non-profit organizations compared with 33 billion NIS in 1999. The MOF reported that the funds have been transferred as block and formula grants, but only about 10 billion are included in the regular budget. The rest are mandatory payments for healthcare and education, jointly paid by national and local governments. The subcontracting of services, from NGOs, not mandated by law amounted to only 6% of all the funds (Tables 2 and 3).

MINORITY NGOs IN ISRAEL CS: JEWISH ORTHODOX AND ISRAELI ARAB NGOS

The data shows that the Jewish orthodox NGOs are almost exclusively engaged in providing services

Table 1 Government grants to NGOs by ministries 2002–2004^a

Department	2002	2003	2004
Prime Minister	46,863	47,905	305,348
Interior	4,697,832	3,963,079	3,839,894
Education	1,388,469	1,196,543	1,513,911
Religion	1,223,054	1,053,534	3,005,485
Welfare	160,425	123,001	98,25
Health	98,089	24,032	6,191
Absorbition	34,002	31,602	28,774
Agricultural education	150,709	119,367	116,468
Science, culture, and sport	438,812	125,570	1,354,638
Total (NIS)	8,238,255	6,684,633	10,180,534

Additional, smaller amounts were granted by various national and local government departments.

[From MOF data. www.app2.mof.gov.il/cgi-bin/ibi.cgi (accessed August 14, 2005).]

^aIn millions of NIS. Grants under clause 3[a] in the regular budget. Numbers do not include subcontracts and indirect support.

Table 2 Funding for NGOs by categories of operation^a

Category	% of the total organizations	% of total public funding
Religious organizations	24	41.3
Education and research	19	17.2
Culture and leisure	14	11.4
Recreation and welfare	12	9.4
Philanthropy	8	3.2
Civic, legal, and political	6	1.3
Housing and development	5	0.6
Health	3	6.2
Labor unions and chambers of commerce	3	4.2
Memoriam	2	1.0
Environment	1	1.5
International activities	1	0.4
Non-Jewish religious NGOs	1	0.4
Others	1	2.4
Total	100	100.5

The categorization of the NGOs follows a universally accepted model developed by Salamon, L.M.; Anheier, H.K.; List, R.; Toepler, S.; Sokolowski, W. *Global Civil Society*; John Hopkins Center for Civil Societies, 1999: 1) Business and professional; 2) Civic and advocacy; 3) Community empowerment; 4) Culture; 5) Education; 6) Environment; 7) Health; 8) International operation; 9) Philanthropy; 10) Relief and social services; 11) Religious and congregations; 12) Unions; and 13) Others (in Israel, memorial organizations).
^aRatio between the volume of the organizations' activities and funds received.

(97%) in three areas: 1) religious activities (60%), mainly operating religious facilities for the exercise of religious ceremonies, e.g., "Miquveh," a bathing facility for women and "Kolel," a Talmudic school for bright young married men; 2) education and research

(24%), including private orthodox day schools for boys and girls, and operating Sunday schools and after-school religious programs; and 3) welfare (16%), including operating soup kitchens, free grocery centers, free kindergartens with three meals per day, among others. Thus, it is not surprising that registration rate of Orthodox education NGOs is growing faster than registration for secular education institutions. However, although Orthodox NGOs constitute almost half of the total number of Israeli NGOs they employ only 25% of all-paid employees.^[3]

The burgeoning trend revealed in the past two decades in the Israel CS is also evident in the Israeli Arab NGOs. The data shows that 65% of all Arab NGOs were registered during 1990–1998. However, the differences between the two ethnic groups are greater than the similarities. Among the Arab NGOs, advocacy organizations are especially noticeable: 7.3% compared to 5.7% in the entire network of NGOs. The number of NGOs providing civil services is also high. The main areas of activity are religion (24.5%), culture and recreation (22.2%), welfare (18.9%), and housing and development (13.9%). The rest 5% are mainly family associations. In contrast, educational activity (8%) is lower compared to the average in the entire NGO community (16%). The rates of Arab NGOs receiving direct government funding are similar to the rates of active organizations in the total CS population (roughly 5%). However, in terms of indirect support (donor's tax deductibles), the number was half compared to that of the entire CS population: 2.5% compared with 5.6%.^[3]

Since the outbreak of the second Intifadah ("Uprising") in 2000, the Israeli Arab NGOs have often adopted a confrontational attitude, in particular criticizing the government's lack of respect for their human and civil rights. The confrontational attitude is a result of a mutual distrust that has developed between the Jewish and the Arab communities in Israel, a mistrust that has yet to be resolved.

Table 3 Government formula grants to NGOs (in billions of NIS)

1992	2.14
1993	2.33
1994	2.28
1995	2.6
1996	3.06
1997	3.56
1998	4.04
2002	8.3
2003	7.29
2004	6.19

(From Berlinski, U.; Sofer, P. Government Support of NGOs and Public Institutions; 2000; Also MOF data 2002; 2003; 2004. Data not available for 1999; 2000; 2001.)

REGULATING, CONTROLLING, AND OVERSEEING THE ISRAEL CS

The massive transfer of resources and responsibilities from the state to CSOs increased in the 1990s, and in the financial year (FY) 2000, transfer of funds to NGOs amounted to 40 billion NIS, or 4.5% of the national budget. However, about half, or 20 billion NIS were payments for medical services (Kupot Holim). To oversee and regulate this process, the Israel Knesset (Legislature) created in 1980, the Office of the Registrar of NGOs within the Ministry of the Interior. The creation of the Office of the Registrar was part of

a new NGO law that addressed mismanagement and lack of accountability in the mushrooming network of NGOs. In 2004, the Registrar's office executed 12,322 random audits of NGOs. In 523 NGOs the Registrar's audit was done by 13 CPA companies. In 403 NGOs the audit found irregularities.

In 1985, the Knesset broadened the government oversight scope by revising the existing Budget Law and adding regulatory measures to eliminate fraud in the funding applications. The revised law imposes strict guidelines and criteria on applicants for public support, including a requirement for all applicants to submit a standard written grant application form that includes the names of the NGO officers and Board of Directors; a review of the NGO's operations in the last two years, and a recent audit approved by a certified public accountant (CPA). The CPA audit has to include a review of the NGO's financial responsibility—proper accounting and budgetary procedures—and show that the NGO executed its operation efficiently and effectively. Accountability has been particularly essential in the process of “subsidiarity,” namely, when social services are “translated into a system whereby private provision of services takes precedent over public efforts.”^[4]

Under “subsidiarity,” the government and the NGOs enter into explicit, legal arrangements that spell out the division of labor, responsibilities, hierarchies, chain of command, accountability, and financial obligations of each of them. For instance, several CS institutions have been authorized to conduct national tests and to certify physicians, CPAs, and lawyers. Burial services are also carried out by non-government, non-profit organizations. In 1995, and later in 1996, the NGO Law was amended to ban registered NGOs from applying for public funding to more than one government ministry for the same activity. The enforcement of this regulation was given jointly to the Registrar's office and the MOF. The MOF is required to issue annual reports, available in print and on the Internet, enumerating in great detail the grants allocated and the contracts awarded to every NGO. Another important reform in the Budget law was passed by the Knesset in 1992, requiring all grants to NGOs to be included in the regular budget under clause 3[a], and the information should be made available to the public. However, following the principles established by the first Israeli government coalition, the law included two exemptions: one for the Orthodox Spharadi, and the other for the Orthodox Ashkenazi educational institutions. Grants for these two CS ultrareligious organizations are not included in the regular budget, thus public exposure of the grants is lacking. The revised Budget law includes strict guidelines for the distribution of funds to the Third Sector. These guidelines require

all ministries and local governments that award grants to use objective and equal criteria to all applicants, and that the criteria be made public.

In 1998, a new regulation was issued requiring every NGO applying for public funds to hold a certificate of “Proper Management,” issued by the Registrar. This unique requirement ensures not only the NGO's adherence to the law, but also its adherence to proper operational norms applicable to all government agencies and departments. These norms ban nepotism, require a reasonable ratio between overhead and operational expenditures, and ensure that unlawful distribution of profits will not be carried out in the form of excessive salaries and fringe benefits.

The “Benchmark of Proper Management” also expects the NGO to operate with utmost efficiency and effectiveness. Only holders of a Proper Management certificate can apply for and receive public support. In addition, the Registrar can issue a “supported organization status,” an incentive issued to NGOs for good behavior. However, the road to the execution of the various important reforms has been rocky, at best.

CRITICISM OF THE ISRAEL CSOs

Since the late 1980s, the Israel CS has encountered some harsh criticism. A special committee appointed in 2000 by the Office of the Prime Minister to investigate possible widespread fraud in the CS aired the most important criticism. The Committee's findings were summarized on 9 August, 2000, and concluded that the government control mechanisms were insufficient partly owing to: 1) bureaucratic shortcomings; 2) lack of coordination among the ministries that either award grants or purchased services from CS; and 3) legal shortcomings. These weaknesses result in: 1) unlawful duplication of grants (an issue that was also raised by the Israel SCO)^[5] and 2) the practice of inequality, namely, the lack of clear and objective criteria for the formula grants,^[6] and the arbitrary manner in which a myriad of agencies, departments, ministries, and local governments have been allocating public funds to unwarranted NGO requests. Legal difficulties were also included in the report. The report enumerated severe shortages of qualified personnel at the Registrar's office, lack of cooperation among the various ministries providing funding for NGOs, and lax control of the subsidizing ministries of the NGOs receiving their support.

The NGO law is vague and allows for inconsistent interpretations. The law incorporates clauses from both the Ottoman and the British rule. Finally, the Third Sector had used the impressive political leverage that it had acquired to circumvent the law by using the

legislator as a tool to issue “special releases” for particular sectors.

CONCLUSIONS

While the activities of Israel CSO can be viewed as intensive and innovative, they are by no means unique. Since the mid-1980s, the processes of delivering social services have changed radically. The transformation of the Third Sector into the main provider of public services is a global phenomenon. Similarly, the NPM approach has become a universal practice. In Israel, as in most developed countries, the Third Sector has become an important economic and social institution, consisting of over 30,000 registered NGOs and employing about 10% of the Israel the workforce. Because public (government) financing is an important source of funding for the CSOs, corruption and misuse of public funds have often been practiced. In the early 1980s, the government launched a major bureaucratic and legal campaign to fight fraud and the unlawful distribution of public funds. The struggle for accountability continues and the government has gained some major achievements: 1) The Knesset and the Government adopted regulations that enhanced the coordination and cooperation among the donor agencies; 2) regulations were adopted to eliminate duplication in grants; 3) restrictions have been imposed on NGOs' overhead expenses and the distribution of

excessive benefits to NGO executives; and 4) the Proper Management Benchmark coerces the CSOs to submit annual activity reports, including financial reports of operations in progress, and completed operations. While the Israeli government maintains the autonomy and independence of its CS, major strides have been made in regulating the CS and in achieving accountability and proper management.

REFERENCES

1. Lester, M. Salomon.; Helmut K. Anheier.; Regina, L.; Stefan, T.; Wojciech Sokolowski, S. *Global Civil Society: Dimensions of the Non Profit Sector*; The John Hopkins University Center for Civil Society Studies: Baltimore, MD, 1999; 8 pp.
2. Gidron, B.; Katz, H.; Bar, M. *The Third Sector in Israel*; Ben Gurion University Press: Beersheva, 2000; 10 (in Hebrew).
3. Gidron, B.; Katz, H.; Bar, M. Characteristics of the Israel civil society. The 30th Arnova Conference, Miami, FL, November, 2001; 24 pp.
4. Helmut, K. Anheier An elaborate network: profiling the third sector in Germany. In *Government and the Third Sector*; Gidron, B., Kramer, R.M., Salomon, L., Eds.; Jossey-Bass: San Francisco, 1992; 31–56.
5. *Annual Report of the Israel State Comptroller*, (“Me'vaker Hamedinah), Report No. 50b; 1998; 727–739.
6. Supreme Court ruling 780/83, quoted in Hartuch, A. de, 82. Also, Report of the State Comptroller, 1998; 728–739.

Israel: Peace and Conflict Resolution Organizations

Raviv Schwartz

Ben-Gurion, University of the Negev, Beer-Sheva, Israel

INTRODUCTION TO P/CROs: INTERNATIONAL PERSPECTIVE

In recent decades, we have witnessed the engagement of a host of international, national, and regional non-state actors seeking to resolve or at least contain the many forms of social, ethnic, religious, and political conflict. Most notable is the role played by non-governmental organizations (NGOs) in this effort, both domestically (e.g., peace movements) and internationally. Such NGO activity has also been evident in regions and countries experiencing protracted, violent conflict such as Israel/Palestine, Northern Ireland, South Africa, former Yugoslavia, etc. In societies subject to what Kreisberg terms *intractable conflict*, understood as “protracted, irreconcilable, violent, and self-perpetuating,”^[1] the identities distinguishing between the conflicted parties are ubiquitous and seemingly indelible and may eventually be “converted into something akin to ontological statuses.”^[2] It is therefore noteworthy that within these same societies, organizations [heretofore referred to as Peace/Conflict Resolution Organizations (P/CROs)] have emerged that reject or at least question the inevitability of both the conflicts and the contentious relations between the protagonists. In doing so, they often defy the prevailing sociopolitical climate in their respective societies.

HISTORICAL OVERVIEW OF ISRAELI CONTEXT

The Early Years

That P/CROs are expected to be active in the Israeli context is by no means a foregone conclusion, since issues of war and peace have traditionally been the sole purview of politicians and generals. In fact, during the first 25 years of Israel’s existence, peace activism was limited to the actions of particular individuals or small groups (rather than mass organizations), who had little effect on either public opinion or policy makers. However in recent decades, this monopoly has gradually eroded, with greater legitimacy afforded to other actors in the discourse around these issues. This shift is related to a more global trend of the reduced public confidence in state institutions.^[3] It is also a function of

the changes underway in Israel’s polity, according to which the public sector, which had initially been imbued with a strong collectivistic ethos, assumed a more open form, allowing for a more pluralistic model of relationships among its different groups and factions.^[4] This marked the transition during the late 1970s and early 1980s from what scholars of Israeli society term the Formative to the Pluralistic Era. It was during this period that the heterogeneous character of Israeli society began finding its expression in the institutional framework and the collectivist ideology gave way to a more individualistic one.^[5]

Another important factor in the loosening of the almost exclusive grip on the security discourse in the hands of the military and political elites was the enhanced sense of security following Israel’s victory in the Six Days War in 1967. The establishment of the nationalist Gush Emunim (Bloc of the Faithful) in the mid-1970s, a movement advocating the settling of Jews, in all the territories occupied by Israel since 1967 and of the dovish Peace Now in 1978 marked yet another milestone in the increased voice of citizens’ associations in the discourse on war and peace.^[6]

Throughout the 1980s, the primary mode of activity of Israeli P/CROs, of which Peace Now was the largest and most dominant, was that of protest—primarily in the form of large-scale demonstrations. From 1986, Israelis were prohibited by law from meeting with members/leaders of the PLO. Hence, dialogue and face-to-face encounters with “the other” was all but impossible—with the exception of encounters which invariably took place outside Israel/Palestine. The first *Intifada*, which commenced in December of 1987, only reinforced the disinclination to conduct any kind of face-to-face activity among Palestinians and Israelis, though a few dialogue initiatives did take place during the late 1980s, such as the “Rapprochement” group, but these were few and far between.^[7]

The Oslo Accords in 1993

In the years following the signing of the Oslo Accords (1993–2000), the dominant tactical form of Israeli P/CROs during the preceding years, i.e., protest, gave way to other modes of activity. The de-emphasizing of protest can be explained in part by the palpable sense on the part of many movement activists that the

government itself, led since 1992 by Labor, would take the lead in implementing the Oslo accords and bringing peace. Furthermore, with the thawing of relations between the Israeli and the Palestinian leadership, the stage was then set for much more extensive contact on a grassroots level between Israelis and Palestinians, referred to by many during this period as “people-to-people” activity. Not only were the macropolitical conditions ripe for the undertaking of such activity, funds from a variety of sources suddenly became available for projects along these lines. A host of private foundations, governments, international agencies, and others began allocating funds to Israeli P/CROs and other institutions that undertook a wide variety of projects grouped under the rubric of “people-to-people” activity. The rationale was that the formal peace agreements, embarked upon by the political elites from both sides, needed to be buttressed by civil society, i.e., the “bottom-up” support of citizens and institutions. During this same period, the Israeli peace movement underwent a process of institutionalization, according to which several P/CROs relied increasingly on professional staff rather than volunteers, assumed more complex organizational structures, and preoccupied themselves with the many administrative and programmatic obligations involved in the application for and receiving of funds and the concomitant financial reporting. This period was marked by the unmistakably enhanced role of foreign funders in the shape and scope of peace activism in Israel.

The 2nd Intifada—October 2000

The 2nd Intifada, which erupted in October 2000, resulted in a drastic reduction in the support among Israeli Jews for the already beleaguered peace process. During the period immediately after the outbreak of violence, demoralization among the ranks of Israel’s peace movement was palpable. The prevailing wisdom at the time, and reinforced through the media, was that the peace camp had been “shell-shocked”—paralyzed into a state of inaction. To be sure, the events that commenced in October 2000 were perceived by many in the field to have undermined the very possibility of peace and reconciliation. During this period then, P/CRO activity by and large ground to a halt and it was only after a year or so that some partial recuperation could be discerned. Those P/CROs that appeared less traumatized and debilitated by this crisis and its aftermath were those featuring a more radical ideological orientation, including some formed soon after the events of October 2000. Also evident during this period were (in most cases new) P/CROs that engaged in various activities designed to provide material assistance to and demonstrate solidarity with

Palestinians, as the collective predicament of the latter deteriorated considerably during this period. These include “Windows” and “Ta’ayush” to name just two.

DIVERSITY OF THE P/CRO COMMUNITY

One very salient feature of the environment of Israeli P/CROs is its diversity—both organizationally and ideologically. Attempts to classify the many types of P/CRO, for the purposes of analysis or for evaluation, have yielded a few typologies. One early effort arrived at the following three categories of P/CRO organizational forms: 1) service providers, which usually feature a specific service technology, serve clients/constituents, and often engage in capacity building in a variety of substantive areas (e.g., education and health); 2) “policy/development,” which typically enjoy a relatively high degree of access to decision makers, possess a particular expertise, interface regularly with negotiators, and often produce outputs such as position papers, etc.; 3) classic peace groups, a category of P/CROs that target ordinary citizens, policy makers, and the media through various public activities and seek to affect change at the level of mass consciousness.^[8] Another scheme developed classified P/CROs based on the nature of the activities conducted and arrived at three parallel categories: *peace building*, *peacemaking*, and *advocacy/protest*.^[9]

The variation in ideological orientations among Israeli P/CROs is also considerable. Until 1993, the dominant trend among P/CROs (particularly the larger more dominant ones) was consistent with the Zionist ethos; that is, they recognized Israel’s right to exist within secure borders as a state defining itself as “Jewish” and linked as such to the world Jewish community. These P/CROs have never adopted an explicitly antimilitarist or pacifist stance, as underscored by the fact that most peace activists serve in the Israeli army reserves.^[10] In recent years however, P/CROs that do not subscribe to Zionism, or may in fact oppose it altogether, have been established and are more visible. Ideological trends, such as anarchism and a blanket opposition to military conscription are presently an integral part of the landscape of P/CRO activity and have succeeded in mobilizing both the necessary material and human resources.

Another aspect of diversity finds expression in the social identity of membership/followers, whereby a number of P/CROs are organized around a particular identity marker or constituency. This phenomenon is mirrored in peace movements throughout the world, many of which also feature identity-specific mobilization.^[11–13] These include groups organized around religion (Orthodox Jews, clergy, ecumenical, etc.), gender (typically women’s groups, not all of which are

feminist in orientation), and professional background (physicians, social workers, retired military officers, and even planners). Yet another category which encompasses a few P/CROs is that which could best be described as “victim.” Included here are organizations formed by and for parents and family members of casualties of the conflict (i.e., terror victims and fallen soldiers). Interestingly, a P/CRO was established only recently composed not of victims of the conflict but of its perpetrators (i.e., Israelis and Palestinians who took part in violence as either soldiers or terrorists, respectively).

REGIONAL SIGNIFICANCE

For a variety of reasons, the phenomenon of P/CRO is not anywhere near as extensive in the Arab world as in Israel. Among the Palestinians as well, the bulk of politically oriented NGO activity focuses much more on the many grievances related to occupation and on the struggle for independence. Nevertheless, it is important to note that there are a number of P/CROs composed of both Israelis and Palestinians. In some cases, they are managed jointly and in others by two separate but parallel governance structures. A few of the P/CROs conforming to this pattern range from the large-scale initiatives of People’s Voice and Geneva to think-tanks such as Israel Palestine Center for Research and Information (IPCRI), to smaller groups with a more interpersonal, reconciliatory emphasis such as Parents’ Circle (bereaved parents) and Combatants for Peace.

CONCLUSION

There are many ways to attempt to gauge the impact of P/CROs. Firstly, looking at their aggregate effect, one must differentiate among the following targets: government policy; public opinion; and popular discourse. In a comprehensive study of Israeli P/CROs from 1967–1998, it was concluded that: “The aggregate efficacy of the Israeli peace movement seems to be quite low in terms of bringing about the desired political transformation in Israeli–Palestinian relations, e.g., the signing of a peace treaty. The efforts to introduce a dramatic change into Israeli public security also produced minor results.”^[14] These impressions are reinforced in a recent evaluation of peace-building projects undertaken primarily by P/CROs.^[9] One notable exception to the lack of P/CRO influence on policy is the experience of a single women’s protest group “Arba Imahot” (Four Mothers or Matriarchs). “Arba Imahot,” which advocated withdrawing Israeli troops from Lebanon. This P/CRO was deemed by many to have been very successful in both altering public opinion and prompting this desired change in policy. It

could be argued however that the success of this group could be traced to its narrow focus on a conflict that ultimately was not about either disputed territory or conflicting national narratives.

The impact of P/CROs does appear to have been more pronounced in the evolving discourse on matters of peace, security, and relations between Israelis and Palestinians. In addition, a strong case can be made for their having successfully challenged the prevailing view of the conflict as inevitable.^[14] Accordingly, the impact of P/CROs on policy can, at best, be conceived of as indirect. Ultimately, it would seem their most significant contribution is in the cultural, symbolic realm. However, in as much as the forces in this realm are brought to bear on public opinion and on government policy, they may be considered efficacious.

The impact of P/CRO activity may also be gauged at the level of those touched by it, i.e., participants in the many activities/projects undertaken. It must be noted that many of the efforts to formally and/or informally evaluate the impact of P/CRO activity are intended primarily for the consumption of those bodies funding the activity. These funders are typically drawn to activities that feature large numbers of participants and offer some opportunity for a larger visibility or “ripple effect”, rather than focusing on the potentially transformative nature of certain P/CRO activities at the microlevel of individual participation.^[15] What is very important to point out is that even during the zenith of P/CRO activity in Israel (1993–2000), the numbers of individuals touched by this enterprise was still quite small. It is difficult to gauge in numbers, but it is reasonable to assume that the percentage of the Israeli population touched directly or indirectly by the P/CRO enterprise never exceeded 10%.

Some have argued that the eruption of violence in 2000, at a time when P/CRO activity was at a height is effectively a confirmation of its lack of efficacy. The pacification of the peace movement during most of the first-half of the 1990s, i.e., their preference for reconciliation, dialogues, and joint projects at the expense of massive peace protest, rendered it ill-equipped to react when the peace process ground to a halt.^[10] Others claim that even if such activity proved unable to buttress a beleaguered peace process, it did nevertheless succeed in keeping open some lines of communication between parties who were transformed almost overnight from partners in a peace process to inveterate enemies in a seemingly timeless violent conflict.^[15]

REFERENCES

1. Bar-Tal, D. Societal beliefs in times of intractable conflict: the Israeli case. *J. Conflict Manag.* **1998**, 9 (1), 23–49.

2. MacDonald, M. The dominant communities and the costs of legitimacy. In *The Elusive Search for Peace*; Giliomee, H., Gagliano, J., Eds.; Oxford University Press: South Africa, 1990.
3. Hermann, T.; Ephraim, Y. When the policymaker and the public meet: Sharon, Israeli public opinion and the unilateral disengagement plan. *Palestine-Israel J.* **2005**, *11* (3/4), 93–100.
4. Gidron, B.; David, B. Self-help awareness in Israel: an expression of structural changes and expanding citizen participation. *J. Voluntary Action Res.* **1986**, *15* (2), 47–56.
5. Gidron, B. The evolution of Israel's third sector: the role of predominant ideology. *Voluntas* **1997**, *8* (1), 11–38.
6. Hermann, T.; David, N. Extra-parliamentarism in Israel: a comparative study of Gush Emunim and peace now. *Middle Eastern Studies* **1992**, *28* (3), 509–530.
7. Cohen, V. Interview. In *Voices from a "Promised Land;"* Rosenwasser, P., Ed.; Curbstone Press: Willimantic, CT, 1992.
8. Schwartz, R.; Lee, P. A preliminary stocktaking of Israeli NGOs engaged in Israeli-Palestinian people-to-people activity. In *Report of the Workshop in Israel-Palestinian Civil Society Comparative Activities*; KATU: Helsinki, 2000.
9. Kahanoff, M.; Walid, S.; Rami, N.; Yana, N. *Assessment of Cooperation Between Palestinian and Israeli NGOs—An evaluation study. Civil societies in dialogue*; UNESCO: Jerusalem, 2006.
10. Hermann, T. "People's diplomacy and people's vigilantism: Israeli grassroots activism 1993–2003. Paper presented at the Workshop on Implementing Peace Agreements: Israel/Palestine, South Africa and Northern Ireland, Van Leer Institute, Jerusalem, May 2–4, 2005.
11. Boulding, E. The early eighties peak of the peace movement. In *Peace Action in the Eighties: Social Science Perspectives*; Marullo, S., Lofland, J., Eds.; Rutgers University Press: New Brunswick, NJ, 1990.
12. Epstein, B. The politics of moral witness: religion and nonviolent direct action. In *Peace Action in the Eighties: Social Science Perspectives*; Marullo, S., Lofland, J., Eds.; University Press: New Brunswick, NJ, 1990.
13. Ruzza, C. Strategies in the Italian peace movement. *Res. Social Movements Conflicts Change* **1990**, *12*, 111–138.
14. Hermann, T. *Final Report: Israeli P/CROs 1967–1998*; International Study of Peace Organizations—ISPO, Aspen Institute: Tel Aviv (unpublished), 1998.
15. IPCRI—Israel Palestine Center for Research and Information. *YES PM: Years of Experience in Strategies for Peace Making*. Jerusalem, November 2002.

Israel: Public Law

Assaf Meydani

*School of Government and Society, The Academic College of Tel-Aviv Yaffo,
Tel-Aviv, Israel*

INTRODUCTION

Public law in Israel reflects the collection of norms that govern the legal relationships between the state and the individuals, and the relations among governmental agencies. Public law covers constitutional law and administrative law. It also incorporates socio-political aspects that infuse it with positive and normative content and play a role in shaping it. The main issues governed by public law include the protection of human rights, the regulation of administrative authorities, and the oversight and monitoring of how administrative authority is used, as well as the formal and material definition of constitutional norms, the legal status of government agencies, and the administrative process. Constitutional law is the body of law that governs the legal status of the basic norms in the Israeli legal system and defines the interrelationships and powers of government agencies and other state entities. The constitutional mechanism is designed around legal norms, each of which has a different normative status. A formal constitution is generally at the top of this pyramid. Unlike countries such as the United States and Canada that have coherent constitutional documents, Israel does not have a formal constitution. However, over the years, a system of basic laws has been enacted, which are designed to address specific constitutional issues. At some point in the future, they may be compiled into a complete constitutional document. The statutes that constitute the legislation entitled Basic Law: Human Dignity and Liberty, passed in 1995, are recognized as having constitutional status that takes precedence over regular law in Israeli judicial interpretation.^[1] The recognition of the basic laws as having constitutional status was dubbed the “constitutional revolution.”^[2] Thus, Israel now has constitutional norms that override normal statutes and provide the legal basis according to which the Supreme Court interprets Israeli legal provisions and conducts judicial review of primary legislation. The basic laws that have had the most significant impact are Basic Law: Human Dignity and Liberty and Basic Law: Freedom of Occupation.

ISRAELI CONSTITUTIONAL HISTORY

According to United Nations General Assembly Resolution 181 of November 29, 1947, Israel was to have its Constituent Assembly adopt a constitution. There has been significant legal debate about whether this resolution is still binding on Israel. Indeed, at the beginning of 1949, a Constituent Assembly was elected. However, the Assembly’s decision was not to ratify a constitution, but instead to become the legislative body, and gradually enact basic laws, which together would form the sections of a constitution at some point in the future. This endeavor has still not been completed. To date, the Knesset has passed 11 basic laws. Five basic laws define the activity of government agencies—Basic Law: The President of the State, Basic Law: The Knesset (amended twice), Basic Law: The Judiciary, and Basic Law: The State Comptroller. Three basic laws define national institutions and assets—Basic Law: Israel Lands, Basic Law: The State Economy, and Basic Law: The Army. One basic law defines the status of Jerusalem as the capital of Israel. Two basic laws—Basic Law: Freedom of Occupation (amended once) and Basic Law: Human Dignity and Liberty—define fundamental human rights and are at the heart of all other legislation. These basic laws assert that their purpose is to establish in law the values of the State of Israel as a Jewish and democratic state. By virtue of this act, those very values of the State of Israel as a Jewish and democratic state became the object of judicial interpretation. Moreover, both of these basic laws restrict the authority of the Knesset to legislate laws that violate the basic rights that they serve to protect. If a law that infringes on these rights is found to have a worthy purpose and is proportional in its infringement, it is deemed constitutional and does not violate the basic law.

Judicial Review of the Constitutionality of Laws: “The Constitutional Revolution”

Judicial review of the constitutionality of primary legislation was not explicitly defined in the basic laws,

and has developed in Israel on the basis of the Supreme Court's interpretation of the constitutional documents. It is through this process that rights such as the freedom of expression, the freedom of election, and the freedom of protest have been established.^[3] In the 1970s and 1980s, the Supreme Court began adopting a strategy of increased intervention in the activities of other government agencies in areas where it had heretofore not inserted itself.^[4] At the procedural level, the High Court of Justice began to accept cases without rigorous investigation of the doctrines of "standing" and "justiciability," which in the past had been the cornerstones in determining the Court's willingness to hear appeals.^[5] At the material-content level, in the 1980s and 1990s, the Supreme Court significantly expanded its level of intervention in the activities of administrative agencies. Thus, it ruled that the exemption from military service granted to yeshiva students was not legal. It disqualified interrogation methods used by the General Security Service and, through its intervention, allowed women to be inducted into the Air Force pilot training course. This intervention drew no small amount of criticism from the legal community and members of the public alike. The expression widely used in public and legal discourse to describe these developments is "activism." Among other things, these developments are the result of deep fissures in Israeli society, along with an institutional structure that creates a situation in which the political system finds it very difficult to cope with problems that require the formulation of public policy. The functional failures of the legislative and parliamentary branch lead to an increasing number of appeals to the judiciary branch.^[6]

The very fact that the Court has the authority to strike down laws is seen by some as dealing a blow to the supremacy of the legislative branch, despite the Knesset's ability to overrule the Court through legislation. However, since the legislation of Basic Law: Human Dignity and Liberty (1992) and Basic Law: Freedom of Occupation (1992), the Supreme Court has invalidated sections of no more than five laws.

Separation of Powers

A fundamental element of the Israeli system of government is the principle of the separation of powers. Over the years, Israel has undergone a change in terms of how the state is perceived by the Court, moving from a pure separation of powers to a model of checks and balances. In the latter model, the Supreme Court recognized its political power and understood that its legitimacy as a partner in collective social decisions in the State of Israel required it to play a number of roles including interventional monitoring of the affairs of other agencies. In this regard, the Court

was required to intervene in matters that deviated from the traditional pattern of separation of powers, for example, stripping Knesset members of their immunity, overseeing regulations, reviewing political agreements, validating political appointments, and allowing government ministers with indictments pending against them to retain their positions.^[7]

The Israeli Judicial Branch

The judicial branch in Israel has the status of first among equals relative to the other branches of government. Thus, for example, it enjoys a substantial degree of independence from political influence. Judges in the Supreme Court are appointed by the Judicial Selection Committee, composed of representatives of all three branches of government, ensuring a professional and apolitical selection process.^[8]

The Israeli Supreme Court serves a wide variety of functions. First of all, it is the highest court of appeals in the Israeli legal system. Furthermore, it is the first venue in which disputes between the individuals and the State are heard. This second role has caused the Supreme Court to be viewed by the general public as the watchdog over the rule of law and the champion in the fight against corruption and the protection of human and civil rights. To reduce the heavy burden of fulfilling this role, a reform has been enacted in which cases brought by individuals against the state will now be heard first by lower courts.

A JEWISH AND DEMOCRATIC STATE

The Israeli public has been grappling with the issue of the Jewish character of Israel as a Jewish and democratic state since the founding of the state. The State of Israel's unique character, as well as the history preceding its establishment, testifies to the explicit desire of its founders to establish a Jewish nation with a Jewish character and to make the State of Israel the homeland of the Jewish people. This desire is reflected in the selection of Hebrew as the official language of the country, the recognition of the Jewish holidays and Jewish symbols as the official national symbols, and the establishment of religious services and immigration legislation that show a decisive preference for Jews.^[9] That said, no law has been passed stipulating who is a Jew, and the formal definition recognizes a Jew as an individual who is born to a Jewish mother or an individual who has converted. It does not specifically define the type of conversion, whether in accordance with Jewish law, nor does it state whether conversion is possible when conducted under auspices that are not Orthodox. While Israel is the national

homeland of the Jewish people, it does not exclude others from becoming citizens; it is, under its laws and founding principles and despite the hostility, war and terror to which it is subject, a state that is a land of equal opportunity for all of its Arab citizens. Thus, for example, the Court has ordered that the writing on street signs must also appear in Arabic and ruled that Arabs and Jews should be treated equally when purchasing residential property.^[10]

Administrative Law

In addition to the provisions of constitutional law, public law also encompasses the provisions of administrative law. This legal field regulates the activities of all governmental administrative agencies. Administrative law deals with the structure of the administration and its authorities, the proper conduct of the administrative process, the amount of discretion the authority is allowed, the oversight of the administrative agency, and the degree of freedom of activity the administrative agency has under private law. At the heart of administrative law is the concept that public administration functions as a public trustee and therefore is expected to behave according to higher standards than those applicable to players under private law.^[11]

THE RULE OF LAW AND PRINCIPLE OF LEGALITY: CONCLUSION

The administrative authorities in Israel generally have a decentralized, hierarchical structure. Thus, for example, there are central administrative authorities (such as the Ministry of the Interior) that have offices around the country (local authorities). Similarly, the different administrative jurisdictions embody various types of authority, among them, quasi-judicial, legislative, and executive authorities. The guideline directing the activity of administrative agencies in Israel is the principle of legality. The principle stipulates that, as opposed to a private individual who is allowed to do anything that is not explicitly prohibited, administrative agencies begin at a higher standard. Implementation of the principle of legality with respect to the administrative agencies assumes that everything that is not explicitly allowed is prohibited. In other words, the activity of the administrative agency must be stipulated explicitly in the law. Furthermore, administrative agencies are subject to the doctrine of natural justice,^[12] which requires fair proceedings in disputes with citizens, including providing a party injured by their action with the right to be heard and to

have his or her case reviewed. Administrative agencies are also subject to the test of reason and proportionality stemming from the constitutional philosophy in Israel.^[13] Thus, the administrative agency must determine clear, general standards for legislation, and is prohibited from enacting legislation that is supposed to apply to an undefined group of people. In addition, the agency must regulate itself internally and create stable legal norms and decisions that must be published and applied from the point of their enactment.

REFERENCES

1. CA 6821/93 United Mizrahi Bank v. Migdal Cooperative Village. In *Decisions*; Vol. 49, section 4, p. 221 (Hebrew).
2. Barak, A. The constitutional revolution: protected human rights. *Mishpat Umimshal* **1993**, 1 (1), 9–10 (Hebrew).
3. Kretzmer, D. The path to judicial review in human rights cases: from Bergman and Kol Ha'am to Mizrahi Bank. *Mishpatim* **1997**, 28, 359–385 (Hebrew).
4. Mautner, M. *The Decline of Formalism and the Rise of Values in Israeli Law*; Ma'agalei Da'at: Tel Aviv, 1993 (Hebrew).
5. Gavison, R.; Kremnitzer, M.; Dotan, Y. *Judicial Activism, For and Against: The Role of the High Court of Justice in Israeli Society*; Yedioth Aharonoth Publishing—Sifrei Hemed and Magnes Press, 2000; pp. 11–13 (Hebrew).
6. Barzilai, G. Courts as hegemonic institutions and social change. *Politics* **1998**, 3, 31–51 (Hebrew); Hofnung, M. Authority, influence and separation of powers—judicial review in Israel in comparative perspective. *Mishpatim* **1997**, 28, 211–238 (Hebrew).
7. Claude, K. Separation of powers in a democracy. In *Basic Issues in Israeli Democracy*; Cohen-Almagor, R., Ed.; Sifriat Poalim: Tel Aviv, 1999; pp. 111–128 (Hebrew).
8. Shetreet, S. *On Adjudication – The Justice System in Law*; Yedioth Aharonoth Publishing, 2004; pp. 85–86 (Hebrew).
9. Rubinstein, A.; Medina, B. *Constitutional Law of the State of Israel*; Schocken Press, 2005; Vol. 1, pp. 318–412 (Hebrew).
10. HCJ 4112/99 Adalah, The Legal Center for the Rights of the Arab Minority in Israel v. Municipality of Tel-Aviv-Jaffa (not yet published); HCJ 6698/95 Qa'dan v. Israel Land Administration et al., PD 54 (1), 258.
11. Zamir, Y. *Administrative Authority*; Nevo Publishing Co., 1996; Vol. 1, pp. 35–36 (Hebrew).
12. HCJ 5848/99 Paritzky, v. District Planning and Building Committee, Jerusalem, 54 (3), PD 5.
13. Har-Zahav, R. *Israeli Administrative Law*. Shenhav Publishers, 1996; HCJ 6406/00 Bezeq v. Minister of Communication, PD 58 (1), 433.

Israel: Spatial Policy

Arie Hershcovich

Department of Political Science, Yezreel Valley College, Emek Yezreel, Israel

INTRODUCTION

One might expect that in a densely populated country, with a high birthrate and a commitment to accept Jewish immigration, spatial policy will be very high on the political agenda. In fact until lately it was shaped within rather closed group of professionals working for governmental institutions, quazi-governmental organizations, interest groups, and the business sector. Those professionals shared a common commitment to the Zionist ideology. That commitment enabled them to pursue common goals, although given only vague directions from the government, which is officially the highest planning authority.^[1,2]

The spatial policy formation system in Israel can be divided into three parts: regulatory, initiatory, and inhibitory. The regulatory part employs zoning—the definition of land use as its main tool. The zones are determined by outline plans, according to the Planning and Building Law 1965. An initiatory system operates alongside the regulatory part. Until the 1990's, it was comprised primarily of “institutional entrepreneurs,” among them government ministries, the Jewish Agency, the Israel Land Administration, and local authorities. The inhibitory part advocates conservation and environmental considerations. It too was comprised mostly by governmental and quasi-governmental organizations. All three parts shared a common commitment to the Zionist ideology, which led to a spatial policy that was intended mainly to disperse the Jewish population of Israel, as a means of confirming the states control of the space.^[2,3] More recently, the balance of the initiatory system shifted toward the private sector, while the inhibitory side has been joined by organizations working to safeguard the territorial rights of the Arabs in Israel, to preserve cultural and historical values, and promote ecological considerations. The rise of these groups along with the decline of the Jewish Agency and less government interventions are manifested in a change of paradigm in the Israeli spatial policy, from population dispersal to a more concentrated, compact, efficient, and environmental friendly policy (Fig. 1).^[4]

THE INSTITUTIONS

Most of the institutions involved with shaping the spatial policy in Israel are part of or affiliated to the

Ministry of the Interior, thus creating a rather closed and elitist system of policy formation. In most cases their constitution and authority are defined by Planning and Building Law 1965, which replaced the British Mandate's Town and Country Planning Act 1936. At the top of the pyramid is the *Government of Israel*, which is authorized to approve national outline plans.

Israel Land Administration

The Israel Land Administration (ILA) was established in 1960 by a special law. It is responsible for managing the “Nation's Land,” which comprise about 93% of the land in Israel (about 15% of it belongs to the National Jewish Fund and the rest to the state of Israel). According to law, the “Nation's Land” cannot be sold to anybody but the state of Israel. It can be only leased through the ILA. Therefore, the ILA has almost complete control over the location, the amount and the price of land available for development.^[6,7] Having a disproportionate representation of the rural sector on its board, the ILA policy is traditionally biased toward the periphery and the agricultural sector. Lately, the ILA started a careful process of privatization by granting full ownership to some of the leaseholders.

The National Planning and Building Council

The National Planning and Building Council prepare national outline plans for government approval. It also approves regional outline plans, advises the government regarding national planning policy, and rules on appeals at the regional and local levels. The Council is chaired by the Minister of the Interior and has 32 members: one-third is government representatives, one-third represent local authorities, and one-third is professionals and representatives of interest groups (Sections 2–6).^[8,9] The National Council's major areas of operations include population dispersal, establishment and expansion of settlements, and planning of national-level infrastructure. Because these areas involve the interests of numerous parties,^[2] the planning process is “lively and competitive... a political process that determines policy results.”^[10]

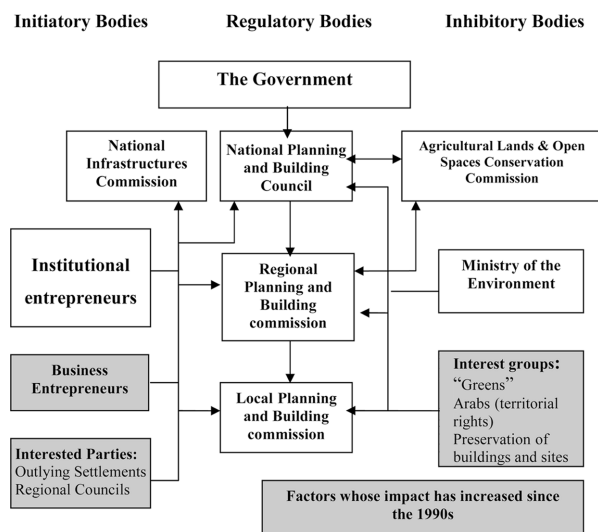


Fig. 1 The Israeli planning system. [From Ref.^[5] (this chart is an updated version of the source).]

The Regional Planning and Building Commission

Israel has six regional planning and building commissions: Northern, Haifa, Tel Aviv, Central, Jerusalem, and Southern. These commissions prepare the regional outline plans and submit them to the National Council for approval.

The Local Planning and Building Commission

The local planning and building commissions are responsible for preparing and publicizing local outline plans, hearing objections, and approving plans. Nevertheless, in most cases a plan still requires the approval of the regional commission. The local commissions issue building permits, collect improvement levies (levies imposed on increases to land value due to planning), and pay compensation for decreases to land value due to planning (Sections 2–6).^[8,9]

The Agricultural Lands and Open Spaces Conservation Commission

All the land in Israel was declared agricultural by default. The Agricultural Lands Conservation Commission alone has the authority to convert the status of land so it can be built on (First Amendment, Section 156B).^[8,9,11]

The Coastal Water Commission

This commission is authorized to approve any plan related to Israel's coastal waters (the 12-mile wide strip of water adjacent to the coast) (Second Amendment, Section 156B).^[8,9,11]

The National Infrastructures Commission

The National Infrastructures Commission was established in 2002 in order to expedite the planning process of national infrastructure. It has all the authorities and responsibilities of the National Council with respect to national infrastructures. A project is defined as a national infrastructure project by a committee of ministers. Such definition makes the project exempt from the regular planning process.^[8,12]

The Planning Administration of the Ministry of the Interior

The Planning Administration is the executive branch of the Minister of the Interior in his capacity as supervisor of the Planning and Building Law 1965. The administration's major functions involve: 1) formulating the national spatial planning policy and 2) professionally advising and running the planning institutions on all levels. According to the law every local outline plan has to be submitted to the Minister of the Interior, which is to the Planning Administration, for inspection.^[8,12]

The government of Israel is able, through these institutions, to control both the land supply and its use. This gives it a formidable capability of forming and implementing spatial policy, which is expressed mainly by the spatial plans.

THE PLANS

In accordance with Section 145 of the Planning and Building Law 1965, every action involving land requires a permit. Such a permit will be granted if the action corresponds with all the outline plans applicable to that land area. Therefore, the outline plans are the most effective means of spatial policy. These plans are ranked hierarchically, from detailed plans through local, regional, and national plans. A plan delimits a given land area within an outline (The Blue Line) and defines the land uses within this area. This action is known as *zoning—defining land uses*. The plan is comprised of a map (or “drawing” as set out in the law, Section 83) and regulations. The drawing must show the land uses, for example residential, commercial, roads, public buildings, open public areas, etc.

The regulations include directives for the various uses and quantity allocations.^[9]

Local Outline Plan^a

The local outline plan's directives are physical only, and concern land usage, road demarcation, spaces, density, etc. (Sections 61–69).^[8,9] The preparation of a local outline plan can involve disputes with many citizens who object to the plan for a variety of reasons. Hence, in most places in Israel the local outline plans are very old, and many do not have outline plans at all. The local commissions cope with this problem in two ways: 1) through partial outline plans, which enables the commission to deal with a smaller number of objections, though their ability to take overall issues into consideration is damaged and 2) By means of “shadow plans,” i.e., master plans (with no legal status) that the commission relates to as if they had validity in order to approve building plans. Because such plans are not official outline plans, they are not publicized in order for people to be able to submit their objections.

Regional Outline Plan

The purpose of the regional outline plan is to set out the requirements for implementing the national outline plan within the region. While the local outline plan relates only to the physical conditions, the regional plan also takes into consideration the “proper conditions for security and employment” (Section 55).^[9]

National Outline Plan

According to the Town and Country Planning Act 1936, there were only local outline plans. The Planning and Building Law 1965 specified also national outline plan (NOP) and regional outline plan for a number of reasons: 1) the small space of Israel, requiring optimization of use; 2) strong competition over land and space; 3) a great deal of government intervention in social and economic areas; and 4) significant emphasis on national considerations.^[10,11,13]

Most of the national plans are geared to a particular topic or area, such as roads, the Ben-Gurion airport, cemeteries, beaches, etc. Three of the plans are comprehensive plans: NOP #6 for Geographic Distribution of the Population (TAMA 6); NOP #31 for Construction,

Development, and Immigrant Absorption, (TAMA 31); and NOP #35 (TAMA 35), which replaced TAMA 31.^[8]

Plan Preparation and Approval

Local outline plans can be prepared by any party, whether private or public and then submitted to the local or regional commission. Regional plans are prepared by the regional planning commission. After a local or regional outline plan is prepared it is publicized in order to enable the public to examine it and submit objections. Preparation of a national plan is initiated by the National Planning and Building Council. The approval of a plan will be next higher institute. NOPs have to be approved by the government of Israel (Sections 77–119).^[8,9]

NATIONAL-LEVEL SPATIAL POLICY

Physical Planning for Israel (The Sharon Plan) 1951

Though not sanctioned by law, “The Sharon Plan” to a large extent has shaped Israel's spatial character. The goals of this plan included establishing a Jewish presence in all areas of the State of Israel and accelerated development.^[14,15] At the core of the Sharon Plan are “The New Towns,” which were built in the periphery for the large number of Jewish immigrants. “The Sharon Plan” was implemented almost in its entirety, at a tremendous cost. The price was paid by the immigrants who were distanced from the countries economic and social centers, and the Arabs whose settlements were destroyed in order to make place for the new towns.^[2,16]

NOP #6 for Geographic Dispersal of the Population (TAMA 6) 1974

TAMA 6 objective were defined as follows: “balanced distribution of the population and increased efforts in populating the Galilee, the Negev, and the other sparsely populated areas of the country, while increasing the Jewish population of these regions in order to preserve the current population ratio between the Jewish population and the minority [Arabs] population in the Galilee region” (on pages 50–51 of Ref.^[17]).

NOP #31 for Construction, Development, and Immigrant Absorption (TAMA 31) 1991

While the official policy is still one of population dispersal, TAMA 31 notes the contradiction between

^aLocal outline plan is sometimes referred to as a city building plan, as set out in the Town and Country Planning Act 1936, which preceded the Planning and Building Law 1965.

the desire to distribute the population and the need to find employment for the large number of immigrants expected to arrive at the beginning of the 1990s, because most jobs were located primarily at the center of the country. Accordingly, it proposes concentrating development in three metropolitan regions: north (Haifa), center (greater Tel-Aviv and Jerusalem), and south (Beer Sheva).^[18] TAMA 31 also presents a new attitude toward the Arab population. While the previous plans ignored the Arabs or saw them as a barrier to achieving their demographic goals, TAMA 31 includes among its objectives “ensuring conditions for equal development of the Arab population of the State of Israel”(on page 2.2 of Ref.^[18]).

Israel 2020—Master Plan for Israel for the 21st Century 1996

Israel 2020 plan has no legal status, but it has had an impact on planning perceptions in Israel. Its major recommendation is to adopt a policy of *centralized dispersal* according to which the population will be concentrated in three multifocal metropolitan areas (urbanized spaces) in the north, the center, and the south of the country. Israel 2020 plan designates three strategic spatial goals: 1) Efficient land exploitation: because Israel is a crowded country with a high natural rate of increase. 2) Enhance economic development: generating benefits of size (agglomeration) through transportation and communication infrastructures. 3) Adapting to peace: the social and economic gaps between Israel and the Arab countries necessitate developing the border areas in order to maximize the benefits for both parties and minimize the danger of antagonism.^[19]

NOP #35 Erez (Land) (TAMA 35) 2005

TAMA 35 is called “Land” against the background of the three planning dimensions in “The Sharon Plan”: land, people, and time. Nevertheless, its basic assumptions, its goals, and the means it employs are totally different from those in “The Sharon Plan.” It asserts that in order to be able to supply the amount of built and open spaces suitable for a developed country, land in Israel must be used more efficiently. Therefore, TAMA 35 recommends “to consolidate the urbanized areas into four^b metropolitan areas and to emphasize the principle of a continuum of open areas” and “to limit suburbanization.”

^bWhile TAMA 31 and 2020 plan count Tel-Aviv and Jerusalem as one metropolitan area, TAMA 35 counts them separately.

TAMA 35 introduces the “language of textures.” A “texture” refers to an area defined in the plan in which a particular type of development is permitted, ranging from urban texture (where major development efforts will be directed. Urban development will be compact . . .) to combined-reserved development (with the goal of creating a continuum of natural resources, landscape, settlement, and tradition . . .). In total, there are six such textures.^[20]

THE USE OF SPATIAL POLICY TO ACHIEVE NATIONAL GOALS

The spatial policy is perceived by the Israeli governments as a central means of promoting national goals.^[21] Until the 1990s, reassuring the control of the land by the Jewish majority was given first priority and was seen as part of the national security policy. Human rights, social fairness, economic development, and environmental considerations were subordinated to it. The association with security matters and the potential contradiction with human rights led to an elitist pattern of policy formation, with minimal public participation. This resulted in an unbalanced allocation of spatial resources. Those who have benefited were mainly the veteran Israelis (the founders), at the expense of the new immigrants, mostly Jews from the Arab countries, who were relegated to the peripheral areas, and of the Israeli Arabs (the indigenous natives), whose towns were destroyed and whose lands were expropriated.^[2,21,22] Recently, a gradual change can be detected, and more attempts are being made to involve the public in the policy formation process.^[2,4]

CONCLUSION—A NEW SPATIAL POLICY FOR ISRAEL

The 2020 Master Plan and the TAMA 35 Plan mark the inauguration of a new spatial policy for Israel. National security considerations are being gradually replaced by environmental considerations and an awareness to different needs determined by gender, ethnic group, socioeconomic status, and life style.^[23,24] In these aspects the spatial policy in Israel is following trends observed in other countries. In fact, the central role of the professional elite in shaping the Israeli spatial policy makes it more susceptible to policy transfer and diffusion in the global era.^[22–26]

REFERENCES

1. Alterman, R. National-level planning in Israel: walking the tightrope between government control and

- privatisation. In *National-Level Planning in Democratic Countries*; Alterman, R., Ed.; Liverpool, UK: Liverpool University Press, 2001; 259–264.
2. Hershovich, A. The Influence of Interest Groups on Plans for Dispersing the Population in Israel. Ph.D. thesis, Israel Institute of Technology: Technion, 1999 (Hebrew).
 3. Hill, M. Israeli planning in an age of turbulence. In *Planning in Turbulence*; Morley, D., Shachar, A., Eds.; The Magnes Press: Jerusalem, 1986.
 4. Hershovich, A. From the Sharon Plan to TAMA 35: the ideology behind three Israeli national plans. In *Horizons in Geography* 67, 2006; pp. 31–40 (Hebrew).
 5. Law-Yone, H.; Gertel, S. *Participation Ideologies in Israeli Planning*; Center for Urban and Regional Studies: Haifa, 1991.
 6. Forman, G.; Kedar, A. From Arab Land to “Israel Land:” The legal dispossession of the Palestinians displaced by Israel in the wake of 1948. *Environment Planning D: Soc. Space* 2003, 22 (6), 809–830.
 7. <http://www.mmi.gov.il>.
 8. Heiman, B. *Planning and Building Law—A Guide for the Public*; New Israel Fund’s Empowerment and Training Center: Shatil, 2002 (Hebrew).
 9. *Planning and Building Law 1965* (Hebrew).
 10. Torgovnik, E. *The Politics of Urban Planning Policy*; Lanham, MD: University Press of America, 1990.
 11. Shlain, E.; Feitelson, E. *The Formation, Institutionalization and the Decline of Farmland Protection in Israel*; The Floersheimer Institute for Policy Studies: Jerusalem, 1996 (Hebrew).
 12. <http://www.moin.gov.il>.
 13. Shahar, A. Planning environment. In *TAMA 35, National Plan # 35*; Ministry of the Interior: Jerusalem, 1999 (Hebrew).
 14. Ministry of the Interior. *Physical Planning for Israel*; Jerusalem: Government Publication, 1951 (Hebrew).
 15. Reichman, S.; Amit, I.; Golan, A. The national plans for population dispersal, In *Territorial Experiments—Readings on the Geography of Settlement in Eretz [land of] Israel*, Tel Aviv: The Open University, 1993 (internal edition) (Hebrew).
 16. Efrat, Z. Mold. In *Constructing a Sense of Place*; Yacobi, H., Ed.; London, UK: Ashgate, 2004; 76–88.
 17. TAMA 6. *National Plan #6, for Geographic Distribution of the Population of Israel*; Ministry of the Interior: Jerusalem, 1974 (Hebrew).
 18. TAMA 31. *National Plan #31, for Construction, Development and Immigrant Absorption*; Edna and Rafael Lerman: Tel Aviv, 1991 (Hebrew).
 19. Mazor, A.; Sofer, M. *The Future’s Picture: A Plan for Organizing the National Space—Israel 2020*; Technion—The Technological Institute of Israel: Haifa, 1996 (Hebrew).
 20. TAMA 35. *National Plan # 35—Erez*; Ministry of the Interior: Jerusalem, 2005 (Hebrew).
 21. Yiftachel, O. Planning and social control: exploring the “Dark Side.” *J. Planning Literature* 1998, 12 (2), 395–406.
 22. Tzfadia, E.; Yiftachel, O. Nation, space and capital: immigrants in Israel and socio-spatial stratification. In *The Ruling Capital*; Filk, D., Ram, O., Eds.; Van Leer Institute: Jerusalem, Hakibbutz Hameuchad, 2004; 197–221 (Hebrew).
 23. Fenster, T. *The Global City and the Holy City*; Prentice Hall: Pearson, 2004.
 24. Fenster, T. Identity issues & local governance: women’s everyday life in the city. *Social Identities* 2005, 11 (1), 23–39.
 25. Alterman, R. National-level planning in Israel: walking the tightrope between government control and privatisation. In *National-Level Planning in Democratic Countries*; Alterman, R., Ed.; Liverpool, UK: Liverpool University Press, 2001; 273 pp.
 26. Levi-Faur, D.; Vigoda-Gadot, E. The international transfer and diffusion of policy and management innovations: some characteristic of a new order in the making. In *International public Policy and Management, Policy Learning Beyond Regional, Cultural, and Political Boundaries*; Levi-Faur, D., Vigoda-Gadot, E., Eds.; Marcel Dekker, 2004; 1–24.

Job Classification

Lyn Holley

Department of Gerontology, University of Nebraska at Omaha, Omaha, Nebraska, U.S.A.

INTRODUCTION

This article begins with a summary description of job classification, then situates job classification within a context of organizing to accomplish work, and discusses the management functions supported by job classification and the roles in classification played by executives, managers, and job classifiers. It discusses “scientific classification,” the impact of different frames of reference (i.e., different employers and different criterion measures) on classification, and describes various job classification techniques. It examines the dual role of job classification as a vehicle for both management and public accountability. Finally, it concludes that job classification codes are vital information for management of work in a wide range of circumstances from stabilizing the civil services of African and Central American nations to maintaining accountability of the “new” public management workforce.

DEFINITION

Job classification is the process of assigning codes to jobs. The code of a particular job represents the type and level of work assigned. Its classification code locates the work of a job in relation to the work of the other jobs within a defined frame of reference. Jobs that are considered to be equivalent in type or level of work assigned have the same classification code. The classification code of a job links the job to a pay range, qualifications needed to do the job, and many administrative requirements and laws.

One example of a classification code is “Counseling Psychologist, (General Schedule) GS-0180-13.” In the U.S. federal government frame of reference, many jobs have this classification code. Jobs with the same classification code are considered equivalent in respect to the type and level of work assigned to the job. Jobs with the same code have the same minimum qualification requirements and are in the same pay range.

Different levels of work have different codes. For example, the job classification code, “Counseling Psychologist, GS-0180-09” indicates a job at a lower level of work than “Counseling Psychologist, GS-0180-13.” A job at level “09” has a lower pay range than a job at

level “13.” A level of work, such as level “09” or level “13,” may cut across many different occupations. For example, the “Counseling Psychologist, GS-0180-13” code is different from the “Information Technology Specialist (Systems Administration), GS-2210-13” code, even though “13” indicates the same level of work in both occupations.

Job classification is used in the United States by the federal government, all state governments, and virtually all local governments.^[1] Large- and medium-size private organizations use job classification. Different employers use different classification systems. The same employer may use different job classification systems for different groups of jobs; for example, a different system may be used for blue-collar (trades) jobs than for white-collar (professional and managerial) jobs.

Government job classification systems tend to be more detailed than private sector job classification systems. Many types or levels of government work have no effective counterpart in the private sector—public rule making and law enforcement are two examples. Pay for these jobs cannot be compared directly with pay in the private sector labor market. Setting pay levels for these jobs in relation to the private sector labor market is achieved by establishing the relationship of jobs that have no private sector counterpart to other jobs in government that do have private sector counterparts.

Despite its widespread application, job classification is not widely understood. Most managers and employees are aware of job classification only as a seemingly arbitrary limitation on pay or staffing flexibility.

ORGANIZING WORK

Job classification is critical to organizing work on a large scale. To plan, coordinate, or direct work, it is necessary to have some way to estimate and track the type and amount of human resources needed to produce the desired results. Public officials or contractors who manage people need to identify the knowledge, skills, and abilities required to produce desired results, recognize persons who have those attributes, and communicate the type of work and pay of each job.

Job classification is critical to demonstrating fairness. The realities of the relevant labor market, law,

and employees' perceptions all pressure managers to pay individual employees appropriately in relation to other employees. Job classification codes permit comparison of pay for similar types and levels of work.

Job classification is a key element in public accountability. Agency heads and legislators or boards are expected to ensure that amounts paid to individuals for work are justified and defensible. Job classification codes are fundamental to recording and comparing the amounts paid for similar types and levels of work by different organization units.

Job classification systems consist of classification codes and rules for how to match work and jobs with classification codes. Classification rules govern how jobs are sorted by type of work (e.g., as secretary or nurse) and level of work (e.g., as trainee, journeyman, or expert). The codes label each type of work and level of work. Different job classification systems have different sets of rules and codes.

One of the results of applying a job classification system is an outline of the work of an organization. The traditional organization chart shows job titles and numbers of workers for each title in each organization unit. Job classification is necessary to assign titles to jobs systematically, so that jobs with the same job titles are actually similar. Job titles allow organizations to interface in a general way with the labor market. The U.S. government supports economic development, record keeping, and workforce research by providing systems of standard job titles and occupational codes for hundreds of different types of work.^a

Job titles are one level of job classification "code." The more precise the "coding" of work into types and levels, the more comprehensive the information on the organization chart about each job. For example, the classification code "Counseling Psychologist, GS-0180-13" includes much information in addition to the job title. It links all the jobs with that code to pay tables and directives about pay, in this case, to pay range 13 in the GS. The classification code also links the job to specific classification standards. The classification standards include information about the nature of the Counseling Psychologist occupation (code 0180), and explain how jobs are classified in the 0180 occupation at the 13 level. The classification code links the job to standards, studies, and directives

about qualifications needed to carry out the work, and to specific procedures for hiring or selecting people to fill the job. The classification code also links the job to laws, regulations, or handbooks outlining other administrative requirements, such as eligibility of the incumbent of the job for various types of overtime or for inclusion in the union bargaining unit. Finally, the classification code links the job to employer and employee records of work, and records of staffing or other personnel decisions.

Job classification is not needed in very small organizations where the mission statement is often adequate for communicating about work. Division of work within the organization (e.g., a team of research scientists, a family farm) can be fluid and informal. The "job" of each person in the organization can emerge from the requirements of each new day. However, if the family farmers seek to hire an equipment operator, they will have to define a "job" to interface with the labor market, locate people with the skills needed, identify appropriate pay, and communicate with applicants about duties to be assigned and work expected.

Larger organizations have more interchange with the external labor market. In larger organizations, relationships are more contractual than familial, and require more formal record keeping. In larger organizations, different laws and regulations apply. Larger organizations are characterized by division and specialization of labor, and require some means for summarizing and combining information about many jobs. Job classification meets these and other needs of larger organizations.

FUNCTIONS SUPPORTED BY JOB CLASSIFICATION

Job classification is a function of the management of work carried out by people. Job classification facilitates movement of people to and among jobs. As it is used in the United States, job classification helps employers to find people and people to find employers on the basis of the work of the job, rather than on another basis such as kinship, ethnicity, or propinquity. Each paragraph in this section describes a management function to which job classification is fundamental.

Estimating, Tracking, and Fulfilling Human Resources Requirements

The concept of "job" is the basis for estimating human resources costs and requirements. "Job" is a unit of analysis for work; It is used to estimate how many of what kind of people are needed to achieve desired

^aThe Standard Occupational Classification System (SOC) classifies occupations on the basis of work performed and required skills, education, and training or credentials. Each occupation is assigned to only one group at the lowest level of the occupation. The SOC issued in Fall 1998 blends previous industry- or work-based classification systems (e.g., the Standard Industrial Classification System) with previous skills and worker attributes-based systems (e.g., the Dictionary of Occupational Titles).

results and to specify how many of what kind of people should be hired or let go. A “job” exists apart from its incumbent.

Identifying the type and level of work in a job is essential to accessing a work-based internal or external labor market. Job titles and further classification codes for type and level of work link the work of the job to qualifications required, and suggest the type and level of work assigned. Job descriptions communicate the work of the job in more detail.

Managing and Planning Work

Distribution of work is guided by job classification. Job descriptions record decisions made about the typical flow of work in an organization, and the typical work of each employee. Supervisors assign and employees accept work within this framework of prior agreement. At higher management levels, classification codes guide assignment of additional or new work (e.g., programs and projects). Classification codes identify the location and number of jobs with qualifications relevant to the new or additional work [e.g., Information Technology Specialists (Systems Analyst)]. The U.S. federal government can access this information for the million or so jobs covered by its central classification system.

Comparing Pay

The concept of “job” is fundamental to most methods for comparing pay. Comparing one person’s pay with another person’s pay has little meaning in the labor market or in court, unless both persons are doing substantially similar work. Work is usually considered “substantially similar” if both persons incumbent jobs with the same classification codes.

Appraising Performance

To appraise the performance of individual employees or to set standards for performance, it is necessary to establish a basis for comparison. Where work performance is the basis for comparison, job classification identifies which types of work are comparable. Comparing the work performance of a secretary with the work performance of an electrical engineer is as inappropriate as comparing circles with squares. Although both a secretary and an electrical engineer might have the same performance rating (e.g., “acceptable”), deciding in each case whether the incumbent has met acceptable work standards requires comparison with work standards that are appropriate for the occupation.

Comparing Unit Performance

Comparing performance of organization units requires a way to represent the types of work performed by the units and levels of resources available in the units. If Unit E carries out engineering aspects of environmental cleanup and Unit L carries out legal aspects of environmental cleanup, direct comparison of their work performance is not appropriate.

In cases where Unit E and Unit L each has a rating that refers to its own performance (e.g., a rating of “met plan”), each unit performance rating has little meaning without comparison to other, similar units. Once similar units are identified, comparison further requires some way to take into account the differences in resources available to each unit. For example, Unit E and Unit E2 both may be engaged in engineering aspects of environmental cleanup. Comparison of these units requires some consideration of resources available in each unit. It may be that Unit E has operated with 50% of its engineer positions vacant, whereas Unit E2 has had no vacancies. It may be that Unit E has had only one supervisor or senior worker for every eight engineers, whereas Unit E2 has had three. It may be that Unit E has 30% of its engineer positions filled at the trainee level, whereas Unit E2 engineers all are at full proficiency. Job classification provides the codes necessary to identify similar units, and within units, to sort out types and levels of resources (jobs) required and available.

Job classification codes outline particular configurations of resources and types of work, and support comparison and evaluation of unit performance.

Identifying Incumbent Rights and Entitlements

The classification of each job links the job with laws, regulations, agency guidance, and precedents that determine the eligibility of its incumbent for Fair Labor Standards Act of 1938 overtime; inclusion in a union bargaining unit, merit pay pools, and particular retirement, health, or paid leave benefits.

Keeping Records of Work and Employment

Job classification codes link records of hiring, paying, and appraising employees and connect the records to laws, regulations, and guidance in effect at specific times.

To fully consider the impact of changing a job classification system, executives and experts need to consider all the ways in which job classification supports management.

EXECUTIVES, MANAGERS, AND JOB CLASSIFICATION PROFESSIONALS

Classification is fundamental to management decisions about what type of people to hire, how much to pay them, which employees are eligible for union membership, how to distribute rewards, and how to report and explain staffing and pay to company executives, regulators, and the courts.

Private sector executives supported by classification experts select job classification techniques, the criterion for the relationships among jobs, and the coverage of job classification used by their organization. Executives use the information provided by job classification to control costs, account for expenditures, plan work, and allocate resources. In the public sector, the legislature makes most of these executive decisions.

Many public sector executives and many managers in both sectors are aware of job classification only as a restraint that limits whom they can hire and how much they can pay. Job classification professionals are the messengers who bring to managers the “bad news” about controls imposed at executive levels. The tension between the responsibilities of job classification professionals and managers sometimes generates an adversary relationship.

Job classification professionals are experts in applying the rules of the job classification system used by their employer. Job classification professionals must often certify the technical accuracy of job classification decisions before personnel actions can be implemented. Making an exception to the system for one manager almost inevitably evolves into an “exception” for all managers. In the federal system and several others, the job classification professional is bound by the ethical implications of classification rules that have a basis in law.

Job classification professionals typically classify jobs in many different occupations. Developing the ability to gather and understand enough information about a particular job, to classify it accurately in the time available to do so requires years of training and experience.

In recent years, the process of job classification has been somewhat simplified by a trend toward using less specific classification codes, and more generic job descriptions. The job classification process has been facilitated by the spectacular enhancement of access to classification resources such as standards, examples of job descriptions, and related information through use of the World Wide Web. Some government agencies have delegated responsibility to managers for classifying their own subordinate positions. In those circumstances, the role of job classification professional becomes one of providing expert advice and training, updating electronic resources, or auditing decisions made by managers.

SCIENTIFIC CLASSIFICATION

Classification is a necessary step in dealing with large amounts of information of all kinds. Ancient Egyptians classified, recorded, and analyzed the behaviors of the Nile River to interface more effectively with nature. Family farms classify income and expenses to more effectively interface with lenders or tax collectors. Modern organizations classify work to interface with external labor markets, as well as coordinate management of their internal workforces.

In the early 1900s, progressive reformers in the United States sought to address government corruption by instituting a merit system for some government jobs. Duties and responsibilities of each job covered by a merit system are typically specified and recorded in a job description. Qualifications required to do the work are specified, and selection of personnel is based on qualifications required to do the work. Pay is set in relation to the work of the job. Job classification is necessary to compare pay for a job with the pay for similar jobs in the labor market, and to compare pay of one job with pay of another job in the same organization. The Classification Act of 1923 established a classification system in the federal government, which has been modified and expanded by successor Acts, and now applies to about 1 million federal jobs.

Achievements of the physical sciences have generated much enthusiasm for scientific methods. Job classification methods echo methods of the physical sciences and, by virtue of establishing systematic procedures for gathering and analyzing information about jobs, have been called “scientific.”

Job classification, however, is more judgmental than chemistry. Chemists observe elements objectively and classify the elements with precision into categories of objective phenomenon, such as the categories in the Periodic Table of Elements. Job classifiers also observe work objectively; however, the classification system itself is a product of past and present values and circumstances, not a description of physical facts. Rather than a Periodic Table of Elements, the fundamental criterion of a job classification system is a hierarchy of jobs or job elements. The hierarchy consists of level relationships considered necessary for legitimacy of the classification system in a particular context. It is this hierarchy of jobs or job elements that any job classification system must reliably reproduce and rationalize. This hierarchy is the criterion used to develop the job classification system.

Within occupations, different levels or hierarchies of work reflect aspects valued more (e.g., the work of training others) and aspects valued less (e.g., working at a trainee level). Between and among occupations, differences depend on customary groupings of skills and knowledge in the labor market and the

educational system. Present groupings of skills and knowledge reflect how past occupations responded to opportunities and accidents of history. On the early Western frontier, the occupation of barber may have included the duties of dentist. When the occupation of barber separated into the two specialized occupations of barber and dentist, salaries of the specializations reflected what was valued more (dentistry) and what was valued relatively less (barbering) by labor markets past and present. Within a specific frame of reference, job classification can be used to systematically, predictably, and rationally replicate the hierarchy of values expressed in the criterion of the job classification system applied.

Results of job classification reflect values embedded in the selection of the job classification criterion, techniques, and coverage. Some societies, labor markets, or employers may value services of fully proficient barbers above those of fully proficient dentists. In those societies or labor markets, job classification systems would place jobs of fully proficient barbers in a higher pay range than jobs of fully proficient dentists.

Every job classification system rests on a criterion measure that embodies values. In the United States, the values typically echo those of the external labor market as it is experienced and interpreted by the employer at the time the classification criterion is chosen and the job classification system is established. In the mid-1970s, an interest in pay equity emerged and prompted some discussion of universal job evaluation systems. The idea of a universally applied classification criterion has not gained solid support in the United States. Relevant laws and judicial rulings typically define the standard for acceptable pay equity practice as the fair and consistent use by an employer of any professionally designed job classification system.^[2]

Job classification is presented in many personnel and human resources manuals as an objective, even scientific, process for determining job worth. This presentation is a partial truth. Job classification is, in essence, a systematic way of identifying correct relationships among jobs. The correctness of the relationships, however, is based on custom and culture as reflected in the criterion used to develop the job classification system. Correctness is not based on some scientifically verifiable fact or on a scientifically established theoretical model. Like the companion disciplines of budget and staffing, the scientific component of job classification is the “science” of management.

JOB CLASSIFICATION SYSTEMS AND TECHNIQUES

Different techniques of job classification involve codes, forms, and formats that appear to be very diverse.

Differences in codes, forms, and formats generate different jargon, and each requires mastery of a different vocabulary. However, if the same criterion is used, application of any of the techniques to the same job information should produce about the same job relationships and job structure. Consider, for example, the criterion of salary in the relevant labor market, and consider two different classification systems based on this criterion—a system using benchmark whole-job comparison, and a system using point factor evaluation. Formats, forms, and codes for classifying jobs under the two systems appear to be very different. However, classification of a group of jobs using the benchmark wholejob comparison should produce about the same job structure as classification of the same group of jobs using the point factor evaluation system. The two systems look different, but are fundamentally the same because both are based on the same criterion.

However, if the same type of classification system (e.g., a point factor system) is developed for two organizations, but each organization uses a different *criterion*, the job structure resulting from classification may be different. If organization A uses as a criterion the judgment of a committee of occupational analysts, and organization B uses the labor market, and if the analysts give “knowledge” more weight than does the relevant labor market, fully proficient librarian jobs may be placed above fully proficient secretary jobs in the structure of organization A, and below fully proficient secretary jobs in organization B, despite both organizations having used the same job classification technique.

Although different criteria may produce different job structures, culture and labor market dynamics influence all criteria toward conformity and stabilize long-term job relationships. Cultural values are shared by groups of experts within an occupation and by different employers, also by legislators in state governments or the U.S. Congress. These shared values influence views of the legitimacy of a job hierarchy. Criteria express values, and values are slow to change. Within occupations, relationships among different levels of work are stable. Fully proficient barbers have been paid more than barbers in training, at least since Medieval Europe.

Among occupations, however, relationships are less stable. Relationships among occupations are more likely to change in response to economic influences, especially in the short term. For example, a limited supply of electricians in an area of high demand could drive the market value of electricians above the market value of accountants. Responding to the opportunity for higher pay, trained electricians may be attracted from other areas or people may take training to become electricians and, over time, electricians may

be oversupplied in that area. Electrician wages may stop rising, and accountants may resume their previous place in the labor market hierarchy. Organizations that hire from the labor market or periodically adjust wages of employees to the labor market must adjust to changes in occupational relationships. Because most market changes are temporary fluctuations, it is often more appropriate to make temporary salary adjustments than to change a job classification system. The federal government's criterion measure for the classification system covering about 1 million white-collar jobs is expressed in legislation (Chapter 51 of Title 5, United States Code). The federal government adjusts to temporary market effects by paying "special rates" to certain occupations in certain areas. However, some of the market changes endure and require changes in the job classification systems to interface effectively with the labor market.

Other things being equal, if the criterion measure is the same, any or all the techniques applied to the same job information should produce about the same structure of jobs. Understanding the role of the criterion is the foundation to evaluation of a classification system. The efficacy of a job classification system is evaluated in terms of how well it supports classification of jobs such that the job hierarchy is an accurate reflection of the criterion. Understanding the role of the criterion is critical to assessing the currency and appropriateness of a job classification system, and to distinguishing between the need for temporary or fundamental system change. Each of the following techniques of job classification can be used to produce job structures consistent with a criterion measure chosen by an organization. A single criterion measure can apply to a broad group of diverse occupations, as in the case of the federal government system for classifying white-collar occupations, or to a smaller group of occupations.

Classifying Whole Jobs and Classifying Jobs Factor by Factor

Each of the following techniques may be used to classify a "whole job" or to classify a job factor by factor. Each job has a potentially limitless number of factors that could be considered in deciding its classification. Each technique discussed typically provides guidance as to whether to use whole job or factor-based classification. The advantages of wholejob classification include its simplicity and flexibility. The disadvantages of wholejob classification include the potential for inconsistency in classifying jobs due to the flexibility for each classifier to consider different factors or to weigh the same factors differently.

In both whole job and factor-based classification, consistency can be improved by having more than one classifier classify the job independently and subsequently reconciling their classification decisions, or by developing more specific guidance for classifying each group of jobs.

Factor-based classification specifies the number and nature of factors to be considered in classifying jobs. The advantages of factor-based classification include improved consistency in classifying jobs, and potentially more comprehensive records of the basis for classification decisions. The improved consistency is achieved by specifically guiding and limiting what is considered. Requirements to explicitly justify and document the basis for evaluation of each specific job factor potentially produce more comprehensive records. These advantages, however, come at a price—factor-based classification standards are more costly to develop and documentation required to record the basis for classification decisions is, overall, more extensive.

Ranking

In ranking systems, jobs are classified by being ranked in relation to each other by a group of raters who can credibly rate the jobs in the organization. Advantages of this technique include potential for minimal paperwork associated with classification and development of the job classification system and minimal cost. A disadvantage is that it is only as credible and defensible as the raters. The number of jobs that can be ranked effectively by any group of raters is limited, and only the original raters can authentically classify jobs consistently—that is, with reference to the same basis for judgments. This type of system may not be sustainable over time.

Benchmark Job Classification Standards

In benchmark systems, jobs are classified by reference to standards. The standard is composed of descriptions of actual jobs that exemplify each level of work. Advantages of this technique include the potential for less paperwork associated with classification of the jobs that match the benchmarks, the potential for application of the standards by as many people as are trained to use them, and the relative ease of communicating about classification of jobs that match benchmarks. A disadvantage of this technique is that it is only as credible as the process used to rank the benchmark jobs; also, jobs that do not match the benchmarks may be difficult to classify. This technique is most efficient for occupations in which there is not much variation within each level of work (e.g., journeyman level).

Narrative Job Classification Standards

Narrative job classification systems refer to standards consisting of narrative descriptions and examples of work at selected levels. There is usually a standard for each major occupational group of jobs, and the occupational standard usually describes each significant level of work within the occupation. Standards are often prefaced by a section giving general information about the occupation and describing features of work important in distinguishing one level within the occupation from another. Advantages of this technique include the potential for application of the standards by as many people as are trained to use them and the standards are typically general and can be applied to large numbers of jobs. Narrative standards allow flexibility for classifiers to weigh unique features of each job. The language distinguishing levels of work is explicit, and the standards provide a uniform basis for communicating about classification of work in an occupation. Disadvantages of narrative standards include the potential for inconsistency due to diverse interpretations of the general language describing each level of work in the occupation, and that the system is expensive to develop. Systems using narrative standards require “evaluation statements” to document the basis for classification decisions, and adequacy of statements may be uneven.

Points-Based Job Classification

Jobs are classified with reference to standards that are narrative descriptions explaining and exemplifying “points” in a range of point values. Although points may be used to express or combine multiple judgments of whole jobs, points almost always are associated with factor-based job classification systems. Advantages and disadvantages of the points-based job classification technique parallel those of narrative job classification. In factor-based systems, points restrict flexibility in judging the relative value of different aspects of work. Restricting flexibility improves consistency of ratings. For example, the work of an historian might match a higher level in respect to knowledge than in respect to authority. If the points for knowledge have greater weight in the overall classification, the job might be classified at the level of its knowledge. If the points for authority have greater weight, the job might be classified at the lower level. In the narrative type of job classification standard points are not included, which gives to whomever classifies the job more flexibility in determining which level is more appropriate for the historian.

Finally, the fundamental nature of all job classification techniques is to guide the process of

observation and decision making. Job classification is largely a judgmental process. Like personality, each job is unique. To classify a job requires reduction of the job to elements used in a particular classification system, and identifying the “best” match between the job and the appropriate classification standard. It is not surprising that different people sometimes reach different conclusions about the same job.

JOB CLASSIFICATION PROCESS

Classifying a job is the process of assigning to the job codes that represent the type and level of work assigned to the job. “Counseling Psychologist, GS-0180-13” is an example of a classification code. Job classification codes include by reference a great deal of information indicating the type and level of work, the qualifications characterizing a job, and the terms and conditions of the employment arrangement.

For the different job classification techniques discussed, the elements of the job classification process are essentially the same. The first step of the process is to describe the work to be assigned to the job. Although the job incumbent is usually consulted, supervisors usually control description of work assigned and may draft the description of the job. Approval of the job description by a second-line supervisor usually is required.

Classifying the job is the middle step of the job classification process. A job classification professional, a manager, a panel of peers, or a consultant may be responsible for matching the work of the job with the appropriate classification codes.

Job classification decisions usually require additional clearances and approvals before classification codes are assigned to a new job (or to a job that has changed). Assigning classification codes may affect other systems supported by job classification. For example, if a job has grown to a higher classification level, necessary clearances may be obtained and its incumbent may be promoted to a higher pay range. Some personnel or human resources systems require open competition for the “new” job at the higher level. If classification of a job reveals that new duties are “confidential management duties,” a decision may be made to withdraw the job from a union bargaining unit, triggering subsequent dealings with the union. When informed about the impacts of impending changes in classification codes, management in some systems has the opportunity to alter assigned duties so classification codes do not change in ways that have undesired consequences.

The final step in the job classification process is documentation of classification activity, reasoning, and outcomes. Documenting classification reasoning

permits detection of errors in observation or classification, and promotes consistency and perceived fairness by making transparent the basis for judgment. Requests for classification review may be frequent, and documenting classification activity helps to manage and monitor the work of classifying jobs. Documentation also facilitates making adjustments in other administrative systems that are caused by new or changed classification of a job. For example, changes in classification of a job may trigger changes in pay, training requirements, or bargaining unit membership. Even if the classification of a re-described job is not changed, when a job description has changed, other changes may be indicated (e.g., performance standards may have to be updated).

MANAGEMENT AND ACCOUNTABILITY

Job classification is, in its essence, an information system. As such, it is an important part of management systems for controlling personnel expenses and practices. The information made available by job classification, however, can be a vehicle for micromanagement by the U.S. Congress, state legislatures and other governing bodies, chief executives, budget directors, and others. Through job classification and its linkages to pay, staffing, and other types of practices, executives can dictate and monitor, among other things, exactly how many jobs will be available to do work, how much each job can be paid, and minimum qualifications required for each job. Managers and workers frequently chafe under the added demands and limitations of these types of top-down administrative controls.

However, job classification is a key instrumentality of government accountability. The merit system for the U.S. civil service initiated by the Pendleton Act of 1883 was an antidote to government corruption. Job classification is at the heart of any merit system. Reformers in the early 1900s saw job classification as a means for controlling patronage abuses. The work of government jobs would be defined, and the work of campaigning and soliciting campaign contributions would not be done on the public payroll. Only those qualified would be hired, and the qualifications required would be based on the work to be performed. Pay levels would be related to work performed, and jobs assigned equal work would be in the same pay range. In 1923, the U.S. government passed the Classification Act, which is the foundation of classification processes and approaches still used in the U.S. federal civil service.^[3]

The importance of job classification to public accountability is supported by its association with the emergence of merit systems in Central American governments following periods of political corruption.^[4]

The Draft Charter for Public Service in Africa implies adoption of merit systems for career civil service jobs and pay consistent with work—functions that are implemented and monitored using job classification codes.^[5]

MANAGING THE NEW WORKFORCE

Since the mid-1980s, management layering and team-based rationalization has reduced the number of middle-income jobs in the United States and restructured much of the workforce.^[6] The flatter, non-unionized, flexible workforce can be reorganized at almost Internet speed into different team configurations. Any given team may include both management and non-management members, both public and private sector members, both contingent and permanent employees, and both part- and full-time employees.

To the extent governments can simply “buy” services in the marketplace without regard to wage equity, credentialed qualifications, or other personnel rules, governments may simply specify the product desired. The challenge of managing work in this new environment would be passed on to contractors or teams agreeing to deliver the product. Whether it occurs inside or outside government organizations, management of work in the new environment requires even more information than managing work in traditional structures. Developing a bid, distributing work among teams, and assigning resources to teams requires some method of estimating amount of work. Recruiting from the labor market and making up teams requires identifying qualifications needed for jobs, matching qualifications with work, and matching work to be done with the “going rate” in the relevant labor market. To compare productivity of teams requires a way to identify teams that have comparable resources and work.

Job codes are essential information for managing and coordinating the new type of workforce, much as they have been for the traditional workforce. Traditional processes for providing the information, however, need to be improved to meet new demands. Flexibility of some linkages characterizing traditional civil service personnel systems might be increased. Some governments are considering “pay banding.” Also called “grade banding,” this type of system increases the number of levels of work covered in each pay range, thus decreasing the precision of job classification and possibly loosening the linkage of pay to the external labor market. Another approach would be to consider relaxing the linkage between level of work and level of qualifications in the manner of U.S. military systems. The military systems permit some assignments of persons to jobs classified at levels

above or below the person's level of qualifications and experience. Another approach to increased flexibility is breaking large, interlinked groups of occupations into subgroups. The U.S. government linkage of all white-collar occupations to the same classification criterion, for example, might be examined to determine feasibility of breaking occupations into smaller categories, each with a distinctive criterion, such as the professional, administrative, technical, and clerical categories widely used in the private sector.

CONCLUSION

Job classification creates the vocabulary that makes possible information linking work with pay, qualifications, and performance. It is the place where management responsibility and accountability often clash with the immediate needs and pressures of the workplace. Used widely in both public and private sectors, it is and has been a vital part of merit systems throughout the United States and around the world.

REFERENCES

1. Henry, N. *Public Administration and Public Affairs*; Prentice-Hall: New Jersey, 2001; 249 pp.
2. Risher, H.; Wise, L. Job evaluation: The search for internal equity. In *New Strategies for Public Pay*; Risher, H., Fay, C.H., Eds.; Jossey-Bass Inc.: San Francisco, 1997; 98–124.
3. Holley, L.M.; O'Connell, J.R. Job classification: The support system for decision making. In *New Strategies for Public Pay*; Risher, H., Fay, C.H., Eds.; Jossey-Bass Inc.: San Francisco, 1997; 76–97.
4. Klingner, D.E.; Nalbandian, J. *Public Personnel Management*; Prentice-Hall: New Jersey, 1998; 29–33.
5. Hon. Hage Geingob. Draft Charter for the Public Service in Africa (Windhoek Agreement). *Annex of Background and Synopsis of the Draft Charter for the Public Service in Africa* (2027) (United Nations Secretariat) <http://www.unpan.org/ST/SG/AC.6/2000/L.3>. Also see and select What's New, then select Windhoek. Articles 15 and 19.
6. Snyder, D.P. Future pay: Reinventing compensation for the information age. In *ACA News*; American Compensation Association: Scottsdale, AZ, 1994; 1, 17–20.

BIBLIOGRAPHY

1. <http://shrm.org>; Website of the Society for Human Resources Management, which responds to a search for "job classification" with topical information that has a private sector focus. Members of the society have access to more information than nonmembers.
2. <http://online.onetcenter.org/>; The O*NET system provides a "crosswalk" between occupational codes used in several, distinct U.S. classification systems such as the Standard Occupational Code System, the Dictionary of Occupational Titles, and the Military Occupational Classification system. The O*NET database houses information on key attributes and characteristics of workers and occupations and O*NET OnLine provides easy access to that information.
3. <http://www.bls.gov/soc/socguide.htm>; *Standard Occupational Code (SOC) System User Guide*; U.S. Department of Labor, Bureau of Labor Statistics explains the nationwide occupational classification system which allows government agencies and private industry to produce comparable data. It is designed to cover all U.S. occupations in which work is performed for pay or profit.
4. <http://www.oalj.dol.gov/libdot.htm>; This is an archived copy of the last (1991) edition of the U.S. Dictionary of Occupational Titles that was created originally in the 1930s to support the work of the U.S. Employment Service.
5. <http://www.opm.gov/fedclass>; U.S. federal job classification standards, guidance for application, and other related material can be read in this section of the website of the U.S. Office of Personnel Management.
6. http://www.un.org/Depts/icsc/ppd/job_clas/index.htm; Job classification standards used by the United Nations Common System of organizations with explanatory material (click on "View").

John Rawls

James R. Heichelbech

Graduate School of Public Affairs, University of Colorado at Denver,
Denver, Colorado, U.S.A.

INTRODUCTION

John Rawls was the most significant political philosopher in the United States during the 20th century. His work revitalized discussions of social equity in public administration and provided a focal point for critical reflection about social institutions. Publishing in over a hundred articles and books between 1950 and 2002, Rawls presented most of his ideas in three books: *A Theory of Justice*,^[1] *Political Liberalism*,^[2] and *Justice as Fairness: A Restatement*.^[3] The following includes a summary of the development of the theory of justice within these three books, a discussion of its significance for public administration and public policy, and a summary of criticisms.

A THEORY OF JUSTICE

A Theory of Justice is presented in three parts dealing with Theory, Institutions, and Ends. The first of these is undoubtedly the most important, as it presents the very idea of justice as fairness. Parts two and three concern the application of principles of justice and the relationship between these principles and *the good* (following the Kantian tradition of distinguishing the “right,” which concerns minimally necessary moral requirements, and the “good,” which concerns maximum positive happiness. Rawls explains the theory by clarifying the subject of justice, offering two principles of justice, and then presenting an argument for those principles from an *original position* behind a *veil of ignorance*.

The subject of justice, he says, concerns the basic structure of society—the way that social institutions distribute rights and duties resulting in division of advantages gained from social cooperation. The primary role of justice is to provide a standard for assessing the distributive aspects of the basic structure of society. Such principles are those that “free and rational persons concerned to further their own interests would accept in an initial position of equality as defining the fundamental terms of their association.” Justice as fairness views principles of justice as the basis for determining which kinds of social cooperation and forms of government are acceptable.^[1]

Rawls defines the basic structure of society as “the arrangement of major social institutions into one scheme of cooperation,” where an institution is understood as “a public system of rules which defines offices and positions with their rights and duties, powers and immunities, and the like,” and says that, ideally, “the rules should be set up so that men are led by their predominant interests to act in ways which further socially desirable ends.” To that end, Rawls offers two principles: “Each person is to have an equal right to the most extensive basic liberty compatible with a similar liberty for others,” and “Social and economic inequalities are to be arranged so that they are both 1) reasonably expected to be to everyone’s advantage and 2) attached to positions and offices open to all.”^[1]

An important feature of these principles concerns equality of opportunity. “The natural distribution,” he says, “is neither just nor unjust; nor is it unjust that persons are born into society at some particular position. These are simply natural facts. What is just and unjust is the way that institutions deal with those facts.” Accordingly, he suggests that a liberal interpretation of these two principles “mitigate the influence of social contingencies and natural fortune on distributive shares,” by requiring, for example, that “free market arrangements . . . be set within a framework of political legal institutions which regulate the overall trends of economic events, and preserves the social conditions necessary for fair equality of opportunity.”^[1]

The Original Position

The argument for these two principles uses the idea of the *original position*, a perspective from which we can devise a “fair procedure guaranteeing that any principle agreed to will be just.” As a preliminary step, Rawls lays out the formal constraints of the concept of right: “Principles should be general . . . It must be possible to formulate them in such a way that they are not tied to particulars . . . Principles are to be universal in application . . . The parties assume that they are choosing principles for a public conception of justice . . . [and] a conception of right must impose an ordering on conflicting claims.”^[1]

The appropriate position from which to arrive at principles within such constraints is that situated

behind a *veil of ignorance*. Although persons in the original position are rational, which means that “in choosing between principles each tries as best he can to advance his interests,” those behind the veil of ignorance do not know how their choice will affect their particular case and they are obliged to evaluate principles solely on the basis of general considerations. In other words, “while they know that they have some rational plan of life, they do not know the details of this plan, the particular ends and interests which is calculated to promote.”^[1]

This strategy for devising principles of justice reflects the deontological character of the theory. Rawls tells us that his aim is “to work out a theory of justice that represents an alternative to utilitarian thought generally,” which he describes as the idea that “society is rightly ordered, and therefore just, when its major institutions are arranged so as to achieve the greatest net balance of satisfaction summed over all the individuals belonging to it.” The primary difference between a theory of justice as fairness and utilitarianism, according to Rawls, is that “utilitarianism is a teleological theory whereas justice as fairness is not.” It follows, then, that justice as fairness “is a deontological theory, one that either does not specify the good independently from the right, or does not interpret the right as maximizing the good.” Priority of the *right* over the *good* is an important deontological feature of justice as fairness, for it restricts the basic structure of society in ways that a utilitarian account could not. Whereas any and every desire has some value from a utilitarian perspective, those in the original position agree “to conform their conceptions of their good to what the principles of justice require.” In other words, principles of justice, as principles of right, allow us to distinguish satisfactions that have value from those that do not and thereby define a reasonable conception of one’s good.^[4]

Political Liberalism

Political Liberalism is also presented in three parts, Fundamental Ideas, Three Main Ideas, and Institutional Framework. And again, the first part is critical for understanding the development of Rawls’s position. Here Rawls attempts to show how it is possible that “there may exist over time a stable and just society of free and equal citizens profoundly divided by reasonable though incompatible religious, philosophical, and moral doctrines,” and how “deeply opposed though reasonable comprehensive doctrines may live together and all affirm the political conception of a constitutional regime.” In other words, Rawls hopes to “work out a conception of political justice for a

constitutional democratic regime that the plurality of reasonable doctrines . . . might endorse.”^[2]

One of the primary tasks is to clarify justice as fairness as a *political* conception. Given the fact of reasonable pluralism, that a “diversity of reasonable comprehensive religious, philosophical, and moral doctrines found in modern democratic societies . . . is a permanent feature of the public culture of democracy,” justice as fairness is presented as “a conception of justice that may be shared by citizens as a basis of a reasoned, informed, and willing political agreement.” Such a conception, he says, should remain “independent of the opposing and conflicting philosophical and religious doctrines that citizens affirm” so that we “gain the support of an overlapping consensus of reasonable religious, philosophical, and moral doctrines.”^[2]

Rawls explains that a political conception of justice has three characteristic features. First, it is a moral conception to the extent that it is worked out for political, social, and economic institutions. Second, it is a “freestanding view,” which means that it is distinguishable from any comprehensive doctrine that it might be part of or derived from. Third, it is expressed in terms of ideas that are implicit in the public political culture of a democratic society.^[2]

Although the remainder of *Political Liberalism* certainly provides important details, most of the discussion serves to clarify concerns related to this political conception of justice. However, it is worth noting the three main ideas of political liberalism described in the second part. The first is the idea of an “overlapping consensus,” in which “the reasonable doctrines endorse the political conception, each from its own point of view.” The second is the priority of the right, which means that a political conception of justice limits conceptions of the *good*. And the third is the idea of public reason. Rawls explains that public reason is public in three ways: as the reason of citizens, its subject is the good of the public and fundamental justice, and its content is public.^[2]

Justice as Fairness: A Restatement

Here Rawls revises the two principles of justice. Second, he reorganizes the argument for those principles. Finally, he revises his understanding of justice as fairness, now understood “as a political conception of justice rather than as part of a comprehensive moral doctrine.”^[3]

While Rawls has quite a bit to say about his revised principles, there are no radical changes and the arguments for them are more of an adjustment than a departure from the original formulation. The revised principles are: “1) Each person has the infeasible

claim to a fully adequate scheme of equal liberties, which scheme is compatible with the same scheme of liberties for all. 2) Social and economic inequalities are to satisfy two conditions: first, they are to be attached to offices and positions open to all under conditions of fair equality of opportunity; and second, they are to be to the greatest benefit of the least-advantaged members of society (the difference principle).’’^[3]

On the other hand, while the explanation of justice as fairness as a political conception is similar to what Rawls presented in *Political Liberalism*, he provides additional clarification of the idea of free and equal persons within such a conception. “Justice as fairness,” he says, “regards citizens engaged in social cooperation, and hence as fully capable of doing so, and this over a complete life.” Accordingly, persons are assumed to have two moral powers, the capacity for a sense of justice and a capacity for a conception of the good. He describes a conception of the good as “an ordered family of final ends and aims which specifies a person’s conception of what is of value in human life or, alternatively, of what is regarded as a fully worthwhile life.” Such persons are equal in having these two moral powers and are free both because they “conceive of themselves and of one another as having the moral power to have a conception of the good,” and because they “regard themselves as being entitled to make claims on their institutions so as to advance their conceptions of the good.”^[3]

RAWLS AND PUBLIC ADMINISTRATION

Rawls’s work had and continues to have a profound impact on the normative dimensions of public administration and public policy. According to Terry Cooper, the publication of *A Theory of Justice* coincided with the development of administrative ethics as a field of study, especially significant with respect to social equity as an important part of the New Public Administration emerging in the late 1960s. In particular, he says, two essays contributed directly to administrative ethics as a field of study: “Social Equity and Organizational Man: Motivation and Organizational Democracy” by Michael Harmon and “Social Equity, Justice, and the Equitable Administrator” by David K. Hart.^[4,5] Both focused on the Rawlsian conception of social equity as applicable to public administration, “thus providing evidence of the practical significance of administrative ethics and building confidence in the possibility of developing it as a field of study.”^[6]

In practical terms, two aspects of the Rawlsian perspective have been important within discussions of ethical public institutions. First, the requirement of impartiality, which is implicit within the very idea of

neutral competence or merit system, gained theoretical grounding as an axiom of ethical decision making in the public sector. Second, the Rawlsian conception of political justice required attention to distributive inequities and helped articulate the basis for various policy initiatives, such as affirmative action. In other words, Rawls provided a practically useful argument for demanding equity both in decision-making processes and in determining who should benefit from them.

CRITICISMS OF RAWLS

The primary sources of criticism have come from communitarian and feminist perspectives. Stephen Mulhall and Adam Swift present a concise summary of various communitarian criticisms of Rawls offered by such theorists as Charles Taylor, Michael Sandel, Michael Walzer, and Alasdair MacIntyre, who argue that the liberal emphasis on the individual implies a “neglect of the formative significance of their social context and the moral significance of relations between them.”^[7] Mulhall and Swift identify four types of criticisms along these lines. First, the Rawlsian conception of the person requires a radical detachment from one’s nature and ends that is psychologically impossible and deprives one of resources needed to reason about social justice. Second, this asocial individualism neglects the extent to which the societies in which people live shape who they are and what values they have. Third, while Rawls requires the theory to be universally applied, it cannot take into account the different ways in which different cultures embody values and practices. Finally, while Rawls emphasizes the neutrality of the theory with respect to comprehensive doctrines, he “smuggles” in particular ideas of the good life for human beings.

Feminist criticisms of Rawls are of two sorts. One is a methodological concern about the legitimacy of the detached perspective of the original position. Another deals with concerns about the application of principles of justice to particular institutions, such as family.

Just as communitarian criticisms focus on the culturally nonneutral conception of rationality reflected in the original position, feminists point out that the original position reflects a gendered conception of rationality. For example, Carol Gilligan’s account of differences in moral reasoning between boys and girls helped clarify the ways in which abstract, detached, and rule-oriented conceptions of moral reasoning can be considered gender-specific. “Listening to people talking about morality and about themselves,” Gilligan “began to hear a distinction in these voices, two ways of speaking about moral problems, two modes of describing the relationship between other and self.” She distinguishes these two voices in terms

of a contrast between an ethic of *justice* and an ethic of *care*—“the logic underlying an ethic of care is a psychological logic of relationships, which contrasts with the formal logic of fairness that informs the justice approach.” The ethic of care “evolves around a central insight, that self and other are interdependent” and “a consciousness of the dynamics of human relationships then becomes central to moral understanding.”^[8]

These insights have been developed further in various ways. Martha Nussbaum provides a succinct overview of a rather diverse set of criticisms. For example, some argue that because “rationality of the parties in the original position is described as self-interested prudential rationality,” and “the parties are characterized as mutually disinterested, unaware of strong ties to others,” it discounts emotionally grounded approaches to moral reasoning and favors an egoistic conception of human beings. Those influenced by the work of Jurgen Habermas also question the “monological” nature of the original position in which “the parties are imagined as basically all alike, and as reasoning on their own, rather than exchanging claims and counterclaims in a dialogue in which different perspectives can be presented and investigated.”^[9]

In contrast to the methodological concern about how the principles of justice are chosen, feminists have also questioned the limited application of those principles of justice.

In *Justice, Gender, and the Family*, Susan Moller Okin suggests that Rawls’s assumption that those in the original position are heads of families “is far from being neutral or innocent,” for it “has the effect of banishing a large sphere of human life and a particularly large sphere of most women’s lives from the scope of the theory.”^[10] Although Rawls assumes that individuals have a sense of justice, he does not explain his assumption that family institutions are just. Still, Okin acknowledges that if we assume that those behind the veil of ignorance do not know their sex, there is implicit in Rawls’s theory a potential critique of social institutions with respect to gender.^[10]

CONCLUSION

Although Rawls has acknowledged that principles of justice have some application to social institutions,

such as family, his responses to both communitarian and feminist criticisms tend to reemphasize the importance of a political conception of justice derived apart from more comprehensive conceptions. And it is worth noting that he consistently offers his account as merely part of a more complete story about social institutions. From the perspective of public administration and public policy, this resonates with the treatment of social equity as merely part of what we need to consider when we think about ethical public institutions. However, he argues that justice as fairness is implicit within a commitment to democratic institutions, and therefore the relative priority of social equity reflects the relative priority of democracy itself.

REFERENCES

1. Rawls, J. *A Theory of Justice*; The Belknap Press: Cambridge, MA, 1971; 7–11, 30–31, 55–60, 73, 102, 130–136, 141–142.
2. Rawls, J. *Political Liberalism*; Columbia University Press: New York, 1993; xviii–xx, 9–13, 36, 134, 174, 213.
3. Rawls, J. *Justice as Fairness: A Restatement*; The Belknap Press: Cambridge, MA, 2001; xv–xvi, 18–34, 42–43.
4. Harmon, M. *Social Equity and Organizational Man: Motivation and Organizational Democracy*; Frederickson, H.G., Ed.; Public Administration Review Symposium on Social Equity and Public Administration, 1974; Vol. 34, 11–18.
5. Hart, D.K. *Social Equity, Justice, and the Equitable Administrator*; Frederickson, H.G., Ed.; Public Administration Review Symposium on Social Equity and Public Administration, 1974; Vol. 34, 3–11.
6. Cooper, T. The emergence of administrative ethics as a field of study in the United States. In *Handbook of Administrative Ethics*, 2nd Ed.; Cooper, T., Ed.; Marcel Dekker: New York, 2001; 11–12.
7. Mulhall, S.; Swift, A. Rawls and communitarianism. In *The Cambridge Companion to Rawls*; Freeman, S., Ed.; Cambridge University Press: Cambridge, UK, 2003; 460 pp.
8. Gilligan, C. *In a Different Voice*; Harvard University Press: Cambridge, 1982; 73–74, 149.
9. Nussbaum, M. Rawls and feminism. In *The Cambridge Companion to Rawls*; Freeman, S., Ed.; Cambridge University Press: Cambridge, UK, 2003; 488–520.
10. Okin, S. *Justice, Gender and the Family*; Harper Collins: U.S., 1989; 95–105.

Judicial Ethics

David A. Yalof

Department of Political Science, University of Connecticut, Storrs, Connecticut, U.S.A.

INTRODUCTION

Our nation's judicial system rests on the theory that a neutral and competent arbiter of disputes (the judge) provides the most effective means of resolving legal disputes among competing parties. Judges are responsible for defining each party's legal rights and responsibilities, allocating resources, and supervising the actions of public officials, among other tasks. In carrying out these duties, the willingness of each judge to respect and honor the judicial office as a public trust is considered fundamental to the success of the system; if members of the judiciary are not held to the highest standards of integrity and independence—or if there is even an appearance of impropriety in their actions—confidence in the judicial system may break down.

For well over a century, much attention has been paid to the ethical conduct of lawyers, who must balance the duty to provide zealous representation of their clients with their sometimes conflicting responsibilities as officers of the Court. Considerably less attention—scholarly or otherwise—has been paid to judicial ethics. In court, judges may be held to even higher standards of integrity than those to which lawyers or other participants in the process normally are held. Outside of court, a judge may still be held to exceedingly high standards of honesty and propriety, as private actions may also reflect on the judge's capacity to perform judicial functions with impartiality. Although judges certainly can (and often do) maintain some personal opinions on broad legal issues, the premise that judges can nevertheless be open minded and impartial regarding the specific legal issues and parties that come before them remains a bedrock principle of our legal system.

EXAMPLES OF ETHICAL DILEMMAS FACED BY JUDGES

Ethical considerations inevitably play a part in many of the decisions made by judges during their tenure on the bench. Unfortunately, the rules and laws that guide judicial conduct only go so far in specifying detailed responses to every possible situation a judge may face. For example, what role can a judge play in participating in public affairs when not actually

presiding over court? To maintain the appearance of propriety, a judge probably should avoid actively campaigning for political candidates or assisting in raising funds for political campaigns. But what about public service activities that call (at least in part) upon a judge's expertise in interpreting facts and the law? In 1946, Justice Robert Jackson temporarily stepped down from his position on the United States Supreme Court to serve as the federal government's chief prosecutor at the Nuremberg War trials. Immediately after President John Kennedy was killed in November of 1963, Chief Justice Earl Warren of the U.S. Supreme Court accepted President Lyndon Johnson's appointment to chair the commission investigating the circumstances that led to President Kennedy's assassination. On a lesser scale, local court judges are often asked to serve on various boards and committees in their home communities. Under what circumstances might such extra-judicial political activities compromise a judge's independence, or at least create the appearance of impropriety?

Financial conflicts of interest pose another ethical minefield for judges. Normally, we would expect judges to disqualify themselves from presiding over cases whenever their impartiality might be "reasonably" questioned. But what does the phrase "reasonably" entail? One judge was replaced in litigation involving the Exxon Corporation when his wife inherited less than one—one billionth of the stock of that company. Another judge stepped down from an antitrust case, because a family member owned less than 10 shares in one of the 2000 corporate parties involved in the litigation. Over 30 years ago, a federal appellate judge (Clement Haynsworth) was denied an appointment to the U.S. Supreme Court, largely because while serving on the corporate board of the Carolina Vend-A-Matic Corporation ("CVAC"), he had supervised a case involving a second company whose parent entity held \$50,000 worth of contracts with CVAC. When is a financial conflict significant enough that it requires a judge to recuse himself from all matters related to the conflict?

A third area of judicial ethics concerns on-the-bench judicial decision making that may, under certain circumstances, constitute an abuse of power. When a judge commits a legal error, it is generally a matter for legal appeal, rather than a question of judicial

misconduct. Still, some errors are so egregious (and perhaps deliberate) that they might be considered unethical, provoking calls for sanctions against the judicial offender. Judges have occasionally been disciplined for disposing of cases summarily without allowing an opportunity for even minimal adversary proceedings. One judge was sanctioned for coercing parties into disposing of their cases out of fear or intimidation. Judges cannot make inappropriate comments that demonstrate bias or insensitivity to the parties before them. In general, judges are required to administer justice through a faithful and competent application of the law. Yet when do judicial decisions cross the line from mere error into an improper use (or even outright “abuse”) of power?

The above examples barely scrape the surface of potential judicial misconduct. Once assuming judicial office, an individual must be prepared to navigate through a maze of ethical considerations, most of which are accompanied by ambiguous instructions and heavily qualified guidelines.

HISTORY OF JUDICIAL ETHICS

Prior to the 1960s, judicial misconduct in the United States was addressed almost exclusively through formal removal procedures, whether by impeachment, a bill of address passed by a legislature, or recall by the people. Because such remedies are quite extreme and cumbersome, they were only rarely utilized. (Legislatures may have also been wary of how frequent threats of removal might hamper judicial independence.) Consequently, many cases of judicial misconduct went unaddressed or ignored for much of our nation’s history.

At least some of the blame for this state of inaction lay with the legal profession’s governing body, the American Bar Association (ABA). The ABA’s lone attempt at formulating standards of judicial ethics proved an almost total failure: the original “Canons of Judicial Ethics” (promulgated by the ABA in 1924) were universally criticized as amounting to mere statements of “moral posturing” that provided little guidance in answering difficult questions of judicial ethics.

Beginning in the 1960s, however, the federal government and various state governments began to adopt new mechanisms for disciplining judicial misconduct. Their efforts focused on the adoption of formal codes of judicial misconduct and the creation of official judicial conduct organizations authorized to investigate and discipline judges. Finally in 1972, the ABA promulgated a new Model of Judicial Conduct to replace the mostly ignored Canons it had formulated almost a half-century earlier. In less than two decades, the 1972 Code was adopted (albeit with minor variations) by 47 states, the District of Columbia, and the

Federal Judicial Conference, which oversees judicial misconduct in federal courts. In 1990, a revised Model Code was adopted, and in the decade that followed, an additional 23 states adopted this new version of the code, which featured the use of gender-neutral language and other minor alterations. To date, 49 of 50 states have adopted some version of one or the other of these codes. (The lone holdout has been Montana, which created its own set of rules of judicial conduct.)

THE CODE OF JUDICIAL CONDUCT

Both the Model Code of Judicial Conduct promulgated in 1972 and the 1990 revised version of the code set forth the same basic standards that govern judicial conduct in and out of the nation’s courts. Although the codes are intended to provide a framework for disciplinary agencies to regulate conduct, they provide no basis for civil liability or criminal prosecution.

In both versions, Canon 1 requires judges to “uphold the integrity and independence of the judiciary.” It is in every sense the “philosophical backdrop” of the entire document; the other provisions of the code give substance to Canon 1.

Canon 2 requires judges to “avoid impropriety and the appearance of impropriety” in all of their activities. In the 1990 version of the code, Canon 2 admonishes judges to adhere to the most broadly accepted norms of social conduct in three specific ways:

- A judge must “respect and comply with the law” and shall act in a manner that “promotes public confidence in the integrity and impartiality of the judiciary.”
- A judge cannot allow personal or political relationships to influence “judicial conduct or judgment” and cannot lend the prestige of the office to advance the private interests of others, whether by conveying the appearance of favoritism or by testifying voluntarily as a character witness.
- A judge may not hold membership in “any organization that practices invidious discrimination on the basis of race, sex, religion or national origin.”

Because a judge’s actions often come under intense public scrutiny, the code specifically imposes burdens that separate a judge from ordinary citizens. According to the official ABA commentary accompanying Canon 2, the test for “appearance of impropriety” is whether the conduct would “create in reasonable minds a perception that the judge’s ability to carry out judicial responsibilities with integrity, impartiality and competence is impaired.” Under this provision, judges are encouraged to disqualify themselves whenever they

maintain a formal financial interest, personal relationship, or any other type of interest in the parties, litigation, or subject matter at hand. Informal appearances matter as well. (Consistent with Canon 2, at least one judge was disqualified for sitting at a lunch table with counsel during breaks, even though there was no evidence of any inappropriate discussion.)

Canon 3 of both codes provides that judges should perform the duties of their office “impartially and diligently.” Those duties include numerous adjudicative and administrative responsibilities, disciplinary responsibilities, and self-disqualification whenever the circumstances of the situation demand it.

The drafters of the 1990 Code of Judicial Conduct combined former Canons 4, 5, and 6 of the 1972 code into one Canon 4 that deals comprehensively with off-the-bench conduct. This new Canon 4 specifically prohibits judges from participating in extra-judicial activities that may cast doubt on their capacity to judge impartially, demean the judicial office, or interfere with the proper performance of judicial duties. Canon 4 does *not* restrict judges from engaging in various avocational activities (writing, lecturing, teaching, etc.); serving as a non-legal officer or advisor for an agency dedicated to the improvement of law, or as a part of a non-profit educational, religious, or charitable organization; or managing family financial affairs. But, judges are restricted from taking on extra-judicial assignments in government that are unrelated to the general improvement of law; engaging in business dealings that may be perceived as exploiting the judge’s position on the bench; receiving improper gifts or loans; and serving as a fiduciary in proceedings that might come before him. A judge may not practice law, except in an unpaid manner to draft or review documents for a family member.

Finally, Canon 5 of the 1990 code essentially reinstitutes Canon 7 of the 1972 code, which orders judges to “refrain from inappropriate political activity.” Judges are admonished to avoid active participation in political organizations and must resign from judicial office upon becoming a candidate for a non-judicial government position. As a candidate for judicial office, a sitting judge must “maintain the dignity appropriate to judicial office” while he is campaigning; thus, the judge cannot personally raise funds for office, even if subject to a public election. And, a judge cannot even speak to a political gathering until becoming a formal candidate for election. This final canon has been a source of some controversy, as judges seeking reelection have had to wrestle with the demands of a hard-fought campaign for office. Specifically, how does a judge defend positions on the campaign trail, while at the same time “maintaining the appropriate dignity” of the judicial office? Certainly, candidates for judicial office enjoy some leeway in criticizing opponents, but such

criticism should not bring the judge’s impartiality into question. As the reporter for the drafting committee of the original code explained, this complicated set of rules amounts to a set of “compromises between political reality and the aim of maintaining the appearance of judicial impartiality.”^a

PROCEDURES FOR DISCIPLINING JUDGES

Along with the proliferation of federal and state judicial ethics codes patterned after the Model Code, numerous new judicial conduct organizations were established in the middle part of the twentieth century to consider legitimate accusations of judicial misconduct and to supervise the application of judicial ethics provisions to offending judges.

In 1960, California became the first state to create a permanent state commission responsible for regulating judicial conduct, and other states quickly followed suit. By the early 1980s, all 50 states and the District of Columbia maintained organizations with the authority to investigate and prosecute judges for misbehavior. Unlike in earlier periods, outright removal was not the only penalty available to public authorities; most of these organizations enjoyed the power to impose or recommend sanctions ranging from simple admonishment to temporary suspensions, as well as removal.

To oversee judicial misconduct in the federal judicial system, Congress in 1980 passed the Judicial Councils Reform and Judicial Conduct and Disability Act, which created judicial councils in each federal circuit charged with reviewing complaints against judges and, where appropriate, ordering sanctions. As with state commissions, these judicial councils can impose a range of sanctions under the act; however, Congress specifically withheld power from such councils to remove a federal judge outright from office, on the theory that Congress enjoys the authority to remove duly appointed federal judges by impeachment under the Constitution.

Although early critics of judicial misconduct commissions and councils feared they would pose a serious threat to judicial independence, the structure of the system acts to allay any such fears. In each jurisdiction, judges play a role at every step of the process (in some states, judges actually comprise a majority of the commission’s membership). Moreover, decisions rendered by such commissions are appealable to another judicial court, transforming the process into a system

^aJeffrey Shaman, Steven Lubet and James Alfini, Eds.; *Judicial Conduct and Ethics*, 3rd Ed.; New York: Matthew Bender, 2000; 365 pp. (quoting E.W. Thode, *Reporter’s Notes to the Code of Judicial Conduct* (1973), p. 96.)

of “self-regulation” in which threats to judicial independence carry little credence.

In addition to the formal sanctions available to a judicial commission or council, Congress and individual state legislatures can also remove judges accused of certain specified levels of misconduct. Under the federal Constitution, a majority of the House of Representatives can impeach federal officers (including judges) for “Treason, Bribery, or other High Crimes and Misdemeanors;” the Senate then enjoys the sole power to try the judge on the impeachment charges and remove the judge from office if found guilty by a two-thirds majority. Most state constitutions refer to “criminal activity,” “serious malfeasance,” or “gross incompetence” as possible grounds for impeachment.

Although the procedures governing systems of judicial discipline vary widely from state to state, there are at least two general categories of systems currently in place. Forty-one states and the District of Columbia have adopted a “one-tier system,” in which a panel composed of judges, lawyers, and even some non-lawyers handles every aspect of the process, from the filing of complaints to the prosecution of charges, the conduct of hearings, and the issuing of a formal recommendation on sanctions. Some states require that the state’s highest court actually impose the sanctions; in other states, the panel may impose sanctions directly. Nine states feature a “two-tiered system” in which a panel investigates and prosecutes formal charges, but an entirely separate panel actually adjudicates the charges and makes a final disposition.

No matter which system is utilized, judicial conduct councils and commissions function as the equivalent of non-criminal investigatory bodies in which discovery occurs and procedural due process rights (including the right to counsel) must be provided. The records of all such proceedings are generally kept confidential, although many commissions at least provide for disclosure to the judge of the identity of the complainant and any witnesses.

CONCLUSION

Just as ethical standards applicable to lawyers must evolve as circumstances in the legal profession and

the court system change, so are standards of judicial ethics shaped by shifting circumstances affecting the selection, performance, and extra-judicial activities of judges. Today, judges participate in a myriad of activities in their local communities; much of that participation might have seemed imprudent as recently as a few decades ago. To stay relevant, the Canons of judicial ethics must therefore be flexible enough to react to these real-world changes, and yet stable enough so that they may provide useful reference points for judges who must rely on them for guidance on a continuing basis. Clearly, an interested observer looking for the most up-to-date understandings and interpretations of judicial ethics would be well-advised to look beyond the language and commentary of the Canons of Judicial ethics. Specifically, a number of Internet sources may be helpful in this regard, including web sites associated with the American Bar Association (<http://www.abanet.org/home.html>); the Federal Judicial Conference (<http://www.uscourts.gov/judconf.html>); and the American Judicature Society (<http://www.ajs.org/ethics1.html>). Additionally, one should actively consult web sites administered by each individual state’s judicial system, which may feature links to state variations on the Model Code, as well as to state judicial opinions interpreting various provisions on judicial ethics.

BIBLIOGRAPHY

1. American Bar Association. In *Model Code of Judicial Conduct*; American Bar Association: Chicago, 1998.
2. Boutros, T., et al. *State Judiciaries and Impartiality: Judging the Judges*; National Legal Center for the Public Interest: Washington, DC, 1996.
3. Eastland, T., et al. *Ethics in the Courts: Policing Behavior in the Federal Judiciary*; National Legal Center for the Public Interest: Washington, DC, 1990.
4. Lubet, S. *Beyond Reproach: Ethical Restrictions on the Extrajudicial Activities of State and Federal Judges*; American Judicature Society: Chicago, 1984.
5. Shaman, J., et al. *Judicial Conduct and Ethics*, 3rd Ed.; LEXIS Law Publishing: Charlottesville, VA, 2000.
6. Volcansek, M. *Judicial Misconduct: A Cross-National Comparison*; University Press of Florida: Gainesville, FL, 1996.

Judicial Independence

Ian Greene

*Department of Political Science, York University,
Toronto, Ontario, Canada*

INTRODUCTION

A key feature of jurisdictions that practice the rule of law is an independent judiciary. This principle means that judges must be in a position, and also perceived by the public to be in a position, to make judgements about the cases before them as impartially as possible. Any attempts to influence a judge's decision outside of the rules of the courtroom are violations of judicial independence, whether from elected politicians, public servants, or private interests. Illicit attempts to influence judges might be direct, such as bribes, threats, or a telephone call from a government official. Indirect breaches of judicial independence might take the form of manipulations of the salaries or working conditions of judges who fail to tow a particular line, or they might involve exploitation of the administrative arrangements or structures of courts with the intention of swaying judicial decisions. Imprudent practices by interest groups, politicians, or the media might also threaten judicial independence.

DEVELOPMENT OF JUDICIAL INDEPENDENCE

Judicial independence evolved along with the struggle for the rule of law. The victors in the Glorious Revolution in England in 1688 had fought for the supremacy of Parliament and its laws over arbitrary decrees by the monarch, and they also demanded an independent judiciary to apply the law evenhandedly. Judicial independence was officially recognized in 1701 in the Act of Settlement, which guaranteed security of tenure for superior court judges of the central courts and protected judicial salaries from tampering by the executive. Prior to 1688, however, there had been a number of attempts to further the cause of the supremacy of the law and judicial independence. In 1329, the Statute of Northampton "declared that no royal command shall disturb the course of the common law, and that if such a command is issued, the judges shall ignore it."^[1] Because judges served at the pleasure of the monarch, however, their desire to keep their judicial appointments and their income often outweighed their resolve to decide independently. Nevertheless, the growing popularity of the ideal of judicial

independence sometimes restrained monarchs from removing independently-minded judges.^[2]

After 1688, the major theoretical defense of the idea of an independent judiciary was provided by John Locke, who argued for "indifferent" judges to settle disputes about the application of the law so that the law would "not be varied in particular Cases, but [there should be] one Rule for the Rich and Poor, for the Favourite at Court, and the Country Man at Plough." For Locke, it was particularly important for members of the government to be subject to the equal application of the law, so that no one would be "a judge in his own case."^[3]

MECHANISMS FOR PROTECTION OF JUDICIAL INDEPENDENCE

The two means for protecting judicial independence in the Act of Settlement were appointment of central superior court judges for life during good behavior and the establishment of judicial salaries by Parliament. Prior to this time, judicial appointments had been at pleasure, meaning that the monarch could remove judges whose decisions he or she disagreed with. "During good behaviour" meant that judges could only be removed by a joint address of both houses of Parliament, and then only because of behavior inappropriate for a judge (such as accepting a bribe, consistently failing to attend at court hearings, or a serious lapse in morals). There has been only one serious attempt to remove a central court judge since 1714, and that was with regard to a judge accused of accepting bribes in the nineteenth century.^[4]

THE AMERICAN REVOLUTION AND THE SEPARATION OF POWERS

It is ironic that the principles of judicial independence that progressive Britons had struggled so hard to achieve near the beginning of the eighteenth century were not applied to the colonies. The failure of the colonial authorities to institute judicial independence was one of the grievances that led to the American revolution. The framers of the 1789 constitution were

familiar not only with the arguments of John Locke but also with those of William Blackstone and Montesquieu. Blackstone described the separation of powers that had developed between the judiciary and the other branches of government.^[5] Montesquieu, writing in the mid-eighteenth century, attributed the apparent success of the English system of government in preserving liberty to the separation of governmental powers into the legislative, executive, and judicial branches.^[6] He seemed unaware of the development of cabinet government in the United Kingdom about that time, which fused the executive and legislative branches, while continuing to treat the judiciary as a separate branch, more or less. The separation of the judiciary from the other two branches of government has never been as complete in the United Kingdom as in the United States. For example, the Lord Chancellor is both a cabinet minister and the head of the judiciary, and the law lords of the House of Lords constitute the final court of appeal. In the United Kingdom as in some other Commonwealth countries, constitutional conventions, or principles of appropriate behavior, limit the intermingling of the judiciary with the other branches of government, and are a major safeguard of judicial independence.^[1]

The theory of the separation of powers was adopted by constitutional framers in the United States. Adams wrote, "The judicial power ought to be distinct from both the legislative and executive and independent upon both, so it may be a check upon both."^[7] According to Madison, "The accumulation of all powers, legislative, executive, and judiciary, in the same hands . . . may justly be pronounced the very definition of tyranny."^[8] And Hamilton, referring to Montesquieu, argued that "liberty . . . would have everything to fear from [the judiciary's] union with either of the other departments . . ."^[9,10]

The result of the constitutional debates was Article 3 of the U.S. Constitution, which describes the "judicial power of the United States." Article 3 was intended to be "a pillar of the separation of powers designed to provide genuine judicial independence."^[11] In Article 3, we see the two mechanisms for protecting judicial independence developed by the Act of Settlement, security of tenure, and establishment of judicial salaries: "The Judges, both of the supreme and inferior courts shall hold their Offices during good Behavior, and shall at stated Times, receive for their Services, a Compensation, which shall not be diminished during their Continuance in Office." Federally appointed judges may be removed through impeachment by the House of Representatives and conviction by the Senate. Article 3 does not apply to the State judiciaries, which have developed their own methods for protecting judicial independence. These usually include safeguards for judicial salaries to prevent arbitrary

tampering and security of tenure during fixed terms (renewable by appointment or election) and sometimes during appointments to retirement.

Between 1803 and 1989, seven federally appointed judges were impeached and convicted, and six others were impeached. Six of those convicted were District Court judges, and one was a Commerce Court judge. The reasons for the convictions included loose morals, frequent drunkenness, conviction for serious criminal charges such as tax evasion or perjury, or accepting bribes. A judge was convicted in 1862 for supporting the Confederate side in the Civil War.^a

Every liberal democracy claims to have an independent judiciary, as well as administrative tribunals which, although not often possessing the same guarantees of judicial independence as judges, nevertheless, are intended to decide the issues that come before them impartially. Judicial independence often has constitutional status and safeguards, although constitutional safeguards are neither a necessary nor a sufficient condition for judicial independence. The test for judicial independence is the degree to which judges feel they can decide impartially, without undue influence from outside the courtroom.

THREATS TO JUDICIAL INDEPENDENCE

Peter Russell has suggested that the threats to judicial independence can be sorted into four categories: structural, personnel, administrative, and direct.^[12] To this list can be added a fifth category: societal.

Structural

Even if courts have constitutional protection, legislatures can often alter the size and the jurisdiction of courts. If such alteration occurs as an attempt to influence the judges' decision making, then judicial independence becomes an issue. Perhaps the most famous example of a structural threat to judicial independence was President Franklin Roosevelt's proposed legislation to increase the number of judges on the Supreme Court by six in 1937. The plan was a reaction to the fact that the Court had struck down 16 New Deal laws in a little over a year. The six new positions would have allowed Roosevelt to appoint additional judges more sympathetic to the New Deal. Although the New Deal was popular, Roosevelt's attempt to erode judicial independence was not, and Roosevelt withdrew the controversial legislation. In the end, after one of the

^aNine other judges resigned before they could be formally impeached (Ref.^[11]).

conservative judges changed his mind about the New Deal legislation, and after Roosevelt had a chance to appoint a judge, the Court stopped striking down the New Deal laws.^[11]

The transfer of jurisdiction away from a court or tribunal so that a government can reduce the caseload of judges appointed by a different government, or the reorganization of a court to make redundant the jobs of judges unpopular with the government, are also examples of the structural threats to judicial independence that occasionally occur even in liberal democracies.^[13]

Personnel

Personnel issues focus around hiring, promotion, remuneration, and firing. The three systems of judicial selection are appointment by the government (either through partisan or nonpartisan procedures), acceptance into a career judiciary through competitive examinations (as in many civil law jurisdictions), and popular election, as in many U.S. States. The non-partisan appointment method and the competitive examination method are the most compatible with the promotion of judicial impartiality and, therefore, independence, while popular election (especially when combined with periodical subsequent confirmation elections) may pose the greatest risk.

The system of judicial promotions may leave open the possibility that judges may be rewarded for pursuing a particular ideological line in their decision making. In Canada, although judicial selection advisory committees have helped to promote more non-partisan systems for judicial appointment to the trial level in most jurisdictions, the federal cabinet's rationale for elevating trial judges to the appellate level remains shrouded in secrecy. And in Japan, it may be that lower court judges hoping for a promotion tend to adopt the ideological orientation of the judicial elite responsible for making judicial promotions.^[14]

It is well established that independent judges cannot have their salaries reduced except during times of economic crisis, and judges must not be treated more harshly than others on the public payroll during such periods. How to determine judicial salary increases is a more thorny issue. It is not easy to conduct salary negotiations between the judiciary and the executive in a way that avoids the appearance of violating judicial independence. In 1997, the Canadian Supreme Court declared that judicial independence requires the creation of independent judicial compensation commissions to recommend adjustments of judicial salaries

on a triennial basis. The Court required governments to be prepared to demonstrate that the recommendations of the commissions were given serious consideration.^b This decision has raised the question of whether judges can sometimes exaggerate the scope of judicial independence in order to protect professional autonomy interests unrelated to judicial impartiality. Whether administrative tribunals and Justices of the Peace also require compensation commissions is a hot topic currently being litigated in Canada.

Independent judges cannot be removed because their decisions are unpopular or because of alleged errors in their reasoning, but only when it has been demonstrated through an impartial process that the judge is not capable of undertaking judicial duties fairly and impartially. In 2001, Zimbabwean President Robert Mugabe encouraged state officials and veterans supporting his cause to threaten the safety of several Supreme Court judges who had found that the government's seizure of farms owned by whites was illegal. The result was the resignation of three judges, paving the way for new appointments who then declared the seizures lawful.^c A less blatant example of an attempt to pressure a judge to resign is the parliamentary enquiry established to consider allegations against a radical judge on Australia's High Court in 1986.^[13]

Administrative

Administrative practices in the courts have the potential to erode judicial independence in a number of ways. For example, prosecutors may be in a position to manipulate the case scheduling system so as to get the cases they are most concerned about before judges with a reputation of being sympathetic to prosecution arguments. Those who control the court's budget may use this power to reward judges who favor the administration's ideology by providing more prestigious working conditions. Worries by judges about the potential abuse of power by the executive branch caused the federal judiciary in the United States to push for and eventually achieve judicial control over court administration in 1939. The result is that the machinery of federal court administration is controlled by committees of judges. Although this system has advanced judicial independence in some ways, the disadvantage is that the courts may not be in a better position than before to obtain a reasonable budget allocation. As well, judges may not be as skilled in administration as those employed by the executive

^bReference re Remuneration of Judges of the Provincial Court of Prince Edward Island, [1997] 3 S.C.R., 1.

^cMugabe seized lands lawfully, court says. *National Post*, Dec. 5, 2001, A12, quoting from *Agence France-Presse* and *The Daily Telegraph*.

branch.^[15] In Canada, the Supreme Court has declared that in order to protect judicial independence, judges must have control over the case scheduling system, the assignment of judges to cases, and court sittings.^[16] The Court was not persuaded that judicial control over other aspects of court administration was necessary to protect judicial independence.

Threats to judicial independence can also come from within the judiciary. Chief judges who want the puisne judges to favor a particular ideological stance might, for example, reward compliant judges with more conference travel. On appellate courts, there is the danger that a chief justice might arrange panels so as to increase the likelihood of a particular outcome.

Direct

Direct threats to judicial independence can be blatant, such as attempts to bribe judges or to intimidate them with physical threats. The public criticism of judicial decisions by elected politicians may also threaten judicial independence, because judges cannot defend themselves through entering into a public debate without impugning their impartiality. As liberal democracies mature, however, conspicuous attempts to influence judges become infrequent. More subtle illicit contacts can, however, be just as damaging. Often, these more abstruse attempts to influence judges take the form of an informal chat or telephone call between a judge and an elected politician or senior administrative official, during which the judge is lobbied to handle a case in a particular way. This kind of contact with a judge represents undue influence, because it might result in an unfair advantage to the party attempting to influence the judge.

Societal

Societal threats to judicial independence include sensationalist media coverage of judicial proceedings or decisions, and demonstrations in front of courthouses that are intended to attract media attention and pressure judges to decide cases in a particular way. Judges in liberal democracies generally do not regard scathing critiques of their decisions in academic journals as a threat to judicial independence, because the writers are usually well-informed about legal issues, and academic analysis is regarded almost as a form of continuing legal education for judges. However, if media criticism is out of context, or seriously unbalanced, this can affect the judge's public reputation and family life. Because of judicial independence, judges may not enter into public debate to defend their decisions. Some judges fear that a conscious or unconscious desire to avoid being pilloried in the media might affect the

judicial reasoning process.^[17] On the other hand, if judges have security of tenure, their salaries may not be tampered with, and they are administratively secure, it is questionable whether societal threats to judicial independence are sometimes more perceived than real.

CONCLUSION

Judicial independence consists of arrangements to ensure that judges are in a position in which they feel they can decide cases as impartially as practically possible. The impartial application of the law is an important principle of liberal democratic government. Arrangements established to protect judicial independence usually include security of tenure until retirement or during a fixed term, the prohibition of tampering with judicial salaries to influence judicial decision making, and providing judges with control over administrative matters that can have an important impact on judicial decision making.

Regardless of the safeguards built into the law or the constitution to protect judicial independence, those wishing to influence judicial decisions outside of the regular courtroom process will attempt to find ways to interfere with judicial independence. Therefore, procedures designed to protect judicial independence need to evolve in order to meet new threats and challenges. Enhancing the protection mechanisms for judicial independence, however, needs to proceed cautiously, lest judicial independence is allowed to encroach on the rightful domains of executive and legislative independence. For example, when judges decree that judicial compensation commissions are constitutionally required to protect judicial independence, have they overstepped the boundaries of judicial independence? More generally, an activist judiciary that imprudently strikes down legislation on constitutional grounds puts judicial independence at risk.

All liberal democracies struggle with the question of the degree to which courts should control their own administrative apparatus in order to promote judicial independence. This issue is particularly important in jurisdictions like the United States that adhere to a strict separation of powers doctrine.^[18] On the one hand, administration—including courts administration—is arguably the responsibility of the executive branch. On the other, executive administration of the courts can be abused so as to endanger judicial independence. However, courts that control their own administration machinery always encounter judicial independence concerns when dealing with budgetary and labor relations issues. As with many administrative conundrums, there are no perfect solutions, and intra-state and international comparisons of attempted resolutions of these issues is always enlightening.

REFERENCES

1. Pucknett, T.F.T. *A Concise History of the Common Law*, 3rd Ed.; Butterworth and Co.: London, 1940; 45 pp.
2. Lederman, W.R. The independence of the judiciary. *Can. Bar Rev.* **1956**, *34*, 769–1139.
3. Locke, J. *Second Treatise on Government*; Library of Liberal Arts: New York, 1952; 51c. 1690.
4. Stevens, R. Judicial independence in England: a loss of innocence. In *Judicial Independence in the Age of Democracy: Critical Perspectives from Around the World*; Russell, P.H., O'Brien, D.M., Eds.; University Press of Virginia: Charlottesville, 2001; 155–172.
5. Blackstone, W. *Commentaries on the Laws of England*, 3rd Ed.; John Exshaw: Dublin, 1768; 269 pp.
6. Montesquieu. *The Spirit of the Laws*, Rev. Ed.; Colonial Press: New York, 1900; 151–160.
7. Adams, J. On government. In *The Works of John Adams*; Adams, C.F., Ed.; Little, Brown: Boston, 1856; Vol. 4, 181–198.
8. Madison, J. The Federalist No. 47. In *The Federalist: or, The New Constitution*; Hamilton, A., Madison, J., Jay, J., Eds.; Dent: London, 1961.
9. Hamilton, A. The Federalist No. 78. In *The Federalist*; 1788.
10. Wheeler, R. *Judicial Administration: Its Relation to Judicial Independence*; National Center for State Courts: Williamsburg, VA, 1988.
11. Abraham, H.J. The pillars and politics of judicial independence in the United States. In *Judicial Independence in the Age of Democracy*; Russell, P.H., O'Brien, D.M., Eds.; 2001; 25–36.
12. Russell, P.H. Toward a general theory of judicial independence. In *Judicial Independence in the Age of Democracy*; Russell, P.H., O'Brien, D.M., Eds.; 2001; 1–24.
13. Williams, J.M. Judicial independence in Australia. In *Judicial Independence in the Age of Democracy*; Russell, P.H., O'Brien, D.M., Eds.; 2001; 173–193.
14. O'Brien, D.M.; Ohkoshi, Y. Stifling judicial independence from within: the Japanese judiciary. In *Judicial Independence in the Age of Democracy*; Russell, P.H., O'Brien, D.M., Eds.; 2001; 37–61.
15. Fish, P.G. *The Politics of Federal Judicial Administration*; Princeton University Press: Princeton, NJ, 1973.
16. Valente v. The Queen, [1985] 2 S.C.R., 673.
17. Greene, I.; Baar, C.; McCormick, P.; Szablowski, G.; Thomas, M. *Final Appeal: Decision-Making in Canadian Courts of Appeal*; Lorimer: Toronto, 1998.
18. Vile, M.J.C. *Constitutionalism and the Separation of Powers*, 2nd Ed.; Liberty Fund Inc.: Indianapolis, IN, 1998.

Judicial Selection and Merit Selection

Cole Blease Graham Jr.

*Department of Government and International Studies, University of South Carolina,
Columbia, South Carolina, U.S.A.*

INTRODUCTION

Based on the principle of dual sovereignty, there are parallel court systems in America—one for the federal or national system and one for each of the 50 states. Generally, these courts are at three levels—an entry level called circuit or district, an intermediate or appeals level, and a final or supreme court. Complementing these general civil and criminal and review courts are specialty courts, such as a regulatory agency court or a court that deals with a single problem such as bankruptcy, as well as courts of lesser or preliminary jurisdiction such as magistrates.

The United States Constitution establishes the process through which the President and Senate select judges for the United States Supreme Court. Congress initiates legislation to create and design all other federal courts. Similarly, individual state constitutions define each state's court system and the roles of the executive and legislative branches or local governments in selecting judges. Reformers have advocated a merit plan for judicial selection to replace the more traditional electoral processes within the states. A merit plan is based on more defined criteria for the qualifications of a judge.

OVERVIEW

When thinking about courts, it is easy to focus on the prominence of the persons or parties on trial and to overlook the fact that courts are really managed by lawyers who are guided by judges. Lawyers have to be guided through court proceedings by judges to create a written record which is respected by other lawyers and by higher courts. A judge must have a sufficient legal background to direct the operation of a court credibly. The American Bar Association (ABA), long an advocate of merit selection, has proposed professional qualifications instead of political party membership as the preferred basis for judicial selection. In addition, the ABA argues that judges should have solid personal attributes such as “superior self-discipline” and “moral courage.”^[1] The ABA also promotes standards for evaluating the performance of a sitting judge

for any needed disciplinary measures and the development of continuing education programs.

MERIT CONCEPT

The merit concept is applied more vigorously in countries where the judiciary is treated as a separate, civil service profession.^[2] In such instances, judges are selected by higher entry qualification standards, such as exceptionally high student grades, success in special examinations, and completion of additional training requirements. A judge's career moves up to more demanding jurisdiction and responsibility based on positive evaluations by higher-level justices rather than by elections or executive appointment. One exception in a civil service system for judges is a constitutional court to which a Parliament may elect some proportion of the judges, even then usually from among already qualified judges and for a limited term.

JUDICIAL SELECTION

Judges in the American federal system are appointed by the President and confirmed by the Senate. The advice and consent clause of the United States Constitution has the intent of promoting an independent, highly professionalized federal judiciary without advancing a specific political ideology or political party. It is nonetheless a highly politicized process.^[3,4] Except for Supreme Court justices, the senior Senator of the President's party from the state in which the vacancy occurs or a new judgeship is created typically controls the appointment. Most federal judges have been members of the same political party as the President who nominated them.

Before confirmation by the Senate, a presidential nomination is reviewed by the Senate Judiciary Committee. The committee will not act until it has received the ABA report on each nominee, a tradition that began in the mid-twentieth century. Each potential nominee gets a questionnaire and returns it in 3 to 4 weeks. The ABA committee evaluates integrity, competence, and judicial temperament and rates each as

“well qualified,” “qualified,” and “not qualified.” In March, 2001, the George W. Bush administration discontinued reliance on the ABA report before announcing its nominations and, instead, decided to rely on the advice of the Federalist Society.^[5]

The political reputation of a judicial nominee often predetermines the outcome of the selection process. Lately, a “blue-slip” policy has emerged in which an individual senator, often, but not always from the home state of a nominee, may veto confirmation.^[6] This practice stalls confirmation of judicial nominees. Critics fear that the integrity of federal courts is threatened by a growing number of judicial vacancies, delays in filling them, and a decrease in legislative respect that may lead to mediocre compromise selections and reduced performance.

States use five approaches to judicial selection: popular election by party (partisan); election without regard for party (non-partisan); legislative election; merit selection; and gubernatorial appointment.^[7] Following the “common man” ideals of Jacksonian democracy that promote the qualifications of the ordinary citizen for government positions, 39 states elect state supreme court justices, and many also elect appeals court judges. Election advocates believe voters should express their preference, and that the ballot box is the best means to select a qualified judiciary.

Judicial candidates in partisan contests must finance their campaigns, make speeches at political rallies, and take positions on issues just like other candidates. Often, campaigns become heated, and candidates are negatively attacked for previous actions as a lawyer or prosecutor, such as defending a drug dealer, or for decisions in previous experience as a judge that may be portrayed as too soft on crime. The ABA Code of Judicial Conduct discourages aggressive partisanship by recommending against discussion of legal issues in partisan campaigns. Challenges by judicial candidates to the ABA ban or other ethical restrictions are often sustained on a First Amendment right to free speech argument.

In June, 2002, the Supreme Court of the United States struck down a canon of judicial conduct adopted by the Minnesota Supreme Court to bar candidates for judicial offices from publicly stating their views on controversial legal or political issues. The nation’s highest court cited two reasons for its action. First, it found that the challenged canon violated the First Amendment of the U.S. Constitution, which protects the content of speech. Second, it decided that this canon infringed upon a category of speech protected by this amendment; namely, the right of judicial candidates to express publicly their qualifications for such offices [*Republican Party of Minnesota v. White*, 536 U.S. (2002)].

Non-partisan elections select candidates without party identification or the participation of a political

party in an effort to reduce negative politics or party influence. It is arguable that non-partisan elections do not achieve the ideals of reformers for an independent judiciary when compared to elections based on party.^[8] Both approaches to popular election of judges are criticized for low rates of voter participation, increased costs of campaigning, and the tendency of incumbents to be reelected continuously. Where judges are elected on a less-than-statewide basis, for example, a state circuit or district judge, the Voting Rights Act of 1965 has been increasingly applied by federal officials to the electoral process if racial discrimination may be present.

Reform of popular election of judges has focused significantly on campaign finance. In February, 2002, the ABA recommended public financing for judicial elections as a way to reduce the fund-raising burden on candidates, the potential influence of contributors, and suspicious criticism of the fairness of courts. Although 14 states provide some level of public financing, recent campaigns have become more expensive and more combative. In a study initiated by merit system appointment advocates, it was found that spending in state supreme court elections doubled from 1994 to 1998 and increased another 60+ % in 2000 to total over \$45 million.^[9]

Among opponents of public financing are business interests which may be adversely affected by court decisions. Some interest groups issue reports on candidates as a way to influence voter decisions. In 2002, the United States Chamber of Commerce and the Business Roundtable launched an effort—the Litigation Fairness Campaign—to raise and direct campaign financing to influence judicial elections in states where previous outcomes in tort litigation in business liability has hurt the business community.

Legislative election remains in only two states—South Carolina and Virginia—where a majority vote of legislators selects a judge. These two former English colonies perhaps continue the tradition of legislative action to offset the potential impact of executive appointment authority or the fear of popular elections. A recent reform is the addition of a preview panel to recommend the qualifications of candidates rather than have candidates appeal directly to the legislature. Legislative voting outcomes show an overwhelming tendency to elect candidates who have previously served in the legislature. Opponents of this approach to selection argue that serving as a judge is very different from legislative service, and that public accountability is reduced by insider pressures for legislative conformity.

About half the American states use a merit selection plan. The typical plan is called the Missouri Plan after the state that first adopted the ABA reform recommendations. It typically involves four steps. First, a

nominating commission composed of judges, bar association representatives, and citizens is established. It is typically non-partisan or attempts to represent both parties equally. Second, the nominating commission generates a list of candidates, conducts hearings or investigations on them, picks three names from the long list, and forwards the names with accompanying information to the governor. Third, the governor appoints one to the judgeship. Fourth, at the next general election, the governor's appointee is confirmed for a full term or rejected.

Some variations occur within the states, as the elements of a Missouri merit selection plan are combined differently. For example, in California, the governor goes first by sending a nomination to a commission on judicial appointments made up of the Chief Justice, an appellate judge, and the Attorney General. If the commission approves, a candidate takes office until the next general election, at which time, the candidate is voted on for a 12-year term. New Mexico features nomination by a commission and appointment by the governor. But, for a full term, the nominee must run in the next general election on a partisan basis and in the subsequent general election on a non-partisan basis.

Gubernatorial appointment as an exclusive process for judicial selection is used in five states. It is an example of checks and balances between two branches of government. The popularly elected executive reflects the will of the people in designating judges who then act independently of the executive. Critics inject that there may be more politics than meets the eye, because a governor's selection may reflect political considerations more than judicial qualifications. The potential for executive political influence is expanded even more, because governors often appoint individuals to fill vacancies before elections or a merit system commission can act, thus giving a specific candidate the benefit of incumbency.

CONCLUSION

The President and the U.S. Senate have played key roles in judicial selection in federal courts during the more than 200-year history of the nation. Recently,

political controversies, often sharply divisive, have made this process more contentious and delayed filling some vacancies. Presidential nominations have become less reliant on the influential role of the ABA. Senatorial advice and consent decisions have diminished senatorial courtesy by giving greater leeway to an individual Senator's ability to veto a presidential nomination.

About half the American states select judges based on merit principles of professionalism as a way to soften what may be called the unnecessary politicization of the courts through campaigns and elections. Advocates of judicial elections maintain that political campaigns give voters substantive choices and enhance the accountability of judicial performance. Many states now include campaign finance reform as a way to orient popular judicial elections toward merit selection concerns.

REFERENCES

1. American Bar Association. *Standards Relating to Court Organization*; American Bar Association: New York, 1974; 43–44.
2. Murphy, W.F.; Pritchett, C.H.; Epstein, L. *Courts, Judges, and Politics: An Introduction to the Judicial Process*, 5th Ed.; McGraw Hill: Boston, 2002; 137–138.
3. Giles, M.W.; Hettinger, V.; Peppers, T. Picking federal judges: a note on policy and partisan selection agendas. *Polit. Res. Q.* **2001**, *54* (3), 623–641.
4. Yalof, D.A. *Pursuit of Justices: Presidential Politics and the Selection of Supreme Court Nominees*; University of Chicago Press: Chicago, 1999.
5. Palmer, E. Senate GOP backs down from dispute over handling of nominees. *Congr. Q. Wkly. Rep.* **2001**, *59* (23), 1360–1361.
6. Taylor, S., Jr. Judicial selection: compromise on ideology, not quality. *Natl. J.* **2001**, *23* (19), 1379–1380.
7. Bowman, A.O'M.; Kearney, R.C. The judiciary. In *State and Local Government*, 5th Ed.; Houghton Mifflin: Boston, 2002; 249–256.
8. Hall, M.G. State supreme courts in American democracy: probing the myths of judicial reform. *Am. Polit. Sci. Rev.* **2001**, *95* (2), 315–330.
9. Stone, P.H. The blitz to elect business-friendly judges. *Natl. J.* **2002**, *34* (7), 480.

Just War

Cyrus Ernesto Zirakzadeh

Department of Political Science, University of Connecticut, Storrs, Connecticut, U.S.A.

INTRODUCTION

Proponents of just war argue that disinterested principles of universal justice should motivate and inform a government's deliberations about whether and when to combat a foreign power. Moral principles should trump preoccupations with international power, prestige, money, or security. Just-war theorists therefore seek to clarify both: 1) the circumstances under which states are morally justified in going to war; and 2) the amounts and forms of violence that, once the battles commence, are morally acceptable.

RIVAL APPROACHES TO MORAL REASONING IN INTERNATIONAL RELATIONS

Because the concept of a just war presumes that moral ideals and standards should play a key role in foreign-policy making, this style of political thinking diverges from so-called "realist" schools of thought, which maintain that moral concepts are (or ought to be) irrelevant to foreign-policy making. Realists believe that state leaders should eschew moral ideas and analyze foreign-policy questions primarily in terms of national self-interest.

Historians of political thought, such as Brian Orend and Jack Donnelly,^[1] have noticed that realist arguments against the concept of just war can take two basic forms. One, often called "descriptive realism," contends that states, whenever they make policies about foreign relations, are always motivated first and foremost by considerations of power, security, and national interest. According to descriptive realists, such as Hans Morgenthau and Karl von Clausewitz, proponents of just war naïvely think that government leaders can abide by moral principles; but this cannot happen for numerous reasons, including the pressures on the state to survive in a hostile and anarchic international environment. A second realist position, sometimes called "prescriptive realism," can be found in the work of Henry Kissinger and George Kennan. A typical prescriptive realist concedes that moral visions sometimes inform government leaders' decisions about war and peace, but this practice is both politically and ethically undesirable. Moral ardor (according to the prescriptive-realist argument) tends to foster bloody

crusades, in which self-righteousness blinds antagonists to possible truces and peaceful solutions. Calm calculations of enlightened self-interest, on the other hand, are more likely to guide state leaders along peaceful paths.

Many contemporary proponents of just war, including Michael Walzer,^[2] find all realists' arguments unconvincing. Historical events, such as the United States' decision to enter World War I, demonstrate that, despite the claims of descriptive realists, moral principles sometimes guide states' decisions about war and peace. The prescriptive-realist argument likewise rests on empirical overgeneralizations (or so just-war theorists contend). True, moral principles sometimes inspire crusades and thus contribute to international havoc; but sometimes, moral reasoning is sober and moderate, especially if the moral system places a premium on tolerance, cooperation, and peace (as the world's major ethical systems do). In other words, the content of a moral vision, and not moral reasoning per se, fosters either war or peace.

In addition, some writers who are sympathetic to just-war theory, such as Orend,^[1a] argue that the unrestrained pursuit of self-interest, untethered to moral principles, may contribute greater havoc to international politics than do moral principles. After all, national self-interest can justify colonialism, imperialism, genocide, and other blatantly unjust political arrangements that in the long run fuel wars. Without moral ballast, national egoism is dangerous (or so just-war theorists maintain).

Besides diverging from realist thought, just-war theorizing differs from several relatively popular ethical approaches to warfare that students of the history of political thought, such as Inis Claude, James Sterba, and Jennifer Teichman,^[3] sometimes call pacifism, militarism, and Christian realism. Twentieth-century pacifists—perhaps the most famous being Gandhi—believe that war is never morally permissible and fear that the concept of a just war provides a convenient cover for large-scale murder. Meanwhile, modern militarists, such as Theodore Roosevelt, have condoned warfare partly because combat both ennobles participants and contributes to the evolution of humanity by eliminating less vibrant people and cultures.^[4] According to this neo-Darwinist line of reasoning, war prevents the human species from stagnating, and just-war theorists,

underestimating the benefits of warfare, are wedded to overly soft-hearted and even effeminate notions of morality. War, in fact, hones important virtues and values, such as discipline and self-sacrifice for the good of the group. Finally, some Christian theologians with a deep commitment to love as a value, such as Reinhold Niebuhr, contend that the concept of a just war is a puzzling, incomplete, and muddled moral position that unsuccessfully blends pacifist and militarist thinking.^[5] For the Christian realist, what is crucial is that just war be reluctantly undertaken to promote peace, and with a spirit of dread about the use of violence, and forgiveness toward one's enemies.

Compared to pacifists, militarists, and advocates of Christian love and mercy, theorists of just war over the centuries have developed a unique moral position. Like pacifists, just-war theorists contend that intentional and large-scale killing is abhorrent and should be avoided. Yet, like militarists and Christian realists, just-war theorists insist that wars sometimes eliminate, or at least contain, significant evils. Therefore (argue the just-war theorists), war can be morally justifiable—despite the large-scale violence and suffering that is entailed—if it is undertaken for morally correct objectives, if it is undertaken with morally correct military methods, and if it ends in a morally appropriate fashion that promotes peace.

EVOLUTION OF JUST-WAR ARGUMENTS

The idea that governments should consider ideas of justice when deploying their armies is more than two millennia old. According to Thucydides,^[6] during the Peloponnesian War, the ancient Athenians exchanged opinions about justice in their assemblies and balanced moral and nonmoral concerns as they debated whether to go to war, whom to fight, and whether to accept peace proposals. To rally support for or against particular courses of action, speakers asked the gatherings whether a proposed military action—for instance, the invasion of a distant country or the execution of defeated enemies—was just. In the ensuing debate, different and sometimes divergent arguments about justice were aired. In the name of justice, some citizens counseled mercy and restraint, while others demanded violent retribution for past offenses and deserved deference from other peoples.

After Rome eclipsed the Greek city-state, discussions about the relationship between justice and warfare evolved in new directions. Church leaders, such as Origen and Lactantius, derived their understanding of justice from sacred text and initially argued that passages in the New Testament—such as Jesus' declarations to turn the other cheek and to love thy enemy—prescribed pacifism. Once Rome adopted

Christianity as its official religion, pacifism became more difficult to defend. The Roman Empire, after all, afforded Christians the modicum of social peace necessary to practice their religion in safety, and many enemies of Rome committed atrocities against the faithful.

Breaking decisively with the Church's initial pacifist position, St. Augustine articulated several themes and arguments that have informed just-war thinking for many centuries.^[7] Augustine argued that Jesus' precepts apply primarily to matters of one's heart. In the privacy of one's heart and home, a true Christian should, of course, unswervingly desire peace and wish to serve others. A Christian, nonetheless, sometimes must physically battle sinful rulers and foreign governments that threaten social peace and that make it difficult, if not impossible, for a Christian to live according to the commandments of the heart. Augustine hastened to add that his just-war doctrine did not mean that all of Rome's wars were justifiable. A government's systematic use of violence against another government is moral only if the ruler's heart is genuinely motivated by love of peace and social order, and not by earthly desires for fame, power, or material goods.

Augustine further argued that only political leaders are entitled to judge whether a war is morally justifiable. If every soldier and subject makes his or her own decision about the morality of a war, chaos might result, which violates the Christian goal of peace. Augustine also insisted that any ruler who is contemplating launching a just war must first explore methods short of armed conflict—such as diplomatic agreements—that might change the behavior of potential enemies. In addition, a declaration of war is just only if the ruler believes that the chances for success are high and that the anticipated amount of violence and suffering in the war will not exceed the expected benefits of victory. Finally, once the war commences, soldiers who are fighting on the side of justice must eschew morally heinous methods. Cruelty and bloodshed should never exceed the minimum needed to prevent future evil behavior on the part of the opponent. Augustine specified that soldiers must refrain from abusing captives. Torture, in particular, is strictly prohibited.

In the centuries following Augustine's death, the idea of a just war gradually became more secular, and proponents of just war increasingly based their moral reasoning not on sacred text but on the self-evident dignity of every human being. Thomas More's *Utopia* (1516), written at the height of the Renaissance, illustrates the newer, humanist views on just wars.

Utopia opens with Hythlodæus, a mysterious traveler, denouncing war in Europe because it is undertaken solely to enrich royal courts. Hythlodæus later

describes the military practices of a fictitious island called Utopia, whose citizens are morally upright despite the fact that they view Christianity as merely one of several beneficial theological systems. According to Hythlodæus, Utopians, believing in the equal and inherent dignity of every citizen, refrain from violence on their own island. Yet they are not pacifists. They go to war for three reasons: to protect their own island from imminent invasion, to aid allies who suffer injustices from intemperate foreign powers, and to rescue innocent persons and communities from arbitrary persecution and tyranny. Even when these conditions exist, the Utopians reluctantly go to war because of the suffering and pain that it entails. War is “an activity fit for beasts . . . they regard it with utter loathing.”^[8] Before taking up arms, the Utopians first try to influence foreign powers through diplomacy, economic incentives, and other non-violent methods. If these methods fail, then the Utopians go to war.

Wishing to fight in a just manner, Utopians avoid large battles and instead try to ambush the opposing army’s commanders. To win with as little bloodshed as possible, Utopian forces refuse to trample upon the crops of any country that they invade, preserve all cities, and respect all private property (except that owned by spies). The soldiers do so in the hopes that by leaving economic arrangements intact, they will drive a wedge between their enemy’s subjects and rulers. Once the hostilities end, the Utopians execute all enemy office holders who, until recently, steadfastly refused surrender and enslave all enemy soldiers who remained loyal to the regime. A detachment of Utopian officials then remains in the territory to observe local affairs and ward off future warlike behavior. To finance this detachment, to defray the costs of the recently completed war, and to finance future wars, the Utopians seize and manage a portion of the properties owned by the killed enemy officials and enslaved soldiers. Other confiscated property is transferred to those citizens within the enemy state who had courageously rebelled against their rulers.

During the 18th century (the age of the U.S. and French revolutions), the notion of a just war evolved in new directions and became associated with the spread of political rights. So-called “enlightenment” writers, such as William Godwin and Johann Herder, desired peace, but they also wished to reform the world. They therefore argued that a war was morally just if it spread liberal institutions and constitutional practices—such as private commercial property, voting rights for the middle classes, and the republican separation of the executive and legislative branches of government—to unenlightened nations.

Immanuel Kant, for instance, denounced war as “the greatest source of the evils which oppress civilized nations” in his essay on the origins of human

history,^[9] and yet insisted that warfare contributes to human progress by forcing despots to grant economic, cultural, and political freedoms to middle-class citizens. Kant further argued in a pamphlet about perpetual peace that to make the world safe for independent citizens, it may be necessary to create a worldwide military alliance of freedom-oriented states.^[10] In theory, each state in the alliance would refrain from attacking the others and would go to war for two reasons only: to protect its own borders from imminent attack, and to protect the liberties of self-directing citizens in other countries. Whenever the armies of the alliance defeat an enemy, the constitution of the vanquished country should be remade so that “every state shall be republican.”^[11] Kant, however, doubted that an international alliance would form in the foreseeable future (because current government leaders are insufficiently enlightened to curb their short-term appetites for territorial expansion and to subject themselves to an international alliance). So he proposed that for now, leaders of all constitutionally liberal states consult their citizens before each and every declaration of war because state leaders, seeking glory and gain, will naturally rush into battles. Citizens, on the other hand, understand the costs of war and will support it only when external threats to freedom are palpable and significant.

During the 20th century, just-war theory changed once more. Michael Walzer, arguably the most influential just-war scholar of the late 20th century, redefined justice in terms of the protection and promotion of individuals’ “rights to life and liberty,” including property.^[12] He argued that such human rights are now recognized worldwide and are beyond serious debate. Walzer also extended his understanding of rights to states, because without a government to provide for social security and stability, individuals would be unable to enjoy their rights to life and liberty.

According to Walzer, in an ideal world, all states would refrain from war, and individuals would be free to conduct domestic affairs as they wish, provided that they do not violate the life and liberty of fellow citizens. In the real world, however, some individuals attack the life and liberties of other citizens (hence, the moral need for the state), and some states oppress their citizens and invade and sack other states (hence, the moral need for just wars). A just war is one undertaken by a state first and foremost to protect, preserve, and restore individuals’ rights to life and liberty.

Drawing on earlier just-war theorizing, Walzer contended that a state is obligated to explore non-military methods of changing the behavior of a rogue state, to consider whether victory will significantly protect or restore of human rights, and to calculate the anticipated costs of a war against the evil that is to

be eliminated. If non-military options prove ineffective, if the likelihood of protecting or restoring individuals' rights appears high, and if the anticipated costs of war appear comparatively low, then a declaration of war is morally justified. Walzer also adopted the proviso, extending back to the writings of Augustine, that a just war must be conducted properly. Non-combatants should never be military targets, except under conditions of military exigency. Rape and pillage are prohibited.

JUST WAR IN PUBLIC POLICY

Since the mid-19th century, various Christian and secular theories of just war have influenced the course of international law and organization.^[13] The 1928 Kellogg-Briand Pact, for example, explicitly condemned war as an instrument of national policy, except for purposes of self-defense. Since 1864, numerous Geneva accords have attempted to control the brutality of war by prohibiting states from using biological and chemical weapons, from torturing prisoners, from intentionally attacking civilian targets, and from fielding soldiers under 15 years of age. The Red Cross, in particular, has played a key role in drafting and enforcing the pacts. Meanwhile, countries belonging to the United Nations have pledged to each other to desist from wars of aggression and to declare wars only to prevent humanitarian catastrophes, such as genocide, or to protect their citizens and borders from imminent attack.

The translation of just war from a moral theory into a set of practical policies has revealed vexing questions about the idea, which so far have received no satisfactory solutions.

First, should leaders of states be the only ones entitled to decide whether a war is morally justified? Or should an international body be given that authority? Walzer has endorsed the first, state-centered position, partly because he believes that groups with shallow moral commitments often manipulate international bodies. Kant, however, viewed state leaders as unreliable; he therefore recommended an international alliance as well as citizen consultations in all republics on the justness of contemplated wars.

Second, exactly which immoral situations, short of actual invasion, justify military actions? Does the refusal of a wealthy country to share its riches with an impoverished country, whose natural resources have been exploited by members of the wealthy country, provide grounds for military action to procure badly needed food? What about acts of genocide—such as the extermination of members of an ethnic minority—or gross human-rights violations—say, widespread use of torture—by an authoritarian regime?

Walzer, for one, has tended to describe genocide and severe and systematic civil-rights violations as sufficient reason to go to war, but has been wary about extending just-war arguments about human rights violations to economic exploitation and cultural discrimination. Still, the question remains open. Orend and Thomas Pogge,^[14] among others, believe that just-war reasoning may be relevant when state leaders contemplate responses to either wealthy countries' refusal to share wealth with less fortunate members of humankind or the systematic destruction of indigenous and minorities' cultures.

Third, once flagrant acts of injustice by a state have been legitimately identified, who exactly is entitled to go to war? Does every nation state have the right to unleash its own forces according to its leaders' own independent judgment? Must international sanction first be secured, and if so, how? Or should the military force in a just war always be broadly international to begin with, and not under the direction of a particular country or state?

Fourth, does the notion of the just conduct of a war make sense in an age of weapons of mass destruction? Should the use of landmines and aerial bombings (not to mention nuclear, chemical, and biological weapons) that indiscriminately hurt soldiers and civilian populations be judged a war crime? Can countries today refine their use of military technology (including nuclear weapons) so as to avoid needless civilian deaths, as James Turner Johnson recommends?^[15] Or, as Philip Acton suggests,^[16] does modern military technology make just-war arguments about sparing non-combatants outdated?

Finally, to ensure a lasting peace, does engagement in a just war require long-term occupation and extensive reconstruction of societies and culture? Is the notion of just war, as Kant suggested, inextricably linked to the difficult process of regime building?

CONCLUSIONS

As we have seen, the idea of just war has metamorphosed considerably over two millennia. This suggests that the idea is capable of further adaptation and variation. When put into practice, the concept raises difficult questions about authority (who is to determine whether a particular declaration of war is justified?), about ethical standards (does "justice" refer to the protection and promotion of Christian peace, to human rights, to republican constitutional arrangements, or to something else?), and about the indiscriminant destructiveness of modern weaponry. As proponents of just war wrestle with these practical problems, theories about the relevance of ethics to war undoubtedly will evolve in new directions.

INTERNET RESOURCES

For further information on the Geneva conventions and later protocols, see the Society of Professional Journalists' Guide to the Geneva Convention at: <http://www.globalissuesgroup.com/geneva/history.html>.

REFERENCES

1. (a) Orend, B. A just-war critique of realism and pacifism. *J. Philos. Res.* **2001**, 26, 441–454; (b) Donnelly, J. Twentieth century realism. In *Traditions in International Ethics*; Nardin, T., Mapel, D., Eds.; Cambridge University Press: Cambridge, 1992; 85–111.
2. Walzer, M. The triumph of just war theory (and the dangers of success). *Soc. Res.* **2002**, 69 (4), 925–944.
3. (a) Claude, I. Just wars: doctrines and institutions. *Polit. Sci. Q.* **1980**, 95(1), 83–96; (b) Sterba, J. Reconciling pacifists and just war theorists. *Soc. Theory Pract.* **1992**, 18 (1), 21–38; (c) Teichman, J. *Pacifism and the Just War*; Basil Blackwell: Oxford, 1986.
4. (a) Hofstadter, R. *Social Darwinism in American Thought*; Beacon Press: Boston, 1955; 170–200; (b) Roosevelt, T. The strenuous life. In *Theodore Roosevelt: An American Mind, A Selection from His Writings*; Penguin Books: New York, 1994; 184–189.
5. Flescher, A. Love and justice in Reinhold Niebuhr's prophetic Christian realism and Emmanuel Levinas's ethics of responsibility: treading between pacifism and just-war theory. *J. Relig.* **2000**, 80 (1), 61–82.
6. Thucydides. In *The Peloponnesian War*; Modern Library: New York, 1982.
7. (a) Deane, H. *The Political and Social Ideas of St. Augustine*; Columbia University Press: New York, 1964; (b) Ramsey, P. The just war according to St. Augustine. In *Just War Theory*; Elshtain, J., Ed.; New York University Press: New York, 1992; 8–22.
8. More, T. *Utopia*; Yale University Press: New Haven, 1964; 118 pp.
9. Kant, I. Conjectural beginning of human history. In *On History*; Bobbs-Merrill: Indianapolis, 1963; 66 pp.
10. Orend, B. Kant's just war theory. *J. Philos.* **1999**, 37 (2), 323–353.
11. Kant, I. Perpetual peace. In *Kant: Political Writings*; Cambridge University Press: Cambridge, 1991; 99 pp.
12. Walzer, M. *Just and Unjust Wars: A Moral Argument with Historical Illustrations*, 2nd Ed.; Basic Books: New York, 1992; 54 pp.
13. Luban, D. Just war and human rights. *Philos. Publ. Affairs* **1980**, 9 (2), 160–181.
14. (a) Orend, B. *Michael Walzer on War and Justice*; McGill-Queen's University Press: Montreal, 2000; 153–179; (b) Pogge, T. *World Poverty and Human Rights: Cosmopolitan Responsibilities and Reforms*; Polity Press: Cambridge, 2002.
15. Johnson, J. Threats, values, and defense: does the defense of values by force remain a moral responsibility? In *Just War Theory*; Elshtain, J., Ed.; New York University Press: New York, 1992; 55–76.
16. Acton, P. The just war tradition and the moral character of nuclear deterrence. *Polit. Stud.* **1991**, 39 (1), 5–18.

Justice through Emancipation: Abraham Lincoln as a Model of Justice

Frank J. Williams

Rhode Island Supreme Court, Providence, Rhode Island, U.S.A.

INTRODUCTION

The statue of Lady Justice, in her various forms, is the modern symbol of justice. Typically blindfolded, she holds a scale in one hand, while the other holds a sword. Statuettes or portraits of Lady Justice are a fixture in courthouses throughout the U.S.A. today, ostensibly because of the virtues she represents.

This article discusses Abraham Lincoln's moral objection to slavery, the issuance of the Emancipation Proclamation, and his role in the passage of the Thirteenth Amendment. His actions combined to create an historical model of justice. In the Proclamation itself, Lincoln maintained it was "an act of justice," and it has come to personify a human rights model of justice.

LINCOLN'S MORAL OBJECTION TO SLAVERY

An understanding of Lincoln's moral objection to slavery is fundamental to any discussion of the Emancipation Proclamation. Lincoln flatly opposed the "institution" by trumpeting the principles of natural law, which broadly called for the infusion of moral principles into government and law. In the political arena, he fought vigorously against the extension of slavery into the western territories.

As early as 1837—some 22 years before he was elected President—Lincoln maintained "[t]he institution of slavery is founded on both injustice and bad policy."^[1] Lincoln believed simply that slavery was wrong and that "like every other wrong which some men commit if left alone, it ought to be prohibited by law."^[2]

He formally built his case against slavery during the course of his campaign in 1858 against Stephen A. Douglas for the US Senate, campaigning on the following moral ground while directly attacking Douglas's advocacy of "popular sovereignty:"

When [Douglas] invites any people willing to have slavery, to establish it, he is blowing out the moral lights around us. When he says that he "cares not whether slavery is voted down or voted up,"—that

it is a sacred right of self-government—he is in my judgment penetrating the human soul and eradicating the light of reason and the love of liberty in this American people.^[3]

Lincoln, quoting his own 1854 printed speech, continued in this vein during the first of their famous debates:

This declared indifference, but as I must think, covert real zeal for the spread of slavery, I cannot but hate. I hate it because of the monstrous injustice of slavery itself. I hate it because it deprives our republican example of its just influence in the world—enables the enemies of free institutions, to taunt us as hypocrites—causes the real friends of freedom to doubt our sincerity, and especially because it forces so many good men amongst ourselves into an open war with the very fundamental principles of civil liberty—criticizing the Declaration of Independence, and insisting that there is no right principle of action but self-interest.^[4]

His courageous stance on the hot button issue of the day was far from politically expedient; he eventually lost that election.

To be sure, Lincoln's stance on slavery was not without its flaws when measured by modern standards. For much of his political life, Lincoln had advanced the concept of colonization, which would have allowed freed slaves to return to Africa. Colonization, as a political solution to the problems presented by emancipation, appealed to him—as it did to his political hero, Henry Clay—because of the "moral fitness in the idea of returning to Africa her children, whose ancestors have been torn away from her by the ruthless hand of fraud and violence."^[5] His support for *voluntary* colonization diminished during his presidency; he said nothing about it publicly after the issuance of the Emancipation Proclamation.

Lincoln's opposition to slavery, however, never waned. As President, Lincoln wrote in a letter to former US Senator Albert G. Hodges of Kentucky: "I am naturally anti-slavery. If slavery is not wrong, nothing is wrong. I can not [sic] remember when I did not so think, and feel."^[6] To the end, Lincoln steadfastly illuminated the moral light through his opposition to slavery.

CONSTITUTIONAL IMPEDIMENTS

The US Constitution conflicted with Lincoln's strong moral beliefs, namely its so-called three-fifths clause. Addressing how taxation and membership of the House of Representatives would be apportioned, it included the phrase "three-fifths of all other Persons." At the very least, this provision recognized slavery as an institution and the reality that the slaves were in some respect human and, in another respect, mere property.

During the seventh debate with Douglas, Lincoln suggested a way to reconcile the end of slavery with the Constitution:

The real issue in this controversy—the one pressing upon every mind—is the sentiment on the part of one class that looks upon the institution of slavery *as a wrong*, and of another class that *does not* look upon it as a wrong. The sentiment that contemplates the institution of slavery as being wrong, is the sentiment of the Republican party . . . They look upon slavery as a moral, social and political wrong; and while they contemplate it as such, they nevertheless have due regard for its actual existence among us, and the difficulties of getting rid of it in any satisfactory way and to all the constitutional obligations thrown about it. Yet having a due regard for these, they desire to see a policy in regard to it that looks to its not creating anymore danger. They insist that it should as far as may be, *be treated* as a wrong, and one of the methods of treating it as a wrong is to *make provision that it shall grow no larger* . . . I have said, and I repeat it here, that if there be any man amongst us who does not think that the institution of slavery is wrong in any one of the aspects in which I have spoken, he is misplaced, and ought not to be with us. And if there is a man amongst us who is so impatient of it as a wrong as to disregard its actual presence among us and the difficulty of getting rid of it suddenly in a satisfactory way, and to disregard the constitutional obligations thrown about it, that man is misplaced if he is on our platform. We disclaim sympathy with him in practical action. He is not placed properly with us.^[7]

The extremes within this quote display Lincoln's political savvy. The first half pays homage to the abolitionists; the second half makes amends for the abolitionists with the South, demanding that those who stand with him take into account, among other factors, the requirements of the Constitution.

The US Supreme Court's infamous Dred Scott decision in 1857 further strengthened the constitutional support for the institution of slavery. Written by Chief Justice Roger Taney, the decision held that blacks were not citizens of the U.S.A. and Congress had no

constitutional power to prohibit slavery in the federal territories.^[8] The opinion went even further and proclaimed that blacks "had no rights which the white man was bound to respect."^[9] Although Lincoln vehemently disagreed with this opinion, he nonetheless recognized the law of the nation when he assumed the presidency in 1860.

Lincoln, however, found support for his antislavery position in the Declaration of Independence. While in Philadelphia on his way to his inauguration in 1860, he reiterated his opposition to slavery and his hope for the preservation of the Union:

It was not the mere separation of the colonies from the mother land; but something in that Declaration giving liberty, not alone to the people of this country but hope to all the world, for all future time. It was that which gave promise that in due time the weights would be lifted from the shoulders of all men, and that all should have an equal chance. This is the sentiment embodied in the Declaration of Independence. Now, my friends, can this country be saved on that basis? If it can, I will consider myself one of the happiest men in the world if I can help to save it. If it cannot be saved upon that principle, it will be truly awful. But, if this country cannot be saved without giving up that principle, I was about to say I would rather be assassinated on this spot than surrender it.^[10]

The distinction Lincoln drew between the Constitution and Declaration of Independence is akin to the primary distinction between law and justice. While law is merely an assemblage of rules that comes to us from the courts or through the enactments of the legislature, justice is based on the integral relationship between people and these rules. When justice is realized, reconsecration of faith in government, community, and the rule of law ensues. The belief that all men are created equal is a basic precept of justice because it resonates universally. That sentiment would become a powerful undercurrent in Lincoln's wartime management of slavery.

THE EMANCIPATION PROCLAMATION

Lincoln's primary objective during his presidency was the preservation of the Union by quelling the rebellion of the southern states. As evidenced by Lady Justice's balancing and weighing of interests, justice is a perpetual search for equilibrium. Lincoln, as President of the U.S.A. in a time of civil war, was in the unique position of determining the point at which it could be established. Ever the resourceful politician-lawyer, he was able to strike at slavery to win the war.

At the outset of his presidency, Lincoln was careful to draw a distinction between his own values and his obligations under the Constitution:

I am naturally anti-slavery. If slavery is not wrong, nothing is wrong. I cannot remember when I did not think so, and feel. And yet I have never understood that the Presidency conferred upon me an unrestricted right to act officially upon the judgment and feeling. It was in the oath I took that I would, to the best of my ability, preserve, protect, and defend the Constitution of the United States. I could not take the office without taking the oath. Nor was it my view that I might take an oath to get power and break the oath in using the power. I understood, too, that an ordinary civil administration this oath even forbade me to practically indulge any primary abstract judgment on the moral question of slavery.^[6]

Lincoln, therefore, came slowly to the idea of emancipating slaves by executive order. For example, his annual message to Congress in 1862 suggested an amendment to the Constitution that would end slavery gradually by 1900. By mid-summer 1862, however, Lincoln was discussing openly with several of his cabinet members the emancipation of slaves. In the middle of the war, Lincoln became convinced he could pursue the goals of emancipation and preservation of the Union simultaneously through the exercise of his presidential powers during wartime.

The Emancipation Proclamation had two manifestations. The preliminary Proclamation on September 22, 1862, declared that the slaves in areas still in rebellion against the U.S.A. on January 1, 1863, 100 days hence, would be then and forever free. It also directed the military not to enforce the fugitive slave laws that required their return.

On January 1, 1863, Abraham Lincoln signed the final Emancipation Proclamation (Appendix). It freed “henceforth and forever” slaves “within any state or designated part of a State whose people . . . shall then be in rebellion against the United States.” Lincoln based the authority of his proclamation on his power as the Commander-in-Chief and specifically termed the measure “a fit and necessary war measure” to suppress the rebellion. He termed it “an act of justice.”

The Emancipation Proclamation demonstrates Lincoln’s concentric grasp of legal and political issues. The Constitution authorized the Congress or the President to confiscate property only with proper compensation. When the preliminary Emancipation Proclamation was issued, the legal basis of the action seemed obscure. When the time came for the final Emancipation Proclamation on January 1, 1863, Lincoln determined that his act was a war measure taken as Commander-in-Chief to weaken the enemy. Thus,

justification for emancipation rested on the war power. His interpretation of the law of war was correct. According to Henry Wheaton’s contemporary treatise on international law, “[f]rom the moment one State is at war with another, it has, on general principles, a right to seize on all the enemy’s property, of whatsoever kind and wheresoever found, and to appropriate the property thus taken to its own use, or to that of the captors.”^[11]

In his final Emancipation Proclamation, Lincoln expressly exempted parts of the South by mentioning particular states and parts of states that were already under the control of the Union army. Always mindful of the constitutional limits on his own power, Lincoln believed his power to wage war did not extend into regions of the former Confederacy that were already in the process of returning to the Union, despite the fact that slavery was still an acceptable institution in those regions. In hindsight, emancipating only the slaves in certain areas may seem like an unacceptable compromise that falls short of the justice of complete emancipation. A modern-day author defines “immoral justice” as “a resolution that makes sense legally and can be justified by judges, lawyers, and law professors simply by conforming, in a very narrow formalistic sense, to precedent and procedure, but ultimately feels emotionally wrong to everyone else.”^[12] The shortcomings of the Emancipation Proclamation, however, would prove to be temporary.

THE THIRTEENTH AMENDMENT

Although the Thirteenth Amendment, stating that “[n]either slavery nor involuntary servitude, except as a punishment for crime . . . shall exist within the United States,” did not become law until after his death, Lincoln also was instrumental in amending the Constitution and abolishing slavery in its entirety. Understanding that the Emancipation Proclamation was a wartime measure, he sought a more permanent determination that would end slavery throughout the country. Even before his 1864 reelection, Lincoln was the amendment’s foremost ally and “[n]o piece of legislation during Lincoln’s presidency received more of his attention than the Thirteenth Amendment.”^[13] Although some have argued that his means may have been dubious, he deserves recognition and praise: “Whereas other political leaders of his day, both home and abroad, fixated only on Machiavellian means and ends, Lincoln practiced the pure art of politics within acceptable *moral bounds*.”^[14] As the next logical step after the Emancipation Proclamation, Lincoln refused to accept the wartime solution of limited emancipation and ultimately achieved the morally just result in the Thirteenth Amendment. In doing so, he solidified his place in American history as the Great Emancipator.

CONCLUSIONS

Secondary to the Declaration of Independence, the Emancipation Proclamation is the most significant statement of moral policy issued by the US government in its history. After nearly two years of armed rebellion, President Abraham Lincoln proclaimed on January 1, 1863 that the slaves in eight of those states, and in large portions of two others, were free people and that the US government and its armed forces would protect their freedom. Despite its enormous significance, the document's legalese, written as a war measure by the Commander-in-Chief to survive potential challenges in court, could not conceal the deeper meaning and purpose of the Emancipation Proclamation—the end of American slavery. The emancipation of slaves and the ultimate abolition of slavery nationwide became the just result.

It is not only the ends, but also the means in the issuance of the proclamation that provide the Lincoln model of justice. Despite his strong moral beliefs, he recognized the constitutional recognition of slavery and only acted in the scope of his wartime power in an effort to quell the rebellion and preserve the Union. He resisted the temptation to act beyond his constitutional powers and free all slaves throughout the nation. Although he would not live to see it, he ardently supported the Thirteenth Amendment and ensured Congressional support for it. In this sense, Lincoln embodies blind justice. He weighs the impact of emancipation on the effort to preserve the Union in one hand—while holding the sword of presidential power in wartime in the other. As a justified and necessary military action, Lincoln's Emancipation Proclamation is one of the single greatest acts of justice in American history.

But Lincoln is more than just the Great Emancipator. He embodies many other virtues that exemplify the foundations of American democracy: commitment, character, leadership, integrity, and accountability. Beyond these virtues, however, Lincoln's efforts to end the rampant injustice of slavery make him the single greatest historical figure of American justice.

ACKNOWLEDGMENTS

The author would like to thank Professor William D. Pederson and Kyle Zambarano, Esq., for their thoughtful contributions.

APPENDIX

By the President of the United States of America:

A PROCLAMATION.

Whereas, on the twenty-second day of September, in the year of our Lord one thousand eight hundred and

sixty-two, a proclamation was issued by the President of the United States, containing, among other things, the following, to wit:

“That on the first day of January, in the year of our Lord one thousand eight hundred and sixty-three, all persons held as slaves within any State or designated part of a State, the people whereof shall then be in rebellion against the United States, shall be then, thenceforward, and forever free; and the Executive Government of the United States, including the military and naval authority thereof, will recognize and maintain the freedom of such persons, and will do no act or acts to repress such persons, or any of them, in any efforts they may make for their actual freedom.

“That the Executive will, on the first day of January aforesaid, by proclamation, designate the States, and parts of States, if any, in which the people thereof, respectively, shall then be in rebellion against the United States; and the fact that any State, or the people thereof, shall on that day be, in good faith, represented in the Congress of the United States by members chosen thereto at elections wherein a majority of the qualified voters of such State shall have participated, shall, in the absence of strong countervailing testimony, be deemed conclusive evidence that such State, and the people thereof, are not then in rebellion against the United States.”

Now, therefore, I ABRAHAM LINCOLN, President of the United States, by virtue of the power in me vested as Commander-in-Chief, of the Army and Navy of the United States in time of actual armed rebellion against authority and government of the United States, and as a fit and necessary war measure for suppressing said rebellion, do, on this first day of January, in the year of our Lord eighteen hundred and sixty-three, and in accordance with my purpose so to do publicly proclaimed for the full period of one hundred days from the day first above mentioned, order and designate as the States, and parts of States, wherein the people thereof respectively, are this day in rebellion against the United States, the following, to wit [sic]:

Arkansas, Texas, Louisiana, (except the parishes of St. Bernard, Plaquemines, Jefferson, St. John, St. Charles, St. James, Ascension, Assumption, Terre Bonne, Lafourche, St. Mary, St. Martin, and Orleans, including the city of New Orleans) Mississippi, Alabama, Florida, Georgia, South-Carolina, North-Carolina, and Virginia, (except the forty-eight counties designated as West Virginia, and also the counties of Berkely, Accomac, Northampton, Elizabeth-City, York, Princess Ann, and Norfolk, including the cities of Norfolk & Portsmouth); and which excepted parts are for the present, left precisely as if this proclamation were not issued.

And by virtues of the power, and for the purpose aforesaid, I do order and declare that all persons held as slaves within said designated States, and parts of States, are, and henceforward shall be, free; and that the Executive Government of the United States, including the military and naval authorities thereof, will recognize and maintain the freedom of said persons.

And I do hereby enjoin upon the people so declared to be free to abstain from all violence, unless in necessary self-defence; and I recommend to them that, in all cases when allowed, they labor faithfully for reasonable wages.

And I further declare and make known, that such persons of suitable condition, will be received into the armed service of the United States to garrison forts, positions, stations, and other places, and to man vessels of all sorts in said service.

And upon this act, sincerely believed to be an act of justice, warranted by the Constitution, upon military necessity, I invoke the considerate judgment of mankind and the gracious favor of Almighty God.

In witness whereof, I have hereunto set my hand and caused the seal of the United States to be affixed.

Done at the City of Washington, this first day of January, in the year of our Lord one thousand eight hundred and [L.S.] sixty-three, and of the Independence of the United States the eighty-seventh.

ABRAHAM LINCOLN.

By the President:
William H. Seward,
Secretary of State.

REFERENCES

1. Kerner, F. *A Treasury of Lincoln Quotations*; American House, Inc.: Chicago, Illinois, 1965; 255 pp.
2. Kerner, F., Ed.; *A Treasury of Lincoln Quotations*; American House, Inc.: Chicago, Illinois, 1965; 260 pp.
3. Basler, R.P., Ed.; *The Collected Works of Abraham Lincoln*; Rutgers University Press: New Brunswick, New Jersey, 1953; Vol. III, 29 pp.
4. Basler, R.P., Ed.; *The Collected Works of Abraham Lincoln*; Rutgers University Press: New Brunswick, New Jersey, 1953; Vol. III, 14 pp.
5. Basler, R.P., Ed.; *The Collected Works of Abraham Lincoln*; Rutgers University Press: New Brunswick, New Jersey, 1953; Vol. II, 132 pp.
6. Basler, R.P., Ed.; *The Collected Works of Abraham Lincoln*; Rutgers University Press: New Brunswick, New Jersey, 1953; Vol. VII, 281 pp.
7. Basler, R.P., Ed.; *The Collected Works of Abraham Lincoln*; Rutgers University Press: New Brunswick, New Jersey, 1953; Vol. III, 312–313.
8. Fehrenbacher, D.E. *The Dred Scott Case: Its Significance in American Law and Politics*; Oxford University Press: New York, 1978; 323 pp.
9. Fehrenbacher, D.E. *The Dred Scott Case: Its Significance in American Law and Politics*; Oxford University Press: New York, 2005; 347 pp.
10. Basler, R.P., Ed.; *The Collected Works of Abraham Lincoln*; Rutgers University Press: New Brunswick, New Jersey, 1953; Vol. IV, 240 pp.
11. Wheaton, H. *Elements of International Law*, 4th Ed.; Stevens and Sons: London, 1904; 480 pp.
12. Rosenbaum, T. *The Myth of Moral Justice*; Harper Collins: New York, 2004; 19 pp.
13. Vorenberg, M. *Final Freedom: The Civil War, the Abolition of Slavery, and the Thirteenth Amendment*; Cambridge University Press: New York, 2001; 180 pp.
14. Williams, F.J. "The End of Slavery: Lincoln and the Thirteenth Amendment—What Did He Know and When Did He Know It?" *Judging Lincoln*; Southern Illinois University Press: Carbondale and Edwardsville, 2002; 143 pp.

Juvenile Stalking

Deborah L. Laufersweiler-Dwyer

*Department of Criminal Justice, University of Arkansas at Little Rock,
Little Rock, Arkansas, U.S.A.*

R. Gregg Dwyer

*Department of Neuropsychiatry and Behavioral Science, University of South Carolina
School of Medicine, Columbia, South Carolina, U.S.A.*

INTRODUCTION

Stalking by adults has been the focus of public agencies for the past several decades. It was not until recently that the social problem of stalking by children and adolescents has been identified as a significant social issue for clinicians, law enforcement, and policy makers. A phenomenon that garnered significant attention during the late 20th century, stalking is again gathering strength as the focus turns from adults to the behaviors of children and adolescents. As with many other phenomena, it was only after the death of several celebrities that serious research by clinicians and practitioners was undertaken to develop typologies, assessment methods, medical protocols, and legal responses to offenders and victims. But during this time, as has often been the case with criminal justice research, the focus was entirely on adults. It was not until the year 2000 that the first empirical study was conducted that described the characteristics of child and adolescent obsessional followers.^[1] What has occurred since then has been a rise in both the academic and the public attention to this problem. Many of the juvenile issues such as bullying, school violence, and dating violence have led researchers to take a closer look at juvenile stalking.

MEDICAL DEFINITION

Definitions for stalking are often divided into three overlapping categories to include medical, behavioral, and legal. Although there is no medical definition of stalking, the diagnosis that is most closely associated with the behavior is Delusional Disorder, Erotomanic Type as found in the *Diagnostic and Statistical Manual of Mental Disorders 4th Edition Text Revision (DSM-IV-TR)*. To be defined as a delusional disorder by *DSM-IV-TR*, the behavior must be nonbizarre, persist for at least 1 month, never have met criteria for schizophrenia, not include prominent auditory or visual

hallucinations, if present tactile or olfactory hallucinations must be consistent with the delusion, except as related to the delusion psychosocial functioning must be intact, if mood symptoms are present they must be brief relative to the delusion's duration, and the etiology cannot be from substance use or a general medical condition.^[2] The term "bizarre" in this context means that the delusion is implausible, cannot be understood, or has no connection to "ordinary life experiences." The subtype known as erotomanic type refers to delusions in which the person believes someone else is in love with him/her. The other person typically holds a higher status than the one with the delusion and the "love" is frequently "romantic" or "spiritual" rather than sexual. The delusional person may or may not attempt contact. These attempts at contact may lead to the label of "stalking" and result in law enforcement involvement.^[2]

BEHAVIORAL DEFINITIONS

The behavioral definitions associated with stalking offer a dissimilar approach in that they are not based on the analysis of the actions as described in the *DSM-IV-TR*. Meloy and Gothard proposed the term "obsessional follower" as a clinical corollary of "stalker."^[3] In an extensive review of the literature, they found that earlier works used very narrow definitions for the term "stalking." Most of society had connected the word stalking with the various sensationalized cases that had commanded popular media attention. By expanding the definition and including the term obsessional follower, Meloy included individuals who engage in stalking (as per the technical definition) as well as persons who engage in deviant or long-term patterns of threat or harassment directed toward specific individuals.^[4] Others involved in the research of stalking choose to steer away from the word "stalking" because it implies following behaviors. Instead, these researchers chose to identify the

behavior as obsessional harassment. This includes the many other tactics used by individuals such as telephone contact, mail, or other methods of communication as opposed to face-to-face contact.^[5] This definition also includes the repetitive nature of the perpetrators preoccupations as well as the victims' perceptions.

Some researchers use subtypes of stalking to include erotomania (the delusional disorder), the paranoid type of schizophrenia, and personality disorders. In addition to identifying the types of stalkers, researchers have also identified types of victims. Meloy distinguishes between two groups of targets, private and public.^[6] This distinction between victim types is significant in that as public awareness grows, the misconception that stalking is only a problem of the rich and famous can be corrected. For many the definition of stalking had been relegated to include only public figures with other behavioral definitions being applied to private individuals. By attaching the definition of stalking to both groups, researchers as well as policy makers have made significant advances in helping the public understand what behaviors constitute obsessional following.

Cupach and Spitzberg analyzed stalking as an obsessive relational intrusion.^[7] Their research used the definition: "repeated unwanted pursuit and invasion of one's sense of physical or symbolic privacy by another person who desires or presumes an intimate relationship" (on pages 234–235 of Ref.^[7]). A key component of this definition is that the goals of the parties are not mutual. In addition, Cupach and Spitzberg identified grades of relational intrusion by placing the acts along a continuum.^[7] This helped provide a subjective boundary between reasonable and unreasonable intrusion in relationships, which is particularly helpful in applying the definitions to the actions of children and adolescents. Much of the previous research has created significant problems in applying adult statutes to juveniles. Unlike statutes such as murder where the behavior is clearly identifiable, many of the behaviors exhibited by children are "normal" adolescent behaviors. The continuum established by Cupach and Spitzberg helps to differentiate between normal behavior and abnormal or problem behaviors.^[7]

The most recent typology is based on relationship between stalker and victim and context of the stalking. Known as RECON, this typology has four categories and was tested on a sample of 1005 stalkers. It is the largest nonrandom sample of stalkers ever studied.^[8] Although juveniles were included in the sample, the average age was 36 years. This model should be considered for future research in juvenile stalking. This particular model lends itself well to practical application by law enforcement and mental health practitioners (Fig. 1).^[8]

Author(s)	Name	Characteristics
Zona et al (11)	Simple Obsessional	most common type; have some prior relationship with the victim, usually an intimate one; most often occur in the context of domestic violence
	Love Obsessional Stalkers	no existing relationship with the victim; many target celebrities
	Erotomaniac Stalkers	delusionally believe that they are loved by the victim; rarest category
	Individual perpetrators may not precisely fit any single stalker category, and often exhibit characteristics associated with more than one category; it is important to remember that these typologies are merely guides.	
Mullen, Pathé and Purcell (18)	Rejected	pursue ex-intimates, either in the hope of reconciliation, or for vengeance, or out of a mixture of both
	Intimacy Seekers	who stalk someone they believe they love and who they think will reciprocate
	Incompetent Suitors	inappropriately intrude on someone, usually seeking a date or brief sexual encounter
	Resentful	pursue victims to exact revenge for some actual or perceived injury
	Predatory	stalking forms part of sexual offending
Mohandie, et. al (8)	Intimate	most likely to be violent, rarely psychotic, escalate in behaviors,
	Acquaintance	persistent, strong desire to start a relationship, law enforcement and mental health should be involved in response to these offenders
	Public Figure	largest group in sample, these offenders are older, less likely to escalate, least likely to threaten, most had a history of mental illness
	Private Stranger	majority of victims are women, more likely to "just want to communicate," high frequency of mental illness, and are less likely to assault the target.

Fig. 1 Behavioral models of stalking.

LEGAL DEFINITIONS

Legal definitions begin by creating a distinction between behaviors that are simply part of a criminal's desire to obtain opportunity and those behaviors that the concept of stalking truly encompasses. California was the first state to enact legislation establishing stalking as a criminal act. By 1993, 48 states and the federal government had created some form of antistalking legislation.^[9] All states have now enacted legislation making stalking illegal.^[10]

Legal classifications of stalking behavior generally focus on the stalker–victim relationship. Perhaps the most frequently referred to is that of Zona, Sharma, and Lane which is based on their study of 74 cases managed by the Los Angeles police.^[11,12] They classify stalkers into a simple obsessional group, where the stalker and victim had a prior relationship; a love obsessional group, where there was no prior relationship; and an erotomaniac group, where the stalker developed a delusion that the victim was in love with him or her. The simple obsessional group is most common and reported to be the most likely to resort to violence.^[13,14] Celebrity stalkers are usually found in the love obsessional group, a high proportion of who

suffer from psychotic illnesses. Those in the erotomaniac group are usually women, with their victims being older men of higher socioeconomic status.^[15]

Although this classification is useful, it has its limitations. In particular, it largely ignores the offender's motivation. It is also not clear how reliably the last two groups can be distinguished, how well the typology differentiates between persistent courtship and stalking, and whether it has any predictive validity.

Most legal definitions have several elements that must be met in order for an act to be classified as stalking. Two chief elements found in most antistalking legislation are threat and criminal intent. Threatening behaviors encompass a variety of acts, including harassment, repeated phone calls or letters, and threats to the victim's family members.^[16] These behaviors are most often manifested in the form of trespassing, displaying a weapon, vandalizing the victim's property, making threats of bodily harm, intimidating the victim and his or her family, and surveillance of the victim. Beyond the simple act, most statutes require police to show that the behavior is ongoing or repeated and that these acts also pose a credible threat to the victim. In establishing a credible case for arrest and prosecution of offenders, criminal justice professionals must establish criminal intent. This aspect of the law is very clear. In order to arrest and prosecute an individual for stalking, it must be shown that the perpetrator's acts meet the standard of intent by being purposeful, intentional, willful, or knowing.^[9]

POLICY ISSUES

A principle question posed to clinicians by policy makers has been "Who stalks?" The answer is needed by legislators to formulate policy and develop remedies to facilitate primary prevention strategies, manage identified offenders, and aid victims. Professionals from various specialties agree that the effective management of stalking cases involves a combination of legal sanctions and mental health intervention. Many clinicians have suggested a multidisciplinary approach to case management that involves clinical intervention and criminal sanctions for the obsessional followers as well as the use of civil orders of protection for the victim.^[12,17] A delicate balance exists between the constituencies calling for new legislation or the reevaluation of existing laws for the protection of stalking victims and the directions of reform that are frequently dictated by behavioral science research. One significant issue that has been placed before policy makers has been driven by the Violence Against Women Grant Office. This office has concluded that in most cases stalkers do not threaten their victims verbally or in writing; therefore, the credible threat

requirements should be eliminated from antistalking statutes making it easier to prosecute cases.^[18] Although perhaps applying to adults, this suggestion is not supported by empirical research for children. Meloy showed that 50–75% of obsessional followers threatened their victims.^[19] McCann found that among children and adolescents the percentage of perpetrators who threatened their victim was just over 50%.^[11] This is particularly important in designing policy that establishes mandatory reporting requirements. Unlike instances of sexual or physical abuse, where reporting criteria are clear, many actions by the obsessional follower may be misinterpreted or simply overlooked due to the age of the offender and victim.

CONCLUSION

We believe stalking is best understood through a multidimensional model involving the relationship of the stalker to the victim, the motivation of the stalker, and the severity of the stalking behavior. Such an approach provides a basis for risk assessment. By developing policy and procedures using comments from each of the domains discussed, medical, behavioral, and legal, administrators can create multidiscipline response to a public health and safety issue. In this way, both victim and offender are addressed, resources of multiple agencies are pooled to converse costs, and expertise possessed by different professions is utilized. By creating a multiagency response from policy inception, one can hope to avoid interagency conflicts regarding jurisdiction and responsibility and promote cooperation.

Much work is needed to develop a comprehensive and yet practical approach to managing juvenile stalking behaviors. Until research is able to fill the knowledge void, administrators must rely on extrapolating from data derived from studying adults and the study of similar behaviors in juveniles, such as general violence and sexual offending behaviors, to generate policy.

This article provides a starting point for administrators faced with developing a response to juvenile stalking behavior or at least coming to a basic understanding of its etiology and trajectory. Given limited research with rigorous methodologies, we approached the issue by examining how the fields of medicine, behavioral science, and law have defined the behavior in question as presented in the recent literature. Although not a comprehensive treatment of the topic given the limits of such an article, we have nonetheless laid the foundation for policy makers to begin formulating a strategy for addressing an important public health and safety concern.

REFERENCES

1. McCann, J.T. A descriptive study of child and adolescent obsessional followers. *J. Forensic Sci.* **2000**, *45* (1), 195–199.
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 4th. American Psychiatric Association: Washington, DC, 2001. Text Revision (DSM-IV-TR).
3. Meloy, J.R.; Gothard, S. A demographic and clinical comparison of obsessional followers and offenders with mental disorders. *Am. J. Psychiat.* **1995**, *152*, 258–263.
4. Meloy, J.R. Stalking (obsessional following): a review of some preliminary studies. *Aggress. Violent Behav.* **1996**, *1*, 147–162.
5. Rosenfeld, B. Assessment and treatment of obsessional harassment. *Aggress. Violent Behav.* **2000**, *5* (6), 529–549.
6. Meloy, J.R. Communicated threats and violence toward public and private targets: discerning differences among those who stalk and attack. *J. Forensic Sci.* **2000**, *46*, 1043–1052.
7. Cupach, W.R.; Spitzberg, B.H. Obsessive relational intrusions and stalking. In *The Dark Side of Close Relationships*; Spitzberg, B.H., Cupach, W.R., Eds.; Lawrence Erlbaum: Mahwah, NJ, 1998; 233–263.
8. Mohandie, K.; Meloy, J.R.; McGowan, M.G.; Williams, J. The RECON typology of stalking: reliability and validity based upon a large sample of North American stalkers. *J. Forensic Sci.* **2006**, *51* (1), 147–155.
9. National Institute of Justice. *Project to Develop a Model Anti-Stalking Code for States*; NIJ Publication No. NCJ 144477; U.S. Government Printing Office: Washington, DC, 1993.
10. Bradfield, J.L. Anti-stalking laws: Do they adequately protect stalking victims? *Harvard Woman's Law J.* **1998**, *21*, 229–266.
11. Zona, M.; Sharma, K.; Lane, J. A comparative study of erotomaniac and obsessional subjects in a forensic sample. *J. Forensic Sci.* **1993**, *38*, 894–903.
12. Meloy, J.R. The clinical risk management of stalking: “someone is watching over me.” *Am. J. Psychother.* **1997**, *51*, 174–184.
13. Kienlen, K.K.; Birmingham, D.L.; Solberg, K.B.; O'Regan, J.T.; Meloy, J.R. A comparative study of psychotic and non-psychotic stalking. *J. Am. Acad. Psychiat. Law* **1999**, *25*, 317–334.
14. Schwartz-Watts, D.; Morgan, D.W. Violent versus nonviolent stalkers. *J. Am. Acad. Psychiat. Law* **1998**, *26*, 241–245.
15. Nadkarni, R.; Grubin, D. Stalking: Why do people do it? *BMJ* **2000**, *320* (7248), 1486–1487.
16. McAnaney, K.G.; Curliss, L.A.; Abeyta-Price, C.E. From imprudence to crime: anti-stalking laws. *Notre Dame Law Rev.* **1993**, *68*, 819–909.
17. Mullen, P.E.; Pathe, M.; Purcell, R.; Stuart, G.W. Study of stalkers. *Am. J. Psychiat.* **1999**, *156*, 1244–1249.
18. Mullen, P.E.; Pathe, M.; Purcell, R. The management of stalkers. *Adv. Psychiatric Treatment* **2001**, *7* (5), 335–342.
19. Meloy, J.R. Stalking: an old behavior, a new crime. *Psychiatric Clin. North Am.* **1999**, *22*, 85–99.

Juvenile Waiver

Craig Hemmens

Department of Criminal Justice Administration, Boise State University, Boise, Idaho, U.S.A.

Benjamin Steiner

Division of Criminal Justice, University of Cincinnati, Cincinnati, Ohio, U.S.A.

INTRODUCTION

The juvenile court was created as an alternate method of dealing with juveniles who ran afoul of the law. Early juvenile court advocates argued that juveniles were less culpable than adult offenders, and thus should be dealt with differently. Dealing with juveniles in a separate court would allow for more attention to the individual needs of the juvenile. This approach differed radically from the punishment-oriented approach of the criminal court system.

By the later part of the 20th century, however, a number of constituencies had become disenchanted with the juvenile court. Conservatives argued that it allowed juveniles who had committed serious crimes to escape with little or no punishment; liberals argued that the juvenile court disregarded the constitutional rights of juvenile offenders. The result was an increase in the use of juvenile waiver. Juvenile waiver is the mechanism by which a juvenile court's jurisdiction is waived, and a juvenile offender is transferred to the jurisdiction of the criminal court. This allows for the imposition of different (and generally greater) sanctions on the juvenile, should he or she be convicted in criminal court.

THE JUVENILE COURT

At common law, persons over the age of 14 years were considered fully responsible for their actions. Children under the age of 7 years were considered incapable of felonious intent, while children between the ages of 7 and 14 years were considered similarly incapable unless it was determined that the child understood the consequences of his/her actions. This was known as the "infancy defense."^[1]

At common law, juvenile offenders received the same punishment as adult offenders and were usually housed in the same facilities. Before the establishment of New York's House of Refuge in 1825, no state bothered to separate children from adults in prison. Soon thereafter, other state reform schools were opened. By 1899 there were 65 facilities for juveniles

in the United States, but juvenile offenders still received the same punishment as adults.^[2]

Several events contributed to creation of a separate juvenile justice system. The Industrial Revolution of the late 19th century transformed America from a rural country to an urban nation. The number of children living in cities increased rapidly, and juvenile delinquency became a problem in many cities. Police and school administrators became concerned about the welfare of these urban children, while others became concerned that the increase in immigrants from Europe was creating social disorganization. The Progressive movement combined these concerns to produce wide-ranging social reforms, including the creation of a separate juvenile court system.

Proponents of a separate juvenile justice system believed that juveniles lacked the maturity and level of culpability that traditional criminal sanctions presupposed, and that juvenile offenders should therefore not only be treated as less blameworthy but also as more amenable to treatment and rehabilitation than adult criminals. The primary justification for creating a separate juvenile justice system was the distinction between punishment and treatment. The criminal justice system at the turn of the century emphasized punishment and deterrence. Separating juvenile offenders from adult criminals would allow the juveniles to be treated instead of punished.^[3]

In 1899 the first juvenile court was established in Illinois, marking the formal beginning of a separate juvenile justice system. Other states quickly followed Illinois' lead. By 1920 all but three states had juvenile courts, and by 1932 all but two states had enacted juvenile codes. By 1945 every state had a juvenile court system.^[4]

The juvenile court system in most states was organized as an entity entirely distinct from the adult court system. Juvenile proceedings were held in their own courtrooms, with judges who heard only juvenile cases. Juvenile court procedure was markedly different from that of the general jurisdiction court. Hearings were private and informal in nature. Juvenile court judges enjoyed enormous discretionary power. Juvenile court jurisdiction was classified as civil rather than

criminal. A whole new vocabulary sought to differentiate juvenile court activities from adult criminal court activities. Juveniles were not arrested; they were “taken into custody.” Instead of indicting a juvenile, prosecutors “petitioned the juvenile court.” Juveniles were not convicted; they were “adjudicated delinquent.” Juvenile court sanctions were not referred to as sentences, but as “dispositions.” Juveniles were not sent to prisons; they were sent to “training schools.”^[3]

The belief that the juvenile court’s primary purpose was to help the child remained a cornerstone of the American juvenile justice system for decades. As late as the 1960s, most state statutes still declared the purpose of the juvenile justice system was to help the child rather than to punish.^[5] In 1967, the President’s Commission on Law Enforcement and Administration of Justice recommended maintaining separate juvenile and adult court systems, even while admitting that the juvenile justice system had so far failed in its task of rehabilitating juvenile offenders.

CHALLENGES TO THE JUVENILE COURT

The juvenile justice system continued to grow in size and power well into the 1960s, when several U.S. Supreme Court decisions forced a change in the form of the juvenile court system and the rising crime rate caused many to re-examine the role of the criminal justice system, including the juvenile court. In *Kent v. United States* (1966), the Supreme Court held that the decision whether to transfer a juvenile to adult court required a full, formal hearing. The Court also established the juvenile’s right to have counsel present at the waiver hearing. In *In re Gault* (1967) the Supreme Court held that whenever a juvenile was charged with an act which could result in his being sent to a state institution, be it a prison or a reform school, he must be accorded due process rights including the right to counsel, the right to confront one’s accusers and cross-examine witnesses, and the right not to incriminate oneself.

At the same time the Supreme Court was forcing juvenile courts to undergo massive internal changes, American society was experiencing a dramatic increase in the juvenile crime rate, which increased by almost 250% between 1960 and 1980.^[6] Between 1960 and 1975, juvenile arrests increased over 140%, while adult arrests during this time increased less than 13%.^[4] Both criminal justice professionals and public sensed that the juvenile justice system was failing to control juvenile delinquency and offending. Some began to blame the juvenile justice system’s emphasis on rehabilitation and treatment for the failure to halt the growth of crime. These “crime control conservatives” believed that the emphasis on treatment was thwarting the

effectiveness of the criminal law, and they called for revision of the methods of handling juvenile offenders.^[7] Chief among the complaints was the criticism of the juvenile courts’ perceived leniency toward juveniles charged with serious crimes.

State legislatures began to respond to the criticisms of the existing juvenile justice system and the calls to “get tough” on juvenile offenders. Many states amended their statutes to emphasize that punishment should either replace, or at least be a coequal concern in the juvenile court’s decision-making process.^[5] One of the most common methods employed to increase the punitive nature of the juvenile justice system was to permit increased use of juvenile waiver process.

FORMS OF JUVENILE WAIVER

Juvenile waiver involves removing a juvenile offender from the jurisdiction of the juvenile court and processing him/her as an adult in criminal court. This process is most commonly referred to “as waiver of jurisdiction” or “transfer,” although it is sometimes referred to as “certification,” “reference,” or “remand.” The two reasons most commonly given for making it easier to transfer juvenile offenders to adult court are: 1) that the juvenile justice system has failed to control these juveniles and 2) these juveniles have demonstrated, either by the seriousness of their offense or by the frequency of their appearances in the juvenile justice system, that they are not amenable to the sort of treatment the system provides.^[8]

There are several different forms of waiver. These include prosecutorial, judicial, and legislative.

Most states employ more than one form of waiver. Additionally, 23 states allow for reverse waiver, wherein a juvenile being prosecuted as an adult may petition to have the case returned to juvenile court. In addition, 31 states have a “once an adult always an adult” provision. This means that once a juvenile has been waived to criminal court, any subsequent accusations are also handled in criminal court.^[9]

Prosecutorial Waiver

Fourteen states and the District of Columbia have prosecutorial waiver. It is sometimes referred to as concurrent jurisdiction or direct file. The prosecutor has the option of filing charges against juvenile offenders in either juvenile or adult court. This method of transfer is perhaps the most controversial, because it vests considerable discretion in the prosecutor, whose primary duty is to secure convictions and who is traditionally more concerned with retribution than

with rehabilitation. Prosecutorial waiver is the least utilized form of waiver.

Judicial Waiver

The most common method of waiver and the one that has the longest history is judicial waiver. The juvenile court judge uses his or her discretionary authority to waive jurisdiction and send the case to adult court. A juvenile court may decide on its own motion to transfer a juvenile to adult court, or the prosecutor may move to transfer and the juvenile court judge must decide the motion. Judicial waiver has received general support from scholars, criminal justice professionals, and professional organizations such as the American Bar Association.

At present, 48 states and the District of Columbia allow for discretionary waiver by juvenile court judges. In the past 20 years, a number of states have modified the judicial waiver process, either by lowering the age at which a juvenile is eligible for waiver or by increasing the number of offenses that are eligible for judicial waiver.

There have been numerous modifications of judicial waiver statutes in the 50 states in the past quarter century. Between 1979 and 1995, 18 states modified their judicial waiver statutes.^[3] Between 1995 and 2002, 27 states modified their judicial waiver statutes.^[10] The nature of these modifications varies. Some states lowered the age at which a juvenile is eligible for judicial waiver. Other states did away with an age limit altogether. Other states modified their judicial waiver statutes by adding offenses to the list of crimes eligible for judicial waiver. Still other states modified their judicial waiver statutes by not only lowering the age at which some offenders could be transferred, but by combining this change in age eligibility with an increase in the number of offenses eligible for transfer.

Legislative Waiver

Another form of waiver, and the one which has received the most attention in recent years, is legislative waiver. Legislative waiver is sometimes referred to as “automatic waiver” because juvenile court jurisdiction is removed automatically, without a motion by the prosecutor or a decision by the juvenile court judge. Statutes exclude specified offenses or offenders from juvenile court jurisdiction. Generally, this exclusion is reserved for the most serious offenses or for repeat offenders.

There are two types of legislative waiver. One type, referred to as offense exclusion, excludes some offenses from juvenile court jurisdiction. These are usually serious, violent crimes such as murder, rape,

and aggravated assault. The reason for excluding juveniles charged with such offenses from the juvenile justice system is that juvenile courts cannot impose sufficiently severe sanctions for such offenses. A second type of legislative waiver excludes from juvenile court those juveniles who possess a particular combination of prior record and present offense. This form of waiver, referred to as “offender exclusion,” is directed at juvenile offenders who have failed to desist from criminal activity despite previous contact with the juvenile justice system.

In the past 25 years, states have made increasing use of legislative waiver, primarily by adding offenses that are excluded from juvenile court jurisdiction or by enacting a waiver statute where none previously existed. In 1979, 14 states had some form of legislative waiver. In 1995, 21 states had some form of legislative waiver. Between 1995 and 2002, 12 states enacted legislative waiver statutes, while 16 states amended existing statutes.^[10]

CONCLUSION

In recent years, the “get tough on crime” movement has greatly affected the juvenile justice system. Liberals decry the system’s occasional disregard for due process and the ease with which the state may intervene in the lives of juveniles. Conservatives complain that the juvenile justice system is too easy on young offenders, that this failure to adequately punish juvenile offenders not only fails to prevent future criminal activity but also fails to adequately address society’s right to punish persons who violate its laws.

The result of this attack on the juvenile justice system has been a tremendous change in the use and application of juvenile waiver, in all its forms. Numerous state legislatures have either lowered the judicial waiver age or have added offenses eligible for judicial waiver to general jurisdiction court. In addition, numerous states have enacted legislative waiver statutes that automatically send an offender to adult court for prosecution based on the offense committed. The increased use of both discretionary and mandatory waiver suggests a continued infatuation with the “get tough” approach to juvenile crime, in the face of crime statistics and research that indicates the practice is not achieving its intended goals.

REFERENCES

1. Gardner, M. *Understanding Juvenile Law*; Matthew Bender: New York, 1997.
2. Howell, J.; Krisberg, B.; Hawkins, J.D.; Wilson, J. *Serious, Violent, and Chronic Juvenile Offenders*; Sage: Thousand Oaks, CA, 1995.

3. Fritsch, E.; Hemmens, C. Juvenile waiver in the United States 1979–1995: a comparison and analysis of state waiver statutes. *Juvenile Family Court J.* **1995**, *46* (3), 17–35.
4. Thomas, C.; Bilchik, S. Prosecuting juveniles in criminal courts: a legal and empirical analysis. *J. Criminal Law Criminology* **1985**, *76* (2), 439–479.
5. Hemmens, C.; Fritsch, E.; Caeti, T. Juvenile justice code purpose clauses: the power of words. *Criminal Justice Policy Rev.* **1997**, *8* (2), 221–240.
6. Hamparian, D.; Estep, L.; Muntean, S.; Priestino, R.; Swisher, R.; Wallace, P.; White, J. *Youth in Adult Courts: Between Two Worlds*; Office of Juvenile Justice and Delinquency Prevention: Washington, DC, 1982.
7. Van den Haag, E. *Punishing Criminals*; The Free Press: New York, 1975.
8. Feld, B. The juvenile court meets the principle of the offense: legislative changes in juvenile waiver statutes. *J. Criminal Law Criminology* **1987**, *78* (3), 471–533.
9. Griffin, P.; Torbet, P.; Szymanski, L. *Trying Juveniles as Adults in Criminal Court*; Office of Juvenile Justice and Delinquency Prevention: Washington, DC, 1998.
10. Steiner, B.; Hemmens, C. Juvenile waiver 2003: Where are we now? *Juvenile Family Court J.* **2003**, *54* (2), 1–24.

BIBLIOGRAPHY

1. Feld, B. Juvenile (in)justice and the criminal court alternative. *Crime Delinquency* **1993**, *39* (4), 403–424.
2. Merlo, A.; Benekos, P.; Cook, W. Waiver and juvenile justice reform: widening the punitive net. *Criminal Justice Policy Rev.* **1997**, *8* (2–3), 145–168.
3. Snyder, H.; Sickmund, M.; Poe-Yamagata, E. *Juvenile Transfers to Criminal Court in the 1990s: Lessons Learned from Four Studies*; Office of Juvenile Justice and Delinquency Prevention: Washington, DC, 2000.

Knowledge Utilization: Issues and Recommendations

Cheol H. Oh

Department of Public Administration, Soongsil University, Seoul, South Korea

INTRODUCTION

Until recently, information technology was viewed primarily as a means of helping organizations to function more efficiently. However, rapid advances in this area are changing the ways in which government governments, business operates, and individuals conduct their daily lives. Thus, understanding, directing, and managing information-related activities within government have become critical to the success of government programs and policies. Despite the importance of information-related activities, such activities have not been systematically examined, nor has an effort been put forth to build a comprehensive theoretical framework. In particular, why and how government officials use information in policy making is still an open issue. This article critically reviews past studies as well as offers suggestions for future studies to advance the theoretical and methodological state of the information utilization field. In this regard, this article attempts to address two types of issues germane to understanding and improving information utilization in government decision making, that is, theoretical and methodological issues. More specifically, this study, first, examines the state of past studies on knowledge utilization, mainly in terms of theory building and methodological rigor. Second, three frequently cited theoretical explanations for information use are discussed. In the field of information acquisition, dissemination, and utilization/impact, the rational actor model has been employed as a major tool for understanding activities of information utilization in policy making. Thus, this study first examines the assumptions dealing with information acquisition and processing in the rational actor model. It then puts forward an organizational interest and a communications perspective as alternative explanations for information utilization in government agencies. Third, a set of methodological issues including conceptualization of “use” are also examined. Finally, some suggestions are made for both theoretical and methodological improvement in the field of information utilization. Because this study does not include an empirical test of the three theoretical explanations, some of theoretical suggestions seem to be hypothetical. However, they are expected to show how to manage information resources and, thus, facilitate use of information by

government bureaucrats. More rigorous studies are needed to improve the current state of art in the field of information utilization.

A BRIEF LOOK AT PAST STUDIES

In general, most early studies in the 1980s were faced with the dilemma created by the perceived gap between social science knowledge and its under- or non-utilization. The authors, thus, tried to refute the dilemma by suggesting that underutilization is counterintuitive,^[1] that models of knowledge utilization oversimplifies reality,^[2] and that more systematic analyses are required. Some of major findings of the early studies in the 1980s are, according to Sabatier and Jenkins-Smith, that research may be characterized by the “two communities metaphor,” that social science knowledge often serves an “enlightenment” function especially in the public sector, and that researchers need to abandon the role of neutral technician to increase the use of knowledge.^[3] More specifically, Rich summarized what we have learned about knowledge utilization during 1980s:^[4]

- Format, style, and timeliness by themselves will not lead to utilization.
- Who provides the information is more important than what the actual information is.
- Distinction needs to be made between use and impact of information.
- Knowledge produces effects, not a single effect.
- The existence or non-existence of the “two cultures” is not a significant factor in predicting utilization.
- It is almost impossible to predict when and where a specific knowledge input is likely to have an effect on a particular decision.
- Policy development and problem solving are affected by many variables other than information.

Further, the extensive research studies produced inconsistent findings, suggesting that the field of knowledge utilization need to resolve conceptual and methodological problems. As a result, it may be argued that the field of knowledge utilization in the 1980s was in its infancy in terms of theory building and methodological rigor.^[5,6] Within academic circles, the field is not

viewed as one in which theory building has taken place; theory with respect to knowledge utilization is not well developed yet (on page 323 of Ref.^[7]). Likewise, the conceptualization of use often leads researchers to a traditional input–output model; there does not, however, appear to be a warrant to adopt this perspective of measuring knowledge utilization. It is an unrealistic conception of how societal problems are solved, nor does it accurately reflect how information enters into the decision-making process (on page 330 of Ref.^[7]).

Is this field still in infancy in the 1990s? What contributions have these studies in the 1990s made to the field of knowledge utilization? From the review, we can find several common patterns concerning knowledge utilization: 1) knowledge utilization is a more complex process; 2) it is explained by many variables; 3) research-based knowledge is only part of knowledge decision makers seek or receive; and 4) interaction between researchers and user is critical.

More specifically, studies in the 1990s show some advancement in terms of research orientation (e.g., explanation- vs. description-oriented) and focus. As shown in Table 1, most studies in the 1990s are more interested in accounting for the use of knowledge than simply describing the event of knowledge utilization. Most studies reviewed began with the simple question “what affects knowledge utilization.” To this end, some studies tried to suggest models for explanation, though they are not comprehensive. In so doing, a few factors are examined repeatedly as major predictors of knowledge utilization:

- Internal source is more likely to influence knowledge utilization than external one.^[5,8,9]
- Potential users’ needs, values, and interests are critical in understanding knowledge utilization.^[5,9–12]
- Knowledge utilization varies depending on types of knowledge.^[10,13–15]
- The context that shapes the decision-making process affects knowledge utilization.^[5,10,11]
- Interactive processes between researchers and users are important.^[8,16,17]
- A combination of various factors affects knowledge utilization.^[6,18–20]
- The content of knowledge (e.g., specific, concrete, and relevant knowledge) affects knowledge utilization.^[15]
- Access to knowledge is a prerequisite for knowledge utilization.^[16,21]

As mentioned earlier, the way authors in the 1990s approach knowledge utilization is more explanation focused. They tried to understand why some use knowledge, while others do not. Their findings,

however, are not that new.^a For example, types of knowledge was discussed in depth in Refs.^[23–25]. Likewise, the notion of interactive processes is implicitly embedded in the discussion of two communities metaphor that dominated the 1980s literature.^[3,26] Information source (i.e., internal vs. external source) and content of information were also suggested as important factors in Refs.^[25,27]. The decision-making context was also proposed as a significant factor in accounting for variations in knowledge utilization.^{[28,29]^b} What the studies in 1990s actually did was to examine to what extent such factors, indeed, affect knowledge utilization, rather than making speculations and conceptualizations. In this sense, their work may be an improvement. But we wonder if the studies in the 1990s actually found a new factor or a set of factors in helping us understand knowledge utilization.

ISSUES OF KNOWLEDGE UTILIZATION

The idea that social science information should be used as a basis for government action and/or social reform has a long-standing tradition in the social science. According to Weiss, examples of such research go back as early as the eighteenth century for purposes such as prison reform or social relief for the poor.^[32] Although policy research has different functions over time, its one timeless and unique feature is that it is expected to aid directly in the policy-making process. The results of policy research are meant to provide feedback (or checking points) that can be used for program development or change as part of a cycle of policy process (see Refs.^[33,34]).

The basic premise of information/knowledge utilization literature is that using social science information for public policy making is a good thing. Use is good, more use is better, and increasing the use of social research means improving the quality of government decisions (the so-called idealized model of utilization in Ref.^[35]; also see Ref.^[36]). Thus, whatever the setbacks, optimism often prevails: Social science research should contribute to sounder and wiser decision making. This optimistic view of information in the policy process inevitably became the subject of investigation by government practitioners and by detached scholars, who have produced the beginnings of a

^aInterestingly, empirical findings in nursing and health services are not diverse either. Individual-related factors (e.g., education, beliefs, age, years employed, etc.) generally dominated most studies (see Ref.^[22]).

^bFor a specific example, Webber (1987) and Whitman (1985) showed that the use of policy information (e.g., policy analysis) in state legislators and congressional committees was fundamentally different from that in other areas such as health or education.^[30,31]

Table 1 Summary of review

Conceptual clarity	Mode of research (explanation focused)	Framework (comprehensive vs. partial)	Areas of research	New questions/perspectives developed or examined
Slappendel (94)	Tyden (96)	Tyden (96): partial factors and interviews	Tyden (96): health	Tyden (96): recipient's perspective
Lester (93)	Slappendel (94)	Slappendel (94): partial factors and case study	Slappendel (94): private organizations in ergonomics	Hutchison (95): social constructionism
Logan and Graham (98)	Florio and Demartini (93)	Florio and Demartini (93): partial factors and case study	Florio and Demartini (93): rural communities	Wingens (90): systems approach
Rich (91)	Hutchison (95)	Hutchison (95): partial factors and multimethods	Hutchison (95): social work	Boggs (92): three-way exchange model
Landry et al. (01)	Lester (93)	Wingens (90): a general model and conceptual	Lester(93): state officials in hazardous waste, economic development, welfare, and education	Cousins and Leithwood (93): interaction process
Webber (91)	Cousins and Leithwood (93)	Boggs (92): partial factors and conceptual	Cousins and Leithwood (93): education	Meillier et al. (97): back pack function
Booth (90)	Backer (91): general issues	Lester (93): partial factors and survey	Hall (97): education	Hultman and Horberg (97): social context
Watkins (94)	Backer and Shaperman (93): general issues	Cousins and Leithwood (93): partial factors and survey	Meillier et al. (97): health	Castill (00): reciprocal process
Oh (97)	Rich (91)	Hall (97): partial factors and descriptive	Hultman and Horberg (97): education (i.e., teachers)	Rich and Oh (00): rationality
Oh and Rich (96)	Castill (00)	Meillier et al. (97): partial factors and interview/survey	Loan and Graham (98): health	Larsen (98): institutional approach
	Rich and Oh (00)	Hultberg (97): two cultures	Castill (00): environment	Leeuw (91): policy theory
	Cunningham et al. (00)	Hultman and Horberg (98): partial factors and interview/survey	Rich and Oh (00): mental health	Watkins (94): critical theory
	Landry et al. (01)	Logan and Graham (98): an integrating model and descriptive	Cunningham et al. (00): health (i.e., addiction)	Conner (98): social ecology
	Huberman (94)	Castill (00): partial factors and survey	Landry et al. (01): education	
	Larsen (98): general issue	Rich and Oh (00): partial factors and survey	Huberman (94): vocational education	
	Webber (91): general issues	Cunningham et al. (00): partial factors and case study	Leeuw (91): juvenile delinquency	

(Continued)

Table 1 Summary of review (*Continued*)

Conceptual clarity	Mode of research (explanation focused)	Framework (comprehensive vs. partial)	Areas of research	New questions/perspectives developed or examined
	Woerkum and Meegeren (90): general issues	Landry et al. (01): partial factors and survey	Nilsson and Sunesson (93): social welfare	
	Leeuw (91): general issues	Huberman (94): partial factors and case study	Zuzovsky (94): education	
	MacRae, Jr. (91)	Oh and Rich (96): an integrating model and survey	Cahill et al. (90): state budget office	
	Nilsson and Sunesson (93)	Webber (91): partial factors and descriptive	Eriksson (90): social work	
	Zuzovsky (94)	Leeuw (91): partial factors and descriptive	Oh and Rich (96): mental health	
	Booth (90): general issues			
	Cahill et al. (90)			
	Eriksson (90)			
	Conner (98)			
	Weiss (98): general issues			
	Oh and Rich (96)			
	Oh (97): general issues			

research literature on the production and utilization of social science knowledge by the government. With more information available, government seems to enjoy the luxury of choice; yet, reality tests of that experience do not guarantee confidence that the quality of government decision making has kept pace.

Past studies of information utilization in organization and/or in decision making have generally taken information for granted or assumed it to be essential and seldom worthy of special attention.^[37] If, in some circumstances, information is artificially restricted, scholars of various stripes generally conclude that it is undesirable. It is also assumed, sometimes naively, that expanded information cannot help but improve the efficiency and effectiveness of any policy-making system, mainly on the basis of simple positivism and/or factual efficiency.^[38] As Webber pointed out, “both providers of information and policy makers presume that policy information contributes to a knowledge base that can assist in making sound policies.”^[27] However, the theoretical rigor and empirical validity of this assumption must be tested. That is, does the assumption accurately describe the actual behavior of policy makers in government?

Despite the tremendous increase in attention to the importance of information in the decision-making process, research generally indicates that governmental policy makers make little use of information (see Ref.^[39]). At best, social science research findings alter policy makers’ understandings and/or definitions of policy problems over a long period of time (the “conceptual use”; see Refs.^[36,40,41]).

It seems clear that both practitioners and scholars think that social science research is not used in policy making as much as is expected. Social scientists have long believed that they can contribute to the pursuit of human welfare, but they also believe that social science information is seriously underused in important policy decisions. Policy makers, however, believe that the reports they receive are unintelligible, do not deal with the immediate problems on their agenda, and are not sensitive to the unique pressure for action under which they must perform. Scholars are not, however, satisfied with fact finding about information utilization in policy making. They became more interested in accounting for variations in the use of social science information and, subsequently, improving the use of it.

By employing a single perspective (e.g., rational choice, organizational interest, or communications) and not considering other important variables and their relationships, most studies provide only a partial explanation of information utilization and processing in policy making. For example, the two communities metaphor has been at the heart of most studies of information utilization, in explaining variations in the

utilization or the underutilization of social science knowledge. In the utilization literature, the gap between researchers and policy makers is related to some (or all) of the following factors: availability of information, relevance of information, the form in which it is delivered, etc. If the gap is appropriately met, then utilization will increase. A competing hypothesis (i.e., the organizational interest perspective) to this two communities metaphor holds that factors affecting information use are quite different from factors traditionally considered important and argues that levels of utilization may be best explained by examining routine bureaucratic or organizational roles and procedures. As a result of single-perspective approaches, this area of research seems to lack a comprehensive framework for information processing and utilization/impact. Some scholars warn that this field of inquiry faces a crisis of identity.^[7] The situation is further complicated by the fact that, despite voluminous work on the use of information in policy making, little is known about its impact on policy making. Practitioners and/or scholars simply assume that use of information automatically leads to changes in outcomes of policy making and do not explore the dynamic processes and causal linkages of factors involved in the impact of information.

Overall, we can easily find that there is the conspicuous absence of a greatly expanded understanding of the use of knowledge from past studies.^[42] Taken as a whole, empirical studies in the area of knowledge utilization have suffered from several critical problems (see Refs.^[7,34]). More important, there appears to be serious conceptual and methodological gaps that need to be filled. Under the circumstances, this article attempts to examine some of significant theoretical and methodological problems—building a theoretical framework from an integrated perspective and measuring “knowledge utilization” from a process perspective. It, then, offers some suggestions for future studies.

EXPLANATIONS FOR INFORMATION UTILIZATION

With some knowledge about how information utilization can be understood in the context of information processing in organization or in the decision-making process, a more germane question to this study is why government decision makers are expected to use information, that is, accounting for government officials’ behavior in using information in their daily business. Among others, there are three frequently discussed explanations for information use that address factors affecting use of information and, thus, can help us develop feasible policy suggestions for improving use of information in government.

Rationality Perspective

As Aldrich noted, virtually all scholars agree with what he calls the “fundamental equation” of political behavior, which is that preferences (or values or goals) determine behavior.^[43] Rational choice is about just how those preferences determine behavior. Although there is no universally accepted definition of rationality, rational actor, or rational choice theory, most social scientists agree that rational decisions should confirm to certain basic requirements. This article does not discuss all variations of rational choice/actor theories; rather, it examines basic elements of rational decision making—“basic” because all rational actor theories agree on it.

All “rational actor” theories in the social sciences try to explain “choice” by individuals or organizations. According to theories of rational decision making (for more detail on rational choice and its limitations, see Refs.^[44–48]), humans are rational actors in the sense that they are engaged in the process of “optimizing” their expected utilities (or simply goals), by selecting the course of action with the highest payoff through a comprehensive analysis. All rational actor theories in the social sciences call for a systematic canvassing of possible alternatives, for a systematic analysis of the consequences of each alternative in terms of the values or goals that one wants to maximize, and for possible choices to be guided by such analysis. Consequently, to choose the best alternative among available options, a decision maker is necessarily required to evaluate and analyze the consequences of available alternatives. This is where information becomes central to rational choice theories. As March pointed out, the main rationale for using information in rational decision making is its role in reducing uncertainty in making a choice among alternatives.^[47] In models that maximize one’s utility function, the lack of information is often perceived as the determinant of seemingly “irrational actions.”^[49] Despite the psychological and other constraints (e.g., cost) in decision making, the notion of bounded rationality also assumes that information is essential in allowing individuals to compare alternatives (see Ref.^[48]). In a similar vein, Elster pointed out that if decision makers had little information, rationality required them to abstain from forming and estimating possible consequences of alternatives.^[50]

Regarding information processing, rational actor theories generally assume that all available information about possible consequences of alternative options will be gathered from various sources, so that individual decision makers are “completely” informed of decision situations (see Ref.^[51]). They also assume that information gathered will be used for making decisions, if it is precise, reliable, and relevant.^[52,53] In other

words, information is being gathered for making decisions rational within the limits of resources and cognitive abilities. It is the structure of the process that makes the decision-making process “rational.” Within this structure, one step leads naturally (and automatically) to the next. Thus, the critical role of information is just assumed and not tested.

Given these underpinnings of rational decision making, what assumptions concerning information acquisition and processing would be needed to maintain the consistency and coherence of rational actor theories? Most of the arguments about the relationship between rational decision making and information revolve around the issue of perfect information (e.g., whether an actor has perfect or incomplete information under constraints of time and resources, for example, perfect information about candidates in election; see Refs.^[54–56]). However, this notion of perfect information (or amount of information) is just one dimension of information processing in organizations and individual decision making. The major underlying assumptions of the rational actor theories with respect to information acquisition, dissemination, and use/impact can be summarized as follows:

1. The human mind is capable of processing all sources of available information.^[48,51]
2. All relevant sources of available information will be sought out and should be applied to a given problem.^[53]
3. Once information is acquired, it will be widely disseminated to all who “have a need to know.”^[7]
4. When the information is collected and disseminated, it will be used if it is sensible and scientifically valid.^[39,47,57]
5. Use of information will lead to a choice among competing alternatives; hence, the information will have an impact.^[47]

Oh systematically examined the empirical validity of rational actor theories as an analytical tool for explaining use of information in policy making.^[20,58,59] By examining each stage of information processing, Oh demonstrates that the assumptions of rational actor theories with respect to information processing in individual decision making do not match with real behavior of decision makers. The empirical findings of information acquisition indicate that individual decision makers search for specific sources of information rather than conducting a wide search of available information. Likewise, at the dissemination stage of information processing, decision makers share information more with those in their own agencies rather than disseminating it to other actors outside their agencies. In a similar vein, at the stage of utilization and impact, Oh found that there was no automatic

linkage between information acquisition and its use. Further, the relationship between information acquisition and its use is not linear as assumed in rational actor theories, suggesting that the amount of information acquired does not necessarily get decision makers to use information in decision making. Although there is a positive relationship between use of information and its impact, use of information explains only a small portion of the variance in the impact of information, and the relationship between the two variables is not linear. This implies that use of information does not necessarily entail simultaneous effects after a certain point of utilization. Even when there is an impact, it is not proportional to use of information. Overall, the findings of Oh's study indicate that individual decision makers do not necessarily process information as assumed in rational actor theories and that, at each stage of information processing, different motives and/or factors may be involved.

Organizational Interest Perspective

This perspective derived from the literature on organizations begins with the assumption that organizational rules, structures, tasks, and cultures are essential for understanding information acquisition, dissemination, and utilization. It also assumes that choices with respect to acquiring and using information are predictable: From this assumption, actors make choices that maximize organizational interests (e.g., budget, personnel, or mission; see Ref.^[60]). The bureaucratic literature of the Weberian tradition notes that once bureaucracies begin to develop a position of independent power and authority, they devote much of their attention to securing and maintaining their own autonomy, which is often called "organizational interest" (see Refs.^[7,61]). This notion of organizational interest is especially important for our analysis of information acquisition, dissemination, and utilization in governmental bureaucracies, because it lies at the foundation of bureaucratic power. Concerned about bureaucratic secrecy and fear of how others might use information if it were to be shared with them, bureaucrats tend to seek monopolies over control of information. This tendency toward monopolistic control over acquisition, dissemination, and utilization of information is consistent with bureaucratic decision makers' reliance on their own information and rejection of information from external sources (for more detail, see Refs.^[62,63]). Consequently, in the case of bureaucratic theories, it seems clear that:

1. Information is essential to the power and prestige of bureaucratic organizations. As a result, use of information is basically instrumental for maintaining or increasing bureaucratic power.

2. The desire to protect organizational interest affects information produced in bureaucracies, as well as types of information disseminated to other bureaucratic agencies and the public.
3. Through incentive systems, organizations can control or facilitate use of specific information.
4. There is, thus, a tendency to search for limited information and to rely on the inventory of information already developed within the bureaucracy. Therefore, information from external sources (i.e., outside the organization) will rarely be consulted.
5. The "trustworthiness" and the "credibility" of the information source are essential for use of information. Therefore, selection of information and consequent use of it will happen when the information confirms a policy position or interest already held by a policy maker.

From this perspective, we can now understand the findings of Oh's study presented above.^[58] For example, in the case of information acquisition, decision makers search for information more from their own agencies, rather than touching their decision base with various sources. These results are quite opposite of the assumptions of information acquisition in rational actor theories. However, these results make sense from the organizational interest perspective. As Weiss aptly pointed out, decision makers in bureaucracies do not always search for all possible sources of information.^[64] Instead, they seek particular types of facts or opinions from particular types of sources in selective and strategic ways, ways that might not correspond to what the rationality model predicts. This is because decision makers can trust information from internal sources and because information from internal sources is believed to support their positions on policy issues or interests of agencies. Likewise, bureaucratic secrecy and monopolistic control of information may make decision makers less likely to share information with other decision makers outside their agencies. This occurs to a large extent, because they do not want to share sources of potential power with others. Further, at the stage of information utilization, even the valid and reliable information has little use for decision makers. This is because the content of information may be, although scientifically valid, against interests or goals of decision makers' agencies or because it simply may not be useful to decision makers (see Refs.^[65,66]).

Communications Perspective

Communication-related studies of information utilization often begin with the assumption that there is a "great divide" between the community of science

and that of government or politics and that the linkage problem between both communities serves to explain low levels of information utilization.^[6,20,33,34,39,59,67-69] Most communication-related studies take the so-called “two communities metaphor” as the underlying conceptual framework for their work.^[68] Typically, the gap between two communities is expressed in terms of a few factors:

1. There is great distrust and even antagonism between the two communities.^[67,68,70] It is often observed that considerable tension exists between program officials and researchers. Program officials believe that they have received little help from research. Likewise, researchers are weary of anti-intellectual program managers and their demands for how-to manuals.
2. There is a preference for and use of alternative jargon and language in each community.^[33,67,68] Researchers communicate to their peers in language designed for scholarly journals. Brief, clearly written reports, free of terms familiar to those who are members of a particular guild, will increase the probability of utilization.
3. Researchers and bureaucrats operate under substantially different conceptions of time and worldview.^[7,39] Government officials are accustomed to working on immediate problems and to meeting deadlines. Researchers who are rewarded for producing high-quality research see no harm in delivering a “better product” a week or even a month after the original delivery date.
4. It is asserted that researchers need to be more concerned with the needs of government officials and the relevance of research to these needs.^[65,67,71] Relevance may be expressed in terms of the questions being investigated; is it one of interest to the potential users?

One of the rational actor model's assumptions is that decision makers use little information when they are provided with little information. However, Oh's study showed that the assumption between acquiring information and use of it is not accurate.^[58] Unlike rational actor theories, the communications perspective, for example, explains little use of information in decision making in terms of the lack of interaction between decision makers and researchers. The lack of interaction between the two actors may produce information, which does not meet the needs of decision makers and is thus less useful, although scientifically valid and reliable. Interestingly, more than one-half of the respondents in Oh and Rich's sample also pointed out that social science information was abstruse and unrelated to the day-to-day problems of administering mental health policy.^[6,20] Respondents went on to

answer that they were, thus, less likely to use information in the formation of policy. Therefore, information utilization is increased in cases where policy makers have the opportunity to clarify their concerns, results, and implications of research through feedback to social scientists.

METHODOLOGICAL ISSUES

Traditions in the Measurement of Utilization

Those who have studied how knowledge, especially research-based knowledge, affects decision making (or other types of outputs or outcomes) tend to focus on the following types of questions:

1. What use does information have in decision making?
2. What types of information are preferred? Are some types selectively passed over or ignored?^[33]
3. To what extent do decision makers consider research in their work?^[70]
4. From the time that information/knowledge enters into an organization (through some channel), what happens to it? What are the diffusion/dissemination patterns (see Refs.^[72,73])?
5. What patterns does information follow, as it flows through an organization or through a specific decision-making hierarchy?
6. To what extent does information confirm or legitimate decisions that have already been made or about to be made?^[53] To what extent does research influence a decision maker's position?^[74,75]
7. At what rate do decision makers adopt new information sources?^[62]
8. What are the characteristics of social science research that make it useful for decision making?^[27,70]
9. To what extent does information benefit the decision maker or the decision-making process?^[34,63]
10. To what extent can a person document levels of premature utilization, deliberate non-utilization, and/or selective utilization of information?^[6,25]

The critical assumption behind many of these questions is that many specific decisions can be attributed (or that decisions are “caused by”) to the use of specific bits of data. In this context, a key question becomes “how does one document the use of information or its impact?” In answering this question, researchers seem to have had an input-output model in mind. Within this framework, it is believed that:

- 1) one can trace the flow of information from the time

it enters an organization to the time that some action is taken (predicated on or influenced by the new information); 2) impact or influence can be measured relative to a single piece or cluster of discrete information that enters an organization (or to which an individual is exposed); and 3) it is appropriate and possible to assess the impact or influence of a single piece of information on the problem-solving behavior of an individual.

In short, for purposes of conceptualization and measurement, researchers/analysts tend to adopt a deterministic view of how knowledge (especially research-based knowledge) is disseminated and used.^[64,76,77] Respondents in empirical studies are conceived of as being able to identify how discrete bits of information affect specific behaviors, choices, or actions. Sometimes specific studies are asked about, and, other times, the influence of social science information or policy-related information in general is assessed.

For example, it is believed that users can say, “the study completed by Jones led me to take an action that I would not (or probably would not) have taken, if I had not been exposed to the study.” Implicitly, it is also being said that the information generated by the Jones study is the key element in reaching a decision. The work environment, past experience, ideology, beliefs, and educational background of an individual do not affect a particular choice or decision in the same way or to the same extent that the Jones study does.

In reviewing the empirical literature on the creation, diffusion, and utilization of knowledge, one quickly learns that the assumptions behind the input–output model are not warranted. They are not based on a realistic conception of how societal problem-solving processes operate (for more detail, see Refs.^[34,63,69]):

1. Knowledge produces effects, not a single effect. If one concentrates on a single effect, one will underestimate the impact that knowledge is having.^[42]
2. Decisions do not generally represent a single event; it is difficult to understand a single decision without examining the series of events that led up to it.^[78,79]
3. It is almost impossible to predict where knowledge is likely to have an effect on a particular policy or decision, or what that effect will be. The problem with the input–output approach for conceptualization or measurement is that the assumptions inherent to the model are rarely met.
4. The approach requires that one is able to attribute (in a causal sense) an action to the dissemination and application of specific bits of data or information.
5. Users and potential users have to be able to abstract utilization processes from overall (more general) problem-solving processes.

6. Problem solving is affected by many variables other than information (e.g., ideology, work environment, and demographics; see Ref.^[80]).
7. Use is not always tied to a particular action or choice.

However, if one wants to reject the input–output approach and “advance” beyond it, there are formidable conceptual and methodological problems that need to be overcome.

First, because knowledge accumulates and builds in organizational memories, some decisions (or outputs) are made that seem independent of any identifiable, discrete inputs. The “process” can simply not be traced in a step-by-step, linear fashion. For the analyst and potential evaluator, this poses the problem of not being able to identify the universe of inputs for a particular output or decision. It is also true that the universe of inputs has accumulated and grown larger over time to deal with long-standing issues. Thus, one is not able to judge levels of utilization and the processes affecting utilization at any certain moment.

Second, because knowledge produces effects, it is often impossible to trace outputs back to their specific inputs, even if it were possible to identify the universe of information inputs. It seems clear that input–output models are only appropriate when one is able to analyze the mechanical application of knowledge to problem-solving activities. For example, statistic X is needed to justify policy Y. In that case, a “tracing” approach could work out well. In cases such as this one, outputs can be traced back to specific, discrete inputs. However, it is difficult to trace the subtler, more naturalistic application of information to the myriad of decision-making channels within organizations.

Analysis Methods

The field of knowledge creation, diffusion, and utilization has employed various approaches to measure utilization, utility, influence, and impact (see Refs.^[33,34,81]).

Survey

The most common method has been surveys that ask respondents to provide self-reports on what information they acquired and used. Moreover, the surveys ask respondents to identify the factors that account for utilization and non-utilization (e.g., Refs.^[70,82]). In many of these surveys, respondents use their own definition of use (they are self-anchored responses). The surveys of a relatively large number of individuals at different levels of government, in different positions of power, represent the most popular approach that has been employed. In addition to large surveys, some studies

have also used a small, focused group of elite interviews, and, in some studies, interviews have been combined with case studies.

Social framework analysis

Another approach that has been used is called “social framework analysis.” This approach analyzes how “schools of thought” and “frames of reference” (e.g., the Chicago School of Economics, Merton’s Columbia School, Paul Lazarsfeld’s approach, and others) are identified as particular ways of defining and analyzing problems. It is clear, for example, that Paul Lazarsfeld had a major influence on the shape of American and European sociology, as well as a substantial impact on the way in which public officials think about social issues. Lindblom and Cohen pointed out that individuals could have a major impact on how science and society approach problems.^[83] Often, according to their analysis, this influence follows their own lifetime and cannot be traced in a direct causal path; Lindblom and Cohen wrote of “indirections” and “lags” in achieving influence. Using the example of Keynesian economics, the authors pointed out that varying forms of “information” might take in varying situations “along the long convoluted road to influence.” This form of analysis does not always take on precise measurement; it relies heavily on reasoning and judgment. However, citation analysis and network analysis are often used to chart influence of ideas and individuals over time.

Case studies

Another example of the way in which researchers/scholars/practitioners have tried to understand the process of knowledge utilization is through the case study method. One of the most thoughtful examples of this approach is Horowitz’s *Rise and Fall of Project Camelot*.^[84] Horowitz reviews the basic documents and related critiques, pertaining to the U.S. Army’s controversial South American social science research projects discontinued in 1965. Horowitz concentrates on the use and abuse of social science research. The case study has the basic advantage of being able to cover an event, a program, or a policy from its initiation through the stage of implementation. Its obvious disadvantage is that it is difficult to generalize on the basis of a single case.

Governmental commissions

We have also learned from special governmental commissions, as well as studies of knowledge utilization

and its related processes. Lyon’s *The Uneasy Partnership*, Brim’s *Knowledge into Action*, Orlan’s *Contracting for Knowledge*, the National Academy of Sciences’ *Federal Investment in Knowledge of Social Problems*, and Bloch’s review of research centers are just a few examples of these kinds of studies. These studies generally consist of testimonials and, in most cases, lament the fact that there is not a closer working relationship between government and the scientific community.

Citation analysis

One quantifiable method of assessing the impact of knowledge that is not subject to respondent reactivity (although still subject to researcher selection bias) is citation analysis (e.g., Ref.^[85]). The particular significance of this method is the power of specifying complex interactions among forms of information and types of use. One is also able to uncover complex communication networks.

Experimental designs

The specification of information, situation, and use and the precision with which they are measured through survey research are crucial to the development of a valuable framework for knowledge utilization. However, through the control of extraneous influences and the purposive manipulation of “treatments,” experimental and quasiexperimental research designs offer the potential elimination of certain threats to validity and reliability, thereby increasing the logical certainty with which one can address the kinds of questions central to knowledge utilization as a field.^[86,87] Whereas the field of utilization has not developed to the point where true experiments have been tried on a systematic basis, there have been preliminary attempts to use experimental and quasiexperimental techniques to isolate and assess influences on knowledge utilization.^[88]

WHERE TO GO: SUGGESTIONS FOR FUTURE STUDIES

Given the history of information utilization/impact and its importance, as well as the tensions between the imperatives of science and those of decision making, the key issues in the development of our knowledge of information utilization and impact must be identified. Despite efforts to explain the gap between the perceived need to use information in decision making and the ability to apply this “need” to “practice,” the literature has several serious theoretical and methodological problems (for more detail, see Refs.^[6,69]). Then, the question is how to resolve the problems. This

study intends to offer a set of suggestions that are expected to help scholars to resolve the theoretical and methodological matters discussed previously.

Directions for Theoretical Improvement

Reducing conceptual confusions

The implicit assumption that use of information makes a difference in decision making is fundamentally flawed (for a similar point of view, see Refs.^[39,89,90]). That is, if information is used in policy making, it should help decision makers to choose a specific policy option and/or change existing programs.

This assumption implicitly or explicitly perceives use of information as having a simultaneous, direct effect on decision making. Assuming a direct link between use of information and its impact explains why few studies have addressed the “impact phase” of the information process as the subject of inquiry.

Empirical observations on how information is produced, disseminated, and used in governmental decision making undermine the assumption, because it is not based on a realistic conception of how information is processed in decision making. Like Innes’s argument that information influences without necessarily being actively used, this author’s view is that information can be used without necessarily having an effect.^[66] In fact, a direct, one-to-one relationship between information utilization and its impact does not need to exist. Many variables other than use of information (e.g., work environment, types of information, or need for information) affect the probability that information used influences the outcome of policy making.

Overall, no systematic studies have been conducted to explain comprehensively the impact of information on decision making (for an example, see Ref.^[63]). That is, how and when does information actually influence decision making? Without studying this impact stage, any investigation of the process of information utilization in organizations and/or in decision making is incomplete.

Building a comprehensive framework

The use of information is a complex phenomenon involving environmental, organizational, and attitudinal components, as well as the specific characteristics of information (e.g., sources, content, or quantity of information received by decision makers). As mentioned previously, most studies lack a comprehensive conceptual framework systematically integrating such factors (see Refs.^[66,91]). This problem of lacking comprehensive research is reiterated as one of four challenges facing the field of knowledge utilization.^[92]

The knowledge utilization field is fragmented and needs new synthesis and cross-cutting activities, such as dealing with issues of overlapping definitions or defining appropriate research paradigms for difference types of knowledge utilization research studies.

The authors of past studies often picked variables that they believed were significant for their studies, with little consideration of other factors. As Rich correctly stated, instead of trying to understand the overall problem-solving situation and assessing where utilization fits in, we are focusing on an artificial utilization process that assumes that all the other factors can be held constant.^[7] In other words, the literature has identified factors that affect one or more parts of the information utilization process, but we do not have an overall perspective. The lack of a comprehensive conceptual framework addressing multiple levels of reality has led researchers to much conjecture about the process of information utilization, but with little subsequent systematic empirical investigation.^[80] For example, a decision maker who distrusts research findings may not use such information and think that it is not helpful in understanding emerging problems (e.g., causality or possible solutions).

However, as Rich and Oh pointed out, decision makers may be likely to seek and use more information if their organization encourages use of information in policy making by offering incentives/rewards.^[33] Thus, a study of information utilization and impact, ignoring various factors and based on a single paradigm, is likely to be counterproductive and may actually contribute to the abuse of information. Larsen succinctly pointed out the need of developing a comprehensive framework:^[93]

[F]or some time, the explanation for inconsistent findings of studies was that not enough research had been conducted to permit valid summarization and that more data were needed. However, in many areas, sufficient research results now are available and *the synthesis of these findings is both timely and necessary.* (Emphasis added.)

A more general framework is thus required that 1) identifies and distinguishes the relevant properties of variables involved and 2) examines how these differentiated components are linked.

Explaining causality

The amount of empirical research on information utilization is distressingly small (for an example, see Refs.^[34,58]). As Barber emphatically and correctly stated, “the predominant present view of the sociology of knowledge is of a separate, philosophical, ideological . . . and non-researchable field” (quoted in Ref.^[94]).

Of the few empirical studies, most deal with utilization in terms of a simple description of respondents' answers to specific questions. Or they simply explain the variance in the dependent variable by employing regression methods (for a detailed review of past studies, see Ref.^[33]). This line of research may help us to understand which variables are more important than others in accounting for use of information. This approach, however, does not investigate the causal linkages among the variables involved in the processes of information utilization. This mode of empirical analysis has left the complexity of relationships among variables unexplained. As Dunn and Holzer noted, one of the operating assumptions regarding the knowledge utilization field is that the processes of knowledge utilization/impact are interdependent in their causes and effects and, thus, complicated.^[95] Therefore, understanding causal linkages among the factors will provide a complex and rich understanding of the direct, indirect, and/or spurious influence that variables may have on the use of information and its impact (see Refs.^[6,62,63]). As Huberman pointed out, we should no longer attempt exploratory studies but rather use explanatory and/or confirmatory studies.^[57]

Considering policy areas

Past studies pay little attention to the effects that differences between policy areas (more specifically, types or modes of policy making) may have on the use of information and its impact. They were generally conducted in a single policy area (e.g., unemployment, education, or health). Even the studies employing more than one policy area did not pay analytical attention to their differences (see Ref.^[79]). As Huberman suggested, we may agree on explanatory variables but disagree that they are equally important across policy areas in accounting for information use and its impact.^[57]

In fact, policy area as the subject of inquiry began to get serious attention from scholars after the publication of Lowi's article in *World Politics*.^[96] Until Lowi, policy was viewed as a product—an outcome—of government. In other words, politics or political factors determined policies. However, Lowi reversed this relationship. His idea is based on the assumption that the perceived attributes of the policy process (or different policy areas) determine the attributes of the political process that makes policy. This is because policy areas differ in the major aspects of the political process, such as power structure, primary political unit, or stability of structure. Therefore, if we understand a certain type or area of policy, we are supposed to understand the political process of factors behind the area or type of policy. By the same logic, we can assume that different policy areas can lead to

different processes of information utilization and impact. As March and Olsen noted, information processing varies according to the context in which decision making takes place.^[97] Contradictory motives are often imputed to decision makers, depending on these various contexts. Thus, because the contexts in which information processing occur share little common essence, the effect of different contexts must be explored (see Refs.^[47,69]).

A Suggestion for Methodological Improvement: Utilization as a Process or an Outcome?

From an evaluation perspective, one of the first questions to address is “how should we specify the nature of knowledge utilization?” That is, to what extent should utilization be considered a “process” and to what extent an “outcome”? As Dunn pointed out, to improve the practical consequences of using knowledge, we need to examine the complex processes of knowledge utilization.^[81] To the extent that utilization is considered to be almost exclusively a process, one should ask, “toward what end is the process oriented?” Is it oriented toward some type of instrumental change, toward planning, toward designing services for consumers or other users, or toward generally facilitating effective problem solving? From a process perspective, it is equally important to determine the extent to which using information is a priori good independent of a specific outcome.

However, if one conceives of utilization as leading to a particular outcome or set of outcomes, then it becomes important to be able to attribute an outcome to the use of some discrete set of information (see Ref.^[93]). To what extent is it possible to attribute a particular outcome to the use of information? Attribution, in this case, may refer to:

- Information contributing to the process of achieving particular objective.
- Information influencing the process of reaching an objective in a positive or negative sense.
- Information “causing” an objective or outcome to be reached

Some other types of “outcomes” that knowledge utilization may be associated with include designing a service delivery system or strategy, developing a strategy, helping to change attitudes of individuals or groups, sensitizing service providers to the needs of consumers, informing the public about a particular problem or issue, helping in the formulation of guidelines or regulations, designing incentives or incentive systems, developing a marketing campaign, and other general and specific problem-solving activities (for more

detail, see Refs.^[39,42,78,81]). As already noted, from an evaluation perspective, it may be difficult to attribute discrete outcomes (i.e., in a cause–effect sense, which Rein and White^[77] call “decisionistic”) to the use of specific information. This is and will continue to be a formidable issue for scholars and practitioners alike, because such an understanding of use is overly narrow and rarely valid.

In this context, it may be more relevant to think about the extent to which information, research, and knowledge are contributing to or “influencing” some particular objectives or outcomes. It is also possible that information has been influential in a user’s thinking about an issue or problem. In this case, the user can say that the information had an influence, but that it could not be traced in cause–effect terms. For purposes of measuring utilization, we need to operationalize what we mean by influence. The question about utilization for what or toward what end can be transformed into the question of “if utilization is the proximate outcome, what is the ultimate outcome?”

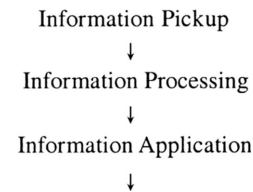
“Use” is also not an all-encompassing concept, which has the same meaning to everyone. There is no commonly agreed-upon definition of what use means. First, use is different from acquiring and disseminating information. Dissemination simply means that some channel of transmission has been used to send data or information from one source to another. Second, in an evaluation context, it is relevant to distinguish between use, utility, influence, and impact:

1. *Use* may simply mean that information has been received and read; it does not necessarily imply that information has been understood. It also does not imply that an action has been taken after the information was received and read.
2. *Utility* represents a user’s judgment that could be relevant or of value for some purpose that has not yet been identified. It also does not imply that an action has been taken.
3. *Influence*, as alluded to above, means that information has contributed to a decision, action, or way of thinking about a problem. In this case, the user believes that using information aided in a decision or an action.
4. *Impact* is more action oriented. In this case, information has been received and understood and has led to a concrete action, even if that action is to reject the information (also see Ref.^[98]).

These distinctions are significant when thinking of utilization as a process rather than an outcome. This article argues that utilization is a process—a series of events that may or may not lead to a specific action by a particular actor at a given point in time. The utilization process may best be described as a series

of “less than discrete” events or what Weiss called “knowledge creep.”^[70]

In measuring utilization, one is not examining a single, discrete event at one point in time. Knowledge utilization occurs gradually over extended period of time and consists of several events:



Depending on the individual and/or organization, this process occurs over a long period of time, or it may occur within minutes. Within each stage or step, there are several variations of how the process may be operating.

CONCLUSION

Students of information utilization share a firm conviction that knowledge/information has become an inextricable element of society that is irreversibly changing public policy and/or administration. They all stress—directly or indirectly—the urgent need for research into practical experience or vice versa. Only in this way can a solid basis be created for a systematic understanding of the field and for the building of more general theories that can integrate existing partial theoretical constructs.

To this end, an important task commands our attention to developing a comprehensive and appropriate theory of information utilization. A first step toward building such a theory may be to integrate the three perspectives mentioned previously. This does not mean, however, that rational choice or actor theories should be rejected from the mainstream of social science (see Ref.^[99]). Rather, to build “more realistic” models of information acquisition and processing in decision making as a potential contributor to scientific progress in social sciences, each stream of studies should be understood as a complementary explanation for information processing in decision making (see Refs.^[49,100]). A more interesting question would then be “under what circumstances can these perspectives be combined together? and how? or which perspective is more appropriate? and why?” Use of information seems to be more easily explained by various factors than by a single perspective. Further, the causality among the variables assumed may depend on certain factors, such as policy areas, policy types, the nature of problems facing decision makers and/or different stages of the policy-making process (for an example,

see Refs.^[34,63,101]). More important, there is a need to extend the generalizations that are currently considered to be “state of art” in information acquisition and processing. For example, some studies suggest that bureaucratic political factors may be more important in explaining and/or predicting information utilization than communication-related factors, factors associated with the quality of information being presented, or the rationality of decision makers. Then, do bureaucratic processes represent the “sufficient condition”? Or is quality a necessary condition, but not a sufficient condition, for information utilization? These questions need to be answered before a framework that integrates the explanatory variables represented by each perspective can be tested.

It is also clear from the discussion presented in this article that more attention needs to be given to conceptualizing and measuring “utilization,” as well as its related concepts. This is essential to constructing an evaluation framework, in which we are interested in specifying appropriate outputs and outcomes and in which we want to construct effective measures of use and its outcomes. In moving toward this evaluation framework, we need to free ourselves from the shackles of the so-called input–output analysis. However, there are formidable problems that first need to be overcome. From a methodological perspective, input–output analysis is easy and is the only tested method of analysis currently available. We, however, need to meet the task of constructing a more appropriate evaluation framework. The issues identified in this article are expected to provide a starting point from which to build.

REFERENCES

- DeMartini, J.A.; Whitbeck, L. Knowledge use as knowledge creation. *Knowledge* **1986**, *7*, 383–396.
- Sunesson, S.; Nilsson, K. Explaining research utilization. *Knowledge* **1988**, *10*, 140–155.
- Sabatier, P.A.; Jenkins-Smith, H.C. Symposium on policy change: editor's introduction. *Policy Sci.* **1988**, *21*, 123–127.
- Rich, R. *The Knowledge Inquiry System: Critical Issues and Perspectives*; Key Note Address at Center for Knowledge Transfer: Alberta, Canada, 2002.
- Lester, J. The utilization of policy analysis by state agency officials. *Knowledge* **1993**, *14*, 267–290.
- Oh, C.H.; Rich, R. Explaining use of information in public policy making. *Knowl. Policy* **1996**, *9*, 3–35.
- Rich, R.F. Knowledge creation, dissemination, and utilization. *Knowledge* **1991**, *12*, 319–337.
- Cousins, J.B.; Leithwood, K.A. Enhancing knowledge utilization as a strategy for school improvement. *Knowledge* **1993**, *14*, 305–333.
- Slappendel, C. Knowledge use versus capability development. *Sci. Commun.* **1994**, *16*, 195–205.
- Florio, E.; DeMartini, J.R. The use of information by policymakers at the local community level. *Knowledge* **1993**, *15*, 106–123.
- Hultman, G.; Horberg, C.R. Knowledge competition and personal ambition. *Sci. Commun.* **1998**, *19*, 328–348.
- Tyden, T. The contribution of longitudinal studies for understanding science communication and research utilization. *Sci. Commun.* **1996**, *18*, 29–48.
- Hall, R. Knowledge use and the dynamics of managing curriculum change. *Sci. Commun.* **1997**, *18*, 342–361.
- Hutchinson, J. A multimethod analysis of knowledge use in social policy. *Sci. Commun.* **1995**, *17*, 90–106.
- Meillier, L.; Lund, A.B.; Gerdes, L.U. The backpack function of health education: use of knowledge types concerning prevention of coronary heart disease. *Sci. Commun.* **1997**, *18*, 216–237.
- Castill, A. Communication and utilization of science in developing countries. *Sci. Commun.* **2000**, *22*, 46–72.
- Hultberg, J. The two cultures revisited. *Sci. Commun.* **1997**, *19*, 194–215.
- Landry, R.; Amara, N.; Lamari, M. Climbing the ladder of research utilization. *Sci. Commun.* **2001**, *22*, 396–422.
- Logan, J.; Graham, I.D. Toward a comprehensive interdisciplinary model of health care research use. *Sci. Commun.* **1998**, *20*, 227–246.
- Rich, R.F.; Oh, C.H. Rationality and use of information in policy decisions: a search for alternatives. *Sci. Commun.* **2000**, *22*, 173–211.
- Boggs, J. Implicit model of social knowledge use. *Knowledge* **1992**, *14*, 29–62.
- Estabrooks, C.; Scott-Findley, S.; Winther, C. Knowledge utilization in nursing and the allied health sciences. In *Multidisciplinary Perspectives on Evidence-Based Decision Making in Health Care*; Lemieux-Charles, L., Champagne, F., Eds.; University of Toronto Press, 2003.
- Weiss, C. Measuring the use of evaluation. In *Utilizing Evaluation: Concepts and Measurement Techniques*; Ciarlo, J.A., Ed.; Sage: Beverly Hills, 1981.
- Knott, S.J.; Wildavsky, A. If dissemination is the solution, what is the problem? *Knowledge* **1980**, *1*, 537–578.
- Rich, R.F., Ed.; *The Knowledge Cycle*; Sage: Beverly Hills, 1981.
- Caplan, N. The two communities theory and knowledge utilization. *American Behavioral Scientist* **1979**, *22* (3), 459–470.
- Webber, D. Legislators' use of policy information. *Am. Behav. Sci.* **1987**, *30*, 612–631.
- Weiss, C.; Gruber, J. Using knowledge for control in fragmented policy arenas. *J. Policy Anal. Manage.* **1984**, *3*, 225–247.
- Nilsson, K.; Sunesson, S. Conflict or control: research utilization strategies as power techniques. *Knowl. Policy* **1993**, *6*, 23–36.
- Webber, D. Legislators use of policy information. *American Behavioral Scientist* **1987**, *30* (6), 612–631.
- Whiteman, D. The fate of policy analysis in congressional decision making. *West. Polit. Q.* **1985**, *38*, 294–311.

32. Weiss, C. Many meanings of research utilization, unpublished, 1978.
33. Rich, R.F.; Oh, C.H. The utilization of policy research. In *Encyclopedia of Policy Studies*; Nagel, S.S., Ed.; Marcel Dekker: New York, 1993.
34. Oh, C.H. *Linking Social Science Information to Policy Making*; The JAI Publishers: Greenwich, 1996.
35. Nachmias, D. The role of evaluation in public policy. *Policy Stud. J.* **1980**, *8*, 1163–1169.
36. Weiss, C.H.; Bucuvalas, M.J. Introduction. In *Using Social Science in Public Policy Making*; Lexington Publishers: Lexington, 1977.
37. Melody, W.H. Information: an emerging dimension of institutional analysis. *J. Econ. Issues* **1987**, *21*, 1313–1339.
38. Jonston, D.F. Cemus concepts as knowledge filters for public policy advisors. *Knowledge* **1983**, *5*, 99–119.
39. Nelson, C.E.; Roberts, J.; Maederer, C.M.; Wertheimer, B.; Johnson, B. The utilization of social science information by policy makers. *Am. Behav. Sci.* **1987**, *30*, 569–577.
40. Weiss, C.H. Knowledge creep and decision accretion. *Knowledge* **1980**, *1*, 381–904.
41. Caplan, N. What do we know about knowledge utilization as field/discipline. In *Paper Presented at the Research Utilization Conference*; University of Pittsburgh, 1978.
42. Mandell, M.B.; Sauter, V.L. Approaches to the study of information utilization in public agencies. *Knowledge* **1984**, *6*, 145–164.
43. Aldrich, J. Rational choice and turnout. *Am. J. Polit. Sci.* **1993**, *37*, 246–278.
44. Allison, G.T. *The Essence of Decision*; Little, Brown: Boston, 1972.
45. Braybrook, D.; Lindbloom, C.E. *A Strategy for Decision*; Free Press: New York, 1972.
46. Majone, G. *Evidence, Arguments, and Persuasion in the Policy Process*; Yale University Press: New Haven, 1989.
47. March, J.G., Ed.; *Decisions and Organizations*; Basil Blackwell: Cambridge, 1988.
48. March, J.G.; Simon, H. *Organizations*; Wiley: New York, 1958.
49. Cook, K.S.; Levi, M. *The Limits of Rationality*; The University of Chicago Press: Chicago, 1990.
50. Elster, J. When rationality fails. In *The Limits of Rationality*; Cook, K.S., Levi, M., Eds.; The University of Chicago Press: Chicago, 1990.
51. Denhardt, R. *Theories of Public Organization*; Wadsworth: New York, 1993.
52. Elster, J. *The Cement of Society*; Cambridge University: Cambridge, U.K., 1989.
53. Feldman, M.; March, J.G. Decision in organizations and theories of choice. In *Prescripts on Organization Design and Performance*; Van De Ven, A., Joyce, W.J., Eds.; John Wiley: New York, 1981.
54. Calvert, R.L. The value of biased information: a rational choice model of political choice. *J. Polit.* **1985**, *47*, 53L–555.
55. Mckelvey, R.D.; Ordeshook, P.C. Elections with limited information. *J. Econ. Theory* **1985**, *36*, 55–85.
56. Riker, W. *The Theory of Political Coalitions*; Yale University Press: New Haven, CT, 1962.
57. Huberman, M. Steps toward an integrated model of research utilization. *Knowledge* **1987**, *8*, 586–611.
58. Oh, C.H. Rationality model of information utilization in policy making: an empirical test. *Korean J. Policy Stud.* **1998**, *7*, 195–225.
59. Oh, C.H. Information utilization in government decision making: why and how. *J. Gov. Stud.* **1999**, *5*, 38–74.
60. Halperin, M.H. Shaping the flow of information. In *Bureaucratic Power in National Politics*; Rourke, F., Ed.; Little, Brown: Boston, 1974.
61. Henry, N. *Public Administration and Public Affairs*; Prentice-Hall: Englewood Cliffs, 1995.
62. Oh, C.H. Information searching in government bureaucracies. *Am. Rev. Public Admin.* **1996**, *26*, 41–70.
63. Oh, C.H. Explaining the impact of information on problem definition. *Policy Stud. Rev.* **1998**, *15*, 109–136.
64. Weiss, C. The use of social science research. In *Organizations and the Human Services*; Stein, H., Ed.; Temple University: Philadelphia, 1981.
65. Macrae, D.; Wilde, J. *Policy Analysis for Public Decisions*; University Press of America: New York, NY, 1985.
66. Innes, J.E. *Knowledge and Public Policy*; Transaction Publishers: New Brunswick, 1990.
67. Caplan, N. The two communities theory and knowledge utilization. *Am. Behav. Sci.* **1979**, *22*, 459–470.
68. Dunn, W.N. The two communities metaphor and models of knowledge utilization. *Knowledge* **1980**, *1*, 515–537.
69. Oh, C.H. Issues for the new thinking of knowledge utilization. *Knowl. Policy* **1997**, *10*, 3–10.
70. Weiss, C.; Bucuvalas, M.J. *Social Science Research and Decision Making*; Cambridge University Press: New York, 1980.
71. Oh, C.H. *Science and Politics: From an Information Utilization Perspective*, Paper Presented at the World Congress of the International Political Science Association. Quebec, Canada, 20040101, 2000.
72. Bardach, E. The dissemination of policy research to policy makers. *Knowledge* **1984**, *6*, 125–145.
73. Rich, R.F.; Oh, C.H. *Knowledge and Public Policy: A Framework for Analysis*, Paper Presented at the Annual Meeting of the Midwest Political Science Association. Chicago, IL, 20180101, 1996.
74. Knorr, K. Policy makers' use of social science knowledge: Symbolic or instrumental? In *Using Social Science Research in Public Policy Making*; Weiss, C.H., Ed.; Lexington Books: Lexington, 1977.
75. Zaltman, G. Knowledge utilization as planned social change. *Knowledge* **1979**, *1*, 82–105.
76. Bulmer, M. Applied social research: a reformulation of "applied" and "enlightenment" models. *Knowledge* **1981**, *3*, 187–209.
77. Rein, M.; White, S.H. Policy research: belief and doubt. *Policy Anal.* **1977**, *3*, 239–271.
78. Booth, T. Researching policy research: issues of utilization in decision making. *Knowledge* **1990**, *12*, 80–100.
79. Pollard, W. Decision making and the use of evaluation research. *Am. Behav. Sci.* **1987**, *30*, 661–676.

80. Bozeman, B.; Bretschneider, S. Public management information system: theory and practice. *Public Admin. Rev.* **1986**, *46*, 475–487 (special issue).
81. Dunn, W. Measuring knowledge use. *Knowledge* **1983**, *5*, 120–133.
82. Van De Vall, M.; Bolas, C. Internal vs. external social policy research. *Knowledge* **1981**, *2*, 461–481.
83. Lindblom, C.E.; Cohen, D.K. *Usable Knowledge: Social Science and Social Problem Solving*; Yale University Press: New Haven, 1979.
84. Horowitz, I.L. *The Rise and Fall of Project Camelot*; Transaction Books: New Brunswick, 1975.
85. Gregg, G.; Preston, T.; Geist, A.; Caplan, N. The caravan rolls on: forty years of social problem. *Knowledge* **1979**, *1*, 31–61.
86. Campbell, D.T.; Stanley, J.C. *Experimental and Quasi-Experimental Design for Research*; Rand McNally and Company: Chicago, 1966.
87. Cook, T.D.; Campbell, D.T. *Quasi-Experimental Designs*; Rand McNally: Chicago, 1986.
88. Pelz, D.C.; Horsley, J.A. Measuring utilization of nursing research. In *Utilizing Evaluation: Concepts and Measurement Techniques*; Ciarlo, J., Ed.; Sage Publications: Beverly Hills, 1981.
89. Nagel, S. *Public Policy: Goals, Means, and Methods*; St. Martin: New York, 1984.
90. Shrivastava, P. Knowledge systems for strategic decision making. *J. Appl. Behav. Sci.* **1985**, *2*, 95–107.
91. Kelly, P., Kranzberg, M., Eds.; *Technological Innovation: An Critical Review of Current Knowledge*; San Francisco Press: San Francisco, 1978.
92. Backer, T. Knowledge utilization: the third wave. *Knowledge* **1991**, *12*, 225–240.
93. Larsen, J. Knowledge utilization. In *The Knowledge Cycle*; Rich, R., Ed.; Sage: Beverly Hills, 1981.
94. Landecker, W.S. The use of survey research in the sociology of knowledge. *Knowledge* **1982**, *4*, 73–94.
95. Dunn, W.; Holzer, B. Anatomy of an emergent field. *Knowl. Soc.* **1988**, *1*, 3–26.
96. Lowi, T.J. American business, public policy, case studies and political theory. *World Polit.* **1964**, *16*, 677–715.
97. March, J.G.; Olsen, J.P. *Ambiguity and Choice in Organizations*; Universitetsforlaget: Bergen, 1979.
98. Leviton, L.C.; Hughes, E.F.X. Research on the utilization of evaluation. *Eval. Rev.* **1981**, *5*, 525–548.
99. Moe, T. On the scientific status of rational models. *Am. J. Polit. Sci.* **1979**, *23*, 215–243.
100. Ostrom, E. Rational choice theory and institutional analysis. *Am. Polit. Sci. Rev.* **1991**, *85*, 237–243.
101. Wingens, M. Toward a general utilization theory. *Knowledge* **1990**, *12*, 27–42.

Labor Policy in Thailand

Andrew Brown

*School of Social Science, University of New England, Armidale,
New South Wales, Australia*

INTRODUCTION

Historically, labor policy in Thailand has emerged from contests that have involved competing constellations of economic, political, and ideological interests^[1] Recently, these contests have been dominated by those particular interests aligned with Thaksin Shinawatra's Thai Rak Thai party (TRT) which formed a government in January 2001. During its first term in office, the TRT government has implemented economic and social policies that embody a broad twofold objective: first, to restructure economy, politics, and society in order to build a globally competitive capitalism; second, to manage the conflicts and tensions that processes of restructuring have generated. Labor policies both reflect and form a significant component of the government's pursuit of these dual objectives.

THAI RAK THAI AND ITS HEGEMONIC VISIONS

Thaksin Shinawatra and the TRT electoral victory signaled the rise to state power of a new alliance of economic, social, and political interests dominated by big domestic capitalist groups.^[2] Over the past 4 years, the new government has engaged in a singular attempt to reform and rejuvenate a political economy devastated by the 1997/1998 economic crisis. In the economic arena, the government has expressed a desire to free Thailand from past growth strategies that emphasized the export of commodities with low-valued added content produced by cheap unskilled wage-labor. Rather, the administration has emphasized the need for new skills, increased productivity, value adding, industrial upgrading, and the creation of a more sophisticated and globally engaged knowledge economy. Recognizing that the attainment of these aims requires long-term political stability, the TRT government has embedded its control over the state bureaucracy, circumscribed the operations of independent regulatory bodies, cemented its domination over parliament, curbed the demands of civil society, and through a range of campaigns, attempted to create a more ordered and disciplined society. Finally, the Thaksin government has also engaged in a political process of developing a new social contract, the functions of

which are twofold: first, to "provide a 'cushion' against social disorder...[and]...prevent political problems destabilizing entrepreneur-led growth;" and second, to "encourage the Thais to be more entrepreneurial." Both are meant to serve as a strategy for "deepening and expanding capitalism."^[3,4]

It is within this context of rebuilding domestic capitalism by pursuing new development directions, entrenching its political rule, instilling discipline in society, and developing a new social contract that the Thaksin government has addressed itself to Thailand's labor problems.

THAI RAK THAI AND LABOR PROBLEMS

Throughout the late 1980s and 1990s, as Thailand became embedded in global patterns of flexible production, labor policy became an increasingly complex arena of debate in terms of both the sheer scale of problems that emerged and the range of competing interests that attempted to shape policy outcomes. In addressing itself to the country's labor problems and navigating its way through the plethora of competing voices and interests, the Thaksin government has produced two clusters of policy interventions. During its first 2 years in office, government responses to labor issues were shaped by its overriding aim to reignite economic growth, promote investment, increase employment opportunities, inculcate a spirit of entrepreneurship, as well as assist business restructuring. During the latter half of its first term, the government has begun to address itself to structural issues such as bureaucratic reform, skilling, labor standards, as well as deal with some of the tensions generated by structural change and labor's deepening exposure to the vagaries of market forces.

Reigniting Growth and Meeting the Needs of Domestic Business

During its first 2 years in office, the Thaksin government's main priority was to reignite economic growth through a strategy of increased government spending. Using funds from the Government Savings Bank and the Government Agricultural Bank, the TRT

government introduced a 1 million baht community fund, established a People's Bank, and also launched the One District One Product Scheme. Other government banks provided credit for small and medium enterprises and a debt moratorium was extended to farmers. Government also endeavored to reduce household expenses, thereby releasing funds that could be used for consumption in other areas. To this end, a new 30 baht health care system was introduced, together with a cheap life insurance scheme, low-cost housing, supplies of cheap computers, and televisions, as well as subsidized credit for the purchase of taxis and bicycles. Combined, these measures aimed to make credit available to households and small and medium size firms, stimulate domestic demand, create employment opportunities, as well as foster a spirit of entrepreneurship.^[5]

While these policies aimed to reignite economic growth, investment, and expand employment, the government also assisted employers as they reorganized and restructured their businesses. For those that survived the 1997/1998 economic crisis, the postcrisis years had seen firms, across a range of industries, pursue strategies designed to reduce costs of production and wages in the face of intense local, regional, and global competitive pressures. Such strategies have included widespread retrenchment and downsizing of workforces, reducing wages and benefits, experimenting with flexible forms of employment, introducing new technologies, implementing new forms of human resource management, as well as spatially relocating operations, often to border areas, in order to take advantage of cheaper foreign labor. These processes of restructuring have been bitterly contested by workers as they have struggled to retain jobs, maintain conditions of employment, and secure fundamental labor relations reform that would better guarantee their legal rights to compensation, wages, benefits, and workplace participation.^[6]

Prior to the 2001 elections, TRT recognized many of the problems that workers faced as business was restructured.^[7] During its first 2 years in office, however, the government failed to deliver on its preelection promises to provide better protection for workers by closing loopholes in key labor laws and to initiate processes of more fundamental legislative and institutional reform. Nor did the government do anything to bring a halt to the union busting that has become a feature of the industrial relations arena since the economic crisis of 1997/1998. Indeed, in what can be seen as an important signal to business generally, after Thaksin's own Shin Corp purchased independent TV broadcaster iTV, it sacked 21 employees all of whom were members of the newly established union, with 9 of those sacked on the union board. Through this leading by example and through policy inertia,

the government effectively gave domestic capital a valuable breathing space within which employers could, without fear of government intervention, restructure their workforces, experiment with more flexible forms of labor, rid themselves of troublesome unions and activists and generally evade their industrial and social responsibilities to provide safe working environments and reasonable remuneration for work performed as stipulated by various laws and regulations. At the same time, this failure to protect basic rights and allow organized labor to be progressively weakened formed part of a wider process through which the Thaksin government sought to reconfigure domestic politics by curbing the demands of civil society groups.

Bureaucratic Reforms and Addressing Structural Issues

While the Thaksin government dawdled on labor relations reform during the first 2 years of its term, it has been more active since late 2002. This coincides with some initial attempt to pursue longer-term goals of utilizing labor skills and other national assets that can be deployed toward the objective of moving up to higher-valued production and a knowledge economy. The TRT's intentions were signaled by the removal of the incompetent labor minister Det Bunlong and the installation of Suwat Liptapanlop. This coincided with a significant period of restructuring as the former Ministry of Labor and Social Welfare was reformed to become the Ministry of Labor (MOL), while a new Ministry of Social Development and Human Services was created. Suwat appointed a team of 10 advisors and promised a solution to 23 key labor problems within 2 years.^[8] He also oversaw significant modernizing reform of Ministry standards, practices, and procedures to ensure that processes of data collection and statistical analysis were streamlined, that staff receive improved training to cope with changing technologies, and that the Ministry will have an enhanced capacity to engage in international dialog on labor issues. In particular, the Ministry assumed as a major function the creation of a better skilled workforce and ensuring that workers receive remunerations in line with international standards and skill levels. Apart from better catering to the skill requirements of the domestic political economy, the Ministry has also adopted an enhanced role to promote employment and create new markets for Thais wishing to work overseas.^[9] At the same time, Suwat presided over the ousting of several senior labor officials who had been involved in a job placement scam involving hundreds of millions of baht.^[10] These reforms to a Ministry that has long been considered to be incompetent and corrupt, in part, reflect the government's desire

to meet the challenges of regional and global competition and develop a more knowledge-intensive economy and a better-skilled labor force.

Developing Skills and Meeting International Standards

In pursuing new growth strategies, TRT recognized the limitations of Export Oriented Industrialisation (EOI) that relied on the use of cheap unskilled labor. Rather, as noted above, the government has stressed the need to develop a better-skilled and more productive workforce that will meet the changing requirements of local industry as well as the lucrative labor export market. In 2002, the government passed a law on the Promotion of Labor Skills and launched a number of programs to foster public-private sector cooperation to develop the nation's human resources. These initiatives have included enhanced cooperation with the private sector in establishing industry skills standards and developing centers to train workers to meet prescribed standards; a cooperative scheme between the Skills Development Department, universities, and employers, to establish workplace schools to upgrade worker's literacy and numeracy; improved workplace training for vocational students; a scheme that allows workers to convert working experience into educational certification; promotion of employment through local knowledge in food production and cooking, massage, child and aged care, and housekeeping; benchmarking and raising standards for 149 targeted occupations to ensure that labor meets local and international market demands; and the establishment of employment centers that will better match workers' skills with employers' needs (see Ref.^[6], p. 17–18).

There are least two motivations for these initiatives. The first is economic. Following the crisis there was a realization that competing economies have been deregulating labor markets, and that there was the potential for Thailand to slip behind, especially as both TRT and the previous government had moved slowly on labor relations reform. The second, related, motivation for these policies was in response to concerns expressed by some politicians, business people, and officials about the ways in which Thailand's poor record on labor standards might negatively impact on foreign investment and access to developed country markets.

This thinking about the competitive need for a more socially responsible business sector saw the government emphasize international standards when it ratified ILO Convention 182 on child labor in 2001 and in May 2004, the government ratified the related ILO Convention 138. These initiatives were accompanied by other measures to protect children from workplace abuse. In December 2001, the government directed the

MOL to establish a Labor Standards Project to educate employers on the need to adopt international practices in order to avoid trade sanctions, improve productivity, and minimize disputation. In 2002, a substantial government budget was earmarked for preparing employers and workers for further trade liberalization. In 2003, a *Thai Labor Standards 2003: Thai Corporate Social Responsibility* document was issued, covering issues such as forced labor, remuneration, working hours, discrimination, penalties and fines, child labor, gender issues, freedom of association, collective bargaining, and occupational health and safety. For each category, basic provisions and standards were established and employers were encouraged to conform.

The New Social Contract and Managing Conflict

The crisis and its social impacts combined with subsequent processes of work and industry restructuring generated significant labor protest (see Ref.^[6], p. 15–16). In the private sector there have been a number of long-running disputes and protests at particular factories some of which have attracted international attention. Over the same period, the Labor Relations Committee has had to mediate hundreds of disputes and grievances that collectively have involved thousands of workers. Moreover, the Labor Courts have become choked with numerous unresolved cases. In league with their unions, labor affiliated nongovernment organizations (NGOs), human rights activists, as well as through links with various labor campaigns outside the country, Thai workers have pressed government to enforce existing labor laws as well as produce a reformed body of legislation that will close the legal loopholes that employers have been exploiting as they have restructured their businesses.

While still not directly addressing many of the issues that have generated industrial tensions, the TRT government considers that its new social contract will offer improved protection to the poor, be acceptable to workers, and deliver continued political support. The social contract is constructed around five TRT election policies: the million baht village and community fund (VCF), a 30 baht universal health scheme (UHS), the people's bank, the one district-one product scheme, and a farmer's debt moratorium. The projected cost of these policies was more than US\$3.6 billion, with the VCF and UHS accounting for about 90% of this, and claiming 52 million beneficiaries by late 2002.

During the last half of its first term, the Thaksin government has also implemented a range of policies that have specifically targeted workers. Social security

coverage has been extended to those employed in small firms as well as to agricultural and some casual workers. In January 2004 a new unemployment scheme was finally introduced. Workers have also been promised low-cost housing as well as improved educational opportunities for their children. The government has also begun to address some of the problems faced by working mothers by establishing child care centers in key industrial areas. A very recent Ministerial Order also extends to home based workers, many of whom are women, improved protection with respect to wages, the delivery of completed work and health and safety. The government has also finally begun to consider a draft labor bill recently released by the Council of State. The proposed new law will replace the long outdated 1975 Labor Relations Act which governs collective bargaining and rights to organize.

CONCLUSIONS

During its first term, the Thaksin government faced the task of balancing the requirements of capital restructuring against the necessity of maintaining social stability and cohesion as well as securing ongoing political support for its policies. The government's interventions on the labor front both reflect and form an important component of this broad endeavor. During the first 2 years of its term, the government acted at a glacial pace on labor reform, which gave domestic business the space to reestablish its competitiveness by reducing labor costs. Over the latter part of its first term, the government has been more active in addressing problems of bureaucratic reform, skilling, and labor standards. It has also attempted to deal with some of the tensions generated by restructuring by extending social security coverage, inaugurating an unemployment scheme, as well as offering some improved education and skilling opportunities,

promoting employment overseas and cleaning up labor recruitment practices.

ACKNOWLEDGMENTS

Thanks are due to Professor Kevin Hewison for comments on an earlier draft of this article.

REFERENCES

1. Brown, A. *Labour, Politics and the State in Industrializing Thailand*; Routledge Curzon: London, 2003.
2. Phongpaichit, P.; Baker, C. *Thaksin: The Business of Politics in Thailand*; Silkworm Books: Chiang Mai, 2004.
3. Phongpaichit, P.; Baker, C. The only good populist is a rich populist: Thaksin Shinawatra and Thailand's Democracy: SEARC Working Paper 36; City University of Hong Kong, 2002; 10 pp.
4. Hewison, K. Crafting a new social contract: domestic capitalist responses to the challenges of neoliberalism. In *Radicalising Thailand: New Political Perspectives*; Ungpakorn, J., Ed.; Institute of Asian Studies, Chulalongkorn University: Bangkok, 2003; 120–151.
5. Phongpaichit, P.; Baker, C. *Thaksin: The Business of Politics in Thailand*; Silkworm Books: Chiang Mai, 2004; 99–133.
6. Brown, A.; Hewison, K. Labour politics: Thaksin's Thailand: SEARC Working Paper 62; City University of Hong Kong, 2004.
7. Thai Rak Thai. The development of human resources and the building of social safety nets (in Thai). Raengnagan Porithat **2001**, 15 (2), 17–18.
8. Bangkok Post Jan 3, 2003.
9. Ministry of Labour. *Policies, Orientation, Commitments* (in Thai); http://www.molsw.go.th/151502e6527d538d19feffbd848cc3e/Language_Template/TH/ministry02w.html (accessed Mar 13, 2003).
10. The Nation Feb 27, 2003.

Labor–Management Partnerships and Cooperation

W. David Patton

*Center for Public Policy and Administration, Boise State University,
Boise, Idaho, U.S.A.*

INTRODUCTION

Labor relations in the United States developed in an atmosphere of conflict between management and labor. The laws and practices of labor–management relations that were established to mitigate this conflict continue to guide labor relations today and reflect our adversarial tradition. However, as a result of labor–management strife and national emergencies, such as the Great Depression and world wars, organized labor and management began working together to resolve their differences. In recent years, there has been increasing pressure for management and labor to work together to solve common problems through labor–management cooperation in the form of committees or partnerships.

BENEFITS OF LABOR–MANAGEMENT COOPERATION

Rancorous battles between management and labor are not only unpleasant to observe but have serious impacts on productivity and efficiency. In the year 2000, over 20 million workdays were lost due to strikes in the United States.^[1] Disputes between management and labor are often settled by formal bargaining, arbitration, or legal processes rather than through collegial discussion. This process is time-consuming and results in settlements that are often unsatisfactory to either side. Labor–management cooperation is a process and a different approach to labor relations. Employees are involved through their representatives in the day-to-day decisions of public and private organizations. Issues of interest to management and labor are brought before joint committees for discussion, study, and resolution.

Evidence shows that the benefits of cooperation as opposed to adversarialism can be substantial. For example, the Social Security Administration improved its customer satisfaction ratings through its partnership with the American Federation of Government Employees (AFGE). The Internal Revenue Service has improved the quality of their services through joint training and streamlining programs in cooperation with the National Treasury Employees Union (NTEU). A joint productivity effort by the Customs

Service and NTEU increased the amount of illegal narcotics seized by 42%. The Social Security Administration, Bureau of Reclamation, Department of Health and Human Services, and other federal agencies have saved over \$10 million largely due to decreases in litigation over grievances and charges of unfair labor practices. Other benefits include improved quality of work life and better labor–management relations.^[2]

Although most people would prefer situations where parties of differing interests cooperate rather than fight one another, there are reasons for the adversarial context of bargaining and benefits similar to the benefits of the adversarial structure of our legal system. An adversarial system allows for a process where the viewpoints of both parties can be heard. The positions and interests of each side can be clearly stated, and accommodations to these interests can be negotiated. In an adversarial system, power is distributed more evenly, and it is more difficult for one side to dominate the other. For this reason, federal law has prohibited employer involvement in union activities. Finally, in the event that the interests of both sides are irreconcilable through negotiation, an adversarial system allows for the exercise of economic and political power through sanctions such as the strike or boycott. Many adversarial labor–management systems call for third-party intervention when both sides are at an impasse, using mediation or arbitration, in which resolutions to disputes can be made by an objective outside party. In some cases, the adversarial nature of traditional collective bargaining can be more appropriate than labor–management partnerships. Labor–management cooperation is not intended to replace collective bargaining but can be an important tool where appropriate to resolve workplace issues.

THE DEVELOPMENT OF LABOR–MANAGEMENT COOPERATION

The earliest forms of labor–management cooperation were committees in the textile and railroad industries—both hotbeds of labor unrest in the early 1900s. The National War Labor Board sponsored labor–management productivity committees during World War I, and during World War II, over 5000 labor–management committees were formed to

improve productivity in American industries.^[3] Following the war, these committees were abandoned by industry, which probably contributed to the record number of strikes in 1946.^[4] More cooperative labor-management relations became the objective of legislation passed between in 1947 and 1978. The Labor-Management Relations Act of 1947 (popularly known as the Taft-Hartley Act) authorized the Federal Mediation Service as an alternate means of settling labor disputes. The Civil Service Reform Act of 1978 allowed collective bargaining on limited subjects in the Federal Government and established the Federal Labor Relations Authority that hears and decides cases on negotiability, unfair labor practices, and exceptions to arbitration awards. The Labor Management Cooperation Act of 1978 authorized the Federal Mediation and Conciliation Service to assist in the establishment of labor-management committees.^[5]

PUBLIC-SECTOR LABOR-MANAGEMENT PARTNERSHIPS

The first public-sector labor-management committees were formed in the mid-1960s and have expanded throughout federal, state, and local levels of government.^[3] Labor-management relations in the federal government, like the private sector, have been marked by conflict since federal employees were allowed some bargaining rights by executive order in 1962 and later in the Civil Service Reform Act of 1978. A 1991 General Accounting Office report concluded that relations between management and labor in the federal government were “overly legalistic, too adversarial, and bogged down by litigation,” and called for reform that would motivate cooperation and productivity. In its report on federal productivity and performance, the National Performance Review recommended that the federal government establish labor-management partnerships to “transform the adversarial relationship that dominates federal union-management interaction into a partnership for reinvention and change.”^[6]

EXECUTIVE ORDERS OF PRESIDENT CLINTON AND PRESIDENT BUSH

Shortly after the NPR report, President Clinton issued Executive Order 12871 establishing the National Partnership Council to promote labor-management partnerships in the federal government. The Order also directed Executive agencies to establish labor-management committees to help reform government, include employees and their union representatives as full partners in these committees, provide training in dispute resolution and interest-based bargaining, and

evaluate their progress. The most controversial directive in Executive Order 12871 required federal managers to include the typically management functions of how employees are assigned and how the work is to be accomplished as “mandatory” bargaining subjects. These subjects had formerly been considered “management rights” and could only be the subject of bargaining if the agency chose to include them in negotiations. President Clinton reaffirmed his commitment to partnerships in a 1999 memorandum to executive agencies of the federal government.

The response to Executive Order 12871 among federal agencies has been mixed. Most of the agencies reporting on their progress with labor-management partnerships^[2] said they had formed partnership councils. However, responses to the other requirements of the Executive Order were less than enthusiastic. Less than half of responding agencies reported substantial union involvement. Nearly half of the agencies were providing dispute resolution training, but less than one-sixth said they had agreed with their unions to bargain over employee assignments and the way work is done. Most agencies were reluctant or had refused to bargain over these subjects, which seriously damaged the confidence of unions in the commitment of management to the partnership concept. Similarly, few agencies had evaluated the effect of labor-management partnerships on the organization or developed quantitative data to support conclusions that partnerships had improved performance and productivity within federal agencies. Of those reporting results, there is some evidence that labor-management partnerships have been successful in improving customer service, quality of service, productivity, efficiency, cost savings (usually a result of less litigation in labor relations), and improvements in the quality of work life and labor-management relations.

On February 17, 2001, shortly after taking office, President George W. Bush issued Executive Order 13203 revoking Executive Order 12871 and subsequent executive orders establishing labor-management partnerships in federal agencies.^[7] President Bush most likely had political, strategic, and philosophical reasons for revoking authorization for labor-management partnerships.^[8] The fundamental reason for revoking labor-management partnerships was ideological. Supporters with influence on the President believed these partnerships “elevated the status of federal unions to a level of equality with agency management.”^[9] This view saw union encroachment on management authority, displacing political accountability from citizens to union members. The U.S. Office of Personnel Management subsequently issued guidance to federal agencies for implementing President Bush’s executive order, saying “Agencies have discretion . . . to adopt a labor relations strategy best suited to their own needs.”^[10]

TYPES OF LABOR-MANAGEMENT COMMITTEES

Every committee developed to formalize labor-management cooperation is designed to meet the needs of the people involved, and therefore, every arrangement is somewhat different. However, successful labor-management committees have some common characteristics. Because labor-management cooperation began spontaneously in reaction to crises or with the objective of being more productive, labor-management partnerships take on many forms. Almost all labor-management committees can be categorized into three structural types: work-site, industry, and area labor-management committees.

The most common form of partnership is at the most local level, the work-site labor-management committees (W-LMC). Work-site partnerships are designed by union and management leaders to address the needs of each workplace and address common issues. W-LMCs may seek to improve communication between management and labor, address issues before they come to formal grievance situations, or study issues so each side has consistent information. An example of a work-site labor-management committee is the agreement between the Port of Baltimore and International Longshoremen's Association and Steamship Trade Association. This partnership seeks to improve the competitive position of the Port of Baltimore and the skills of its workforce. A second form of cooperative arrangement is the industry labor-management committee (I-LMC). These committees are defined by a geographical area and specific industry and provide a forum for discussion of issues affecting both management and labor. An example of an I-LMC is The Construction Industry Service Corporation (CISC) in Northeastern Illinois. CISC provides a scholarship fund, apprenticeship program, and construction financing services. The next level of partnership is the area labor-management committee (A-LMC) that serves a community or regional area. A-LMCs assist in developing work-site committees, promoting communication, improving work conditions, encouraging job retention, skill training and economic development. The Philadelphia Area Labor-Management Committee (PALM) is a partnership between the AFL-CIO and the Greater Philadelphia Chamber of Commerce. PALM provides labor-management relations assessments, team-building training, work-site productivity, quality and safety designs, and third-party issue resolution.^[11]

Labor-management committees are the formal structures of cooperation between management and labor, but these structures cannot sustain themselves without the commitment of organizational leaders from management and labor to the philosophy,

principles, and processes of labor-management cooperation. Typically, the committee organization would allow for a structured meeting of representatives from management and labor, open access to information, and cooperative problem solving. Employees are included in the process of day-to-day decisions where their jobs are affected. Leaders and employees in the organization operate under the philosophy that cooperation will allow for higher productivity and a more enjoyable work environment.

CHARACTERISTICS OF SUCCESSFUL LABOR-MANAGEMENT COOPERATION

The experiences of labor-management cooperation efforts throughout government have produced a consistent set of characteristics for successful partnerships. These characteristics include attributes common to any collective activity and highlight the context, processes, and purposes of the partnership. To operate a successful labor-management partnership, it is helpful to incorporate the following principles.^[12,12]

1. Common understanding of the nature of the partnership: All interested parties to the partnership should understand the need for cooperation and the purposes for the creation of the partnership. The roles of management and labor are different from typical labor relations and need to be resolved before progress can be made.
2. Mutual respect and trust: Successful partnerships display respect for the positions held by both sides and respect for the people involved. Trust is developed over time but is crucial to cooperative efforts to solve problems and prepare plans for the future. Often, the partnership is based on a strong personal relationship between management and labor.
3. Broad-based support: Support for a cooperative approach to labor relations must come from all levels of the involved organizations. Commitment to the philosophy and processes of collaborative problem solving needs to come from the leadership of union and management as well as mid-level managers and front-line representatives who will be involved in the day-to-day activities of the partnership. The principles of the partnership should be incorporated into the everyday operations of the organization.
4. Long-term outlook: The current system of adversarial labor relations developed over many decades of trial and error. A major change to a cooperative approach will take

time. The parties to the partnership have to understand that trust and respect are earned and will come with experience. Complex problems require sufficient time for deliberation and for solutions to gain acceptance. Participants in the partnership should develop a positive vision of the future and work toward that goal.

5. Investment in the process: To effectively work in partnerships, management and labor representatives should undergo training in group processes, problem solving, and interest-based negotiation. In addition to training, management should be prepared to allow for the time needed for effective meetings of the partnership.
6. Mutual objectives: The objectives of the partnership should be in the interest of all those involved in the process. Agreement should begin with the selection of objectives throughout the duration of the partnership. Everyone should benefit from his or her participation.
7. Appropriate issue selection: It is important to the success of the partnership that it experiences quick, even if small, wins to demonstrate the viability of the partnership. Eventually, however, it is important that the partnership work on more complex problems such as productivity improvement or reorganization. The cooperative approach to problem solving works best when there is some convergence of the interests of the participants. When there is little room for compromise, it is sometimes better to rely on other methods of dispute resolution such as collective bargaining. Successful partnerships have separated their work from the collective bargaining process and carefully decided which issues to address in either forum.
8. Proactive rather than reactive: The philosophy of labor-management partnerships lends itself to thoughtful deliberation and strategic planning. The committee should consider issues, opportunities, and threats that present themselves in the future for action of the partnership.
9. Political neutrality: The influence of political ideology or organizational factions can seriously damage the credibility of labor-management partnerships. Cooperation is neither conservative nor liberal; it is a philosophy addressing an effective approach to solving issues of mutual concern to employers and employees in organizations. Similarly, the existence of hidden agendas of those participating in partnerships can harm the relationships needed for effective cooperation.

10. Flexibility: One of the most enduring characteristics of public organizations is the element of change—as the shift in executive orders on labor-management partnerships demonstrate. Partnerships need to be adaptive and responsive to the needs of the employer and employees. As changes occur, the partnership needs to be ready to take on issues as they present opportunities or threats to the organization.

CONCLUSION

For more information, consult the following resources: the National Labor-Management Association, the Labor Management Partnership Clearinghouse—Center for Labor Research and Education, the University of California at Berkeley, or the U.S. Office of Personnel Management Partnership and Labor-Management Relations.

REFERENCES

1. Bureau of Labor Statistics. *Work Stoppages Involving 1000 Workers or More*; 1947–2000; <http://www.bls.gov/news.release/wkstp.t01.htm> (accessed August 2001).
2. United States Office of Personnel Management. *Labor-Management Partnership: A Report to the President*; December 2000; <http://www.gov/lmr/report> (accessed August 2001).
3. National Labor Management Association On-Line Manual for Area and Industry Labor-Management Committees, <http://www.nlma.org/infkit/histor1.htm> (accessed August 2001).
4. NLMA Present Forms of Labor-Management Committees, <http://www.nlma.org/infkit/presen1.htm> (accessed August 2001).
5. U.S. Code 29 Section 175a.
6. General Accounting Office. *Federal Labor Relations: A Program in Need of Reform*; 1991.
7. *Executive Order 13203, Revocation of Executive Order and Presidential Memorandum Concerning Labor-Management Partnerships*; February 17, 2001.
8. Hyde, A.C. The death and life of the labor-management partnership. *Public Manag.* **2001**, 30 (1), 57–59.
9. Nesterczuk, G.; Devine, D.; Moffit, R. *Taking Charge of Federal Personnel. No. 1404*; Domestic Policy Studies Department, The Heritage Foundation: Washington, DC, 2001.
10. Cohen, S.R. *Memorandum for Heads of Departments and Agencies, Guidance for Implementing Executive Order 13203*; U.S. Office of Personnel Management: Washington, DC, 2001.
11. Philadelphia Area Labor Management Committee, 2000; <http://www.palmnet.org> (accessed August 2000).
12. NLMA. *Ten Pointers for Creating an Effective Community-Wide Labor-Management Committee*; available at; www.nlma.org (accessed August 2000).

Labor–Management Rights and Responsibilities

D. S. Chauhan

*Department of Political Science, Bowling Green State University,
Bowling Green, Ohio, U.S.A.*

INTRODUCTION

The discussion on the issue of management rights in collective bargaining has undergone a significant change during the past 40 to 75 years, both in the public and private sector, respectively. The change in our understanding of the application of the concept in the public sector is influenced by various factors involving scope of bargaining, civil service rules, merit-based personnel policies, organizational democracy, participative management, and the changing nature of the work environment. The emerging social, economic, political, and technological changes have also contributed to our thinking about the management rights, and managerial control and authority.

THE CONCEPT OF MANAGERIAL AUTHORITY

Conceptually, the literature on labor–management relations continues to use various terms to describe managerial authority. These terms include “management rights,” “managerial authority,” “managerial control,” “management prerogatives,” and “managerial incentives.” At the base of all these concepts is the thinking that the hierarchical structure of the organization gives the managers the privilege to direct, control, and supervise the activities of the employees or workers to achieve the agency’s mission and statutory goals. However, the term “administrative or managerial responsibility” is lacking the measurement of the moral or ethical dimensions of administrative responsibility as characterized by either inadequate “action” or “inaction” in dealing with policy or personnel issues affecting labor–management relations. The traditional labor relations literature articulated two views of the management rights concept usually applied by the arbitrators and courts as “residual rights doctrine” or “implied obligations” doctrine. The residual rights doctrine is more popular with management practitioners and the “implied obligations” doctrine is more attractive to public managers “to negotiate changes in terms and conditions of

employment even in the absence of an express contract provision covering the issues involved.”^a There has been increasing popularity among behavioral scientists to support a position based on the theory of employment contract, suggesting that the management may have property right to allocate the resources needed to run the organization. However, the management of human resources requires the consent of the managed. Workers have no legal compulsion to cooperate with the management.^[1] Killingsworth made a similar argument, noting that the reserved rights doctrine does not have adequate value as a legal principle because it ignores the power relationship between the employer and employees.^[2] Similar views of management rights were articulated by behavioral theorists such as Chester I. Barnard and Herbert Simon, who stated that “the authority of the management depends on the employees acceptance of management’s right to direct employee behavior.”^[3] Whether they accept a narrow or broad degree of managerial authority, it may depend on the attitudes of workers or work groups. As Professors Keohan and Katz stated: “In practice, therefore, the work force’s acceptance and practical definition of management rights may deviate considerably from the specific rights provided in the labor agreement and from the philosophical arguments found in the management or administration literature.”^[4]

NATURE OF EMPLOYER–EMPLOYEE RELATIONSHIP IN GOVERNMENT

The nature of employer–employee relationship in government has changed over a period of time, bringing into existence a new set of realities, policies, and organizational structures. The attitude of government as a sovereign authority has also changed from

^aFor a discussion of these two views, see Frank Elkowri and Edna Asper Elkowri, *How Arbitration Works*, 3rd edition, Bureau of National Affairs, Washington, DC, 1976; 412-435.

negative control and confrontation to accommodation and continuing adjustment giving rise to the notion of bilateral decision making based on mutual respect and cooperation. The management rights in this collaborative environment are influenced by various legal, organizational, and environmental constraints. The other important factor that affects employer-employee relationship is the question "where does the management start?" It is quite different in the public sector, and the term "management" has to be used more loosely because the decision-making authorities of government (both legislative and executive) can grant or deny what the unions want. Also, the management authority is so diffused that varying amounts of authority are given to executives by city or county charters.

The discussion on "relationship clauses" in various contracts may show that management rights are legibly defined. In many cases, a union cannot control or bargain over a major decision on funding, manpower utilization, and operating standards, but can negotiate or file grievances over the details of their implementation. The boundary is uncertain and the distinction is fuzzy because the management directs the work with the employees' consent. The problem is further complicated by the political nature of the relationship.^[5] These uncertainties and blurred boundaries between the authority of management and the rights of the union (members) led Gene Huntley to observe that management rights' clauses were originally intended to keep unions out of specific management decisions entirely. As the unions grew in power and strength, management rights clauses served less as fences and more as signposts.^[6]

OBJECTIVES OF MANAGEMENT RIGHTS

The development of industrial conflict is basically economic and arises from the persistent urge of working people, and from resistance by owners and managers to many of the changes that the resolution of the conflict requires.^[7] A review of labor relations' literature in the United States and Europe reveals four major patterns of relationships. These include 1) *confrontation* characterized by attack and resistance techniques very common in early history of the labor-management relations in the United States; 2) *cooperation* characterized by working together under law to achieve common purpose (legal contract); 3) *collaboration* viewing conflict as dysfunctional for all, and reinforcing the genuine desire to work together aimed at eliminating conflict and embracing the health of the organization; and 4) *codetermination*, mostly common in Europe, requires that primary policies and operating decisions be made in a

codeterminate manner between labor and management.^[8] The decision-making process is influenced by a unity of function and the interdependence of such other factors as employee needs, expectations, perceptions, and the relative power advantage available to the parties impinging upon the decision-making process.^[9]

Why Management Rights/Prerogatives?

Almost all labor-management contracts incorporate the management rights clauses to:

1. Provide management an authority or incentive to manage.
2. Carry out the mission of the agency.
3. Promote administrative/managerial efficiency.
4. Establish policies concerning hiring, firing, demoting, and reassigning employees.
5. Discipline employees for their willful failures or negligence in following agency rules or government laws.
6. Take corrective actions to modify the behavior of employees.

Protecting Individual Rights in Collective Bargaining

The scope of the constraints imposed on management rights may differ from jurisdiction to jurisdiction and among states or federal government, depending on the comprehensive nature of their labor relations program. In the context of labor-management rights, another important issue that requires adequate attention is how to protect the individual rights in collective agreements? Normally, union represents individual interest as part of the collective agreement. However, it is the grievance procedure in the contract that provides for a protection clause. It is through the grievance procedure that the union can share the rights of management. A grievance may be defined as any complaint—legal, imagined, or contrived—on the part of the employee with regard to wages, hours, or any other terms and conditions of employment.

The Section 14(b) of the Taft-Hartley Act allows both states and union territories to enact legislation prohibiting union security arrangements. Accordingly, any individual will have the right to work in a bargaining unit that is represented by a union without joining the union or supporting the union financially.^[10]

LABOR–MANAGEMENT RIGHTS AND OBLIGATIONS

Collective bargaining is a psychological process in which each party (labor or management) tries to take advantage of the other by the use of statutory, contractual, and customary rules in fulfilling its enduring collective interests. These rules of bargaining influence the balance of power. Because of the sovereign nature of governmental authority and the enumeration of management rights in statutory law, the union as a protector of collective rights of the workers and maintaining the balance of power relies heavily on the mutually agreed-upon contractual provisions. These provisions are related to employee rights, grievances, procedures, and unfair labor practice(s). Employee unions also make use of the threat of strike or binding arbitration as potential means of maintaining balance of power. If this balance of power is lost significantly by the union, it may make the collective bargaining process less effective or redundant.

Employee Rights

Unions seek preservation of the balance of bargaining power by seeking incorporation of employee rights provisions in the labor–management relations statutes, contracts, ordinances, and agreements of understanding concluded between the labor and management in the public sector. These employees rights clauses normally stipulate that employees have rights to:

1. Form unions.
2. Join and assist unions.
3. Participate in good-faith collective bargaining.
4. Engage in lawful concerted bargaining activities or refrain from doing so.
5. Be represented by a union of their choice.
6. Maintain collective security.

There are several union security procedures available to employee unions. The management often prefers to negotiate an “open shop” arrangement, but unions normally prefer statutory language providing for a “union shop” through “dues check off” process, which allows unions to charge membership fees for representation.

Management Rights

In the public or private sector, management rights are either enumerated in the enabling

legislation^b or specified in the management rights clause in the contract. These rights are not free from mutual obligations. The careful analysis of labor–management rights reveals a balancing among management’s rights to manage, employees needs for security and fair treatment, union responsibility to the people, and the impact of these concerns on the collective bargaining process. For instance, the management rights specified in the state of Indiana, gives management authority to:

1. Manage and direct schools.
2. Direct work of employees.
3. Establish policy.
4. Hire, promote, demote, transfer, assign, and retain employees.
5. Discipline employees.
6. Maintain efficiency.
7. Relieve employees due to lack of work.

^bManagement Rights (*Source: Civil Service Reforms Act, 1978, Section 7106*):

- Subject to subsection (b) of this section, nothing in this chapter shall affect the authority of any management official of any agency—
 - to determine the mission, budget, organization, number of employees, and internal security practices of the agency; and
 - in accordance with applicable laws—
 - to hire, assign, direct, layoff, and retain employees in the agency, or to suspend, remove, reduce in grade or pay, or take other disciplinary action against such employees;
 - to assign work, to make determinations with respect to contracting out, and to determine the personnel by which agency operations shall be conducted;
 - with respect to filling positions, to make selections for appointments from—
 - ◇ among properly ranked and certified candidates for promotion; or
 - ◇ any other appropriate source; and
 - to take whatever actions may be necessary to carry out the agency mission during emergencies.
- Nothing in this section shall preclude any agency and any labor organization from negotiating—
 - at the election of the agency, on the numbers, types, and grades of employees or positions assigned to any organizational subdivision, work project, or tour of duty, or on the technology, methods, and means of performing work;
 - procedures which management officials of the agency will observe in exercising any authority under this section; or
 - appropriate arrangements for employees adversely affected by the exercise of any authority under this section by such management officials.

- Take actions necessary to carry out mission of schools.^c

Similarly, the city of Phoenix, Arizona, grants to its city manager the right to determine purpose of agency, set standards of service, exercise control and discretion, direct and discipline employees, and determine the means and personnel for carrying out operations. The state of Washington statute also carries a similar statement, specifying the scope of management rights by stating that management is required to carry out statutory goals using most appropriate means and personnel; manage employees; hire, fire, and transfer employees; establish work rules; and discipline employees.^[11] The enumeration of management rights either in the state statute or in the municipal ordinance demonstrates a remarkable similarity in specifying the nature and scope of managerial authority.

Toward Mutual Understanding

1. The history of labor-management relations demonstrates that the authoritarian or autocratic managerial leadership style cannot succeed in anchoring its goals in an environment filled with mutual distrust, disrespect, suspicion, threat, reprisals, and coercive behavior. What is needed is an understanding based on mutual respect, trust, partnership, and cooperative leadership. The current practices involving participative management style, use of labor-management committees, and the consultative leadership process based on the cooperative and reform-oriented approach advocated by the federal labor relations program under the E.O. 12871, 1993.^[12] Consequently, it is important to recognize the symbiotic functional relationship between labor-management rights.

CONSTRAINTS ON MANAGEMENT RIGHTS

Since the adoption of labor-management programs in the state of Wisconsin in 1959, the comprehensive

labor-management policies have been established in more than two-thirds of the states over a period of 45 years. During this period, unions have grown in strength, maturity, and negotiating skills. The changing work environment has also led to significant changes in organizational structures, personnel policies, working conditions, scope of bargaining, employee training and development, work standards, and the health and safety of the workers. All these structural, functional, and policy changes have had a significant impact on “management rights” or “managerial authority.” The following factors have either directly or indirectly contributed to influencing the nature and scope of management rights in the public sector:

1. *Expanding scope of bargaining:* As a matter of strategy, unions prefer a wide range rather than a management seeking to narrow the scope of bargaining. Traditionally, the scope of bargaining included wages, hours of work, and working conditions. However, the scope of bargaining today has expanded well beyond the traditional boundaries. For instance, teachers have been able to bargain successfully to participate in policy and other institutional matters.^d
2. *Grievance procedures:* The provision grievance action is common in most statutory laws, administrative appeal procedures, and negotiated contracts. A grievance procedure gives the employee union or management an opportunity to complain or grieve against each other to redress or seek compliance to any violation of stipulated conditions specified in contractual agreements. The employee unions invoke grievance procedures more often than management involving the issues related to contract interpretation, enforcement of contractual provisions, and the tenure and conditions of employment. In general, the causes of grievances may be classified into three broad categories: 1) misunderstandings; 2) intentional violations; and 3) sympathetic grievances. However, the most

^cThe State of Indiana Statute, Act 75 (1.c.20, Sec. 1 *et seq.*, 1973), states:

1. manage and direct schools;
2. direct the work of employees;
3. establish policy;
4. hire, promote, demote, assign and retain employees;
5. discipline employees;
6. maintain efficiency;
7. relieve employees due to lack of work;
8. take actions necessary to carry out the mission of schools.

^dTypically, teachers have been more successful in broadening the scope of bargaining than any other occupational groups mainly because of their joint participation in management affairs of the organization. The following demonstrates the nature and scope of working conditions negotiated by the teachers as part of their collective bargaining agreements: 1) class size provisions; 2) the number and functions of teaching aids; 3) school calendar; 4) teacher representation on instructional policy committees; 5) teacher arbitration procedures; 6) special programs; 7) curriculum and grading criteria; 8) textbook selection and the allocation of federal grant money; 9) teaching methodologies and peer evaluation; 10) staff-pupil integration and day care facilities for employees; and 11) teacher participation in school site selection and student disciplinary procedures.

frequently arbitrated issues in government pertain to:

- a. discipline and discharge
- b. arbitrability
- c. just cause
- d. management rights
- e. past practice
- f. overtime
- g. reinstatement and reimbursement from suspension
- h. leaves of absence
- i. promotion
- j. union security
- k. disciplinary procedures and practices.^[13]

The nature and scope of the grieved issues listed above are suggestive of the fact that employee unions—through written or negotiated grievance procedures—have significantly influenced or created constraints on the effectiveness of management rights clauses incorporated in the collective bargaining contracts or agreements.

3. *Personnel practices*: Personnel practices pertaining to recruitment, selection, and testing were traditionally considered as part of management prerogatives. In some cases, unions have penetrated these traditional areas of management rights through their efforts to influence the determination or modification of job qualifications and position descriptions.^[14] Similarly, unions have also recognized the importance of employee training and development for career enhancement. They vigorously support special training needs of women and minorities. Consequently, the public policy on equal employment opportunity/affirmative action has also increased the influence of unions in managing human resources in public organizations.
4. *Merit systems and civil service rules*: To resolve issues or problems associated with partisan politics adversely affecting personnel practices in the United States, the government established the merit system to select, promote, retain, and reward public employees based on merit principles. A merit system also requires civil service rules and regulations that not only strengthen the hands of the management in certain ways, but they also constraint management rights through added security to union employees, based on such considerations as seniority, due process, equity and fairness, and the reduction or layoffs. The civil service system covers merit as well as non-merit appointments. It is possible that “a civil service system may or may not

contain a merit system, but a merit system is always part of a civil service system.”^[14]

5. *Unfair labor practice(s)*: The unions have also gained strength by grieving through the unfair labor practices (ULPs). A ULP involves the violation of any right granted to employees, unions, or employers by a collective bargaining law or contract.^c More specifically, a ULP constitutes: 1) employer violations of employee rights; 2) employer violations of union rights; 3) union violations of employer rights; and 4) union violations of employee rights. Remedies to ULPs normally include warnings or orders such as cease and desist, reinstatement by a regulatory administrative agency, back pay, and postwritten notices. The fear that either

^cULPs:

1. interfere, restrain or coerce employees”—interfere with, restrain, or coerce employees in or because of the exercise of their rights as guaranteed in the statute.
2. “dominate unions”—dominating or interfering with the formation, existence, or administration of any employee organization; making financial contributions to any employee organization.
3. “discriminate on account of union membership”—discriminating in regard to hire or tenure of employment or any term or condition of employment to encourage or discourage employees in the exercise of the fights guaranteed to them by the statute.
4. “discriminate on account of testimony”—discharging or otherwise discriminating against any employee for signing or filing an affidavit, petition, or complaint, or given any information or testimony, under the statute.
5. “refuse to bargain”—refusing to negotiate in good faith with the exclusive representative of the employees in an appropriate unit (or with the employer representative), or refusing to process grievances.
6. “interfere with employer choice of representative”—interfering with, restraining, or coercing a public employer in the selection of a representative for the purposes of negotiations or the adjustment of grievances.
7. “cause employer to commit an unfair labor practice”—coercing, intimidating, or inducing any officer or agent of the employer to engage in any practice with regard to its employees that would constitute ULP if undertaken by the officer or agent own initiative; or attempting to do so.
8. “violate collective bargaining agreement”—violating the provisions of any written agreement with respect to terms and conditions of employment affecting employees, including an agreement to arbitrate or accept the terms of an arbitration award, where previously the parties have agreed to accept such awards as final and binding upon them.
9. “strike”—the concerted failure to report for duty, the willful absence from one’s position, unauthorized holidays, sickness unsubstantiated by a physician’s statement, the stoppage of work or the abstinence in while or in part from the full, faithful, and proper performance of the duties of employment, for the purpose of inducing, influencing, or coercing a change in the conditions, compensation, rights, privileges, or obligations of employment.

party can subject the other to a charge of facing a ULP becomes a constraining factor.

TOWARD ESTABLISHING MANAGERIAL RESPONSIBILITY

The labor-management process emerging from the growing emphasis on the labor-management cooperation, mutuality of their rights and obligations, and partnership in managing public employer-employee relationship seem to be gradually moving toward decline of management rights affected by the demand of the unions to participate jointly in policy areas that were traditionally considered to be within the exclusive domain of the management. For instance, the development of personnel policies in such areas as evaluation guidelines, pay determination, advancement, recruitment practices, testing for substance abuse, and the health, safety, and welfare of the employees under the general rubric of the "terms and conditions of employment" now often requires union consultation. Consequently, the management may be tempted to devise a coping strategy to avoid frequent confrontation with unions. This kind of bureaucratic behavior or managerial behavior in government may raise a fundamental question: Are the legal definitions of management rights incorporated in the negotiated contracts or statutory provisions adequate to provide effective safeguards and incentives to manage? The answer to this question in the public labor relations literature is pointing to the display of an attitude on the part of the bureaucrats either to adopt a posture of "inaction" or "circumvention" of public policy caused by the failure to act or take appropriate action. The case in point is the failure of the university bureaucrats to enforce the Ohio State Teacher Retirement System (STRS-Ohio) regulations to protect the rights of its faculty members.^f The administrative bureaucracy at the university under consideration was

particularly deceptive. Similarly, in its initial stage, the STRS also failed to take appropriate corrective measure in a timely manner.^[15]

In the declining state of management rights and the changing work environment in the public sector, what is needed is to safeguard the legally defined management rights and augment them by an ethical component so that the "public interest" is fully protected. The evolving cooperative leadership can help to a certain degree in promoting labor-management harmony, but it cannot guarantee an ethical correctness of mutual obligations. What is needed is to establish the process of managerial responsibility that is based on mutual respect, obligations, and ethical integrity and responsibility.

CONCLUSION

The evolving emphasis on labor-management cooperation and partnership is sound, but the attitude of management to keep the unions out of its domain of influence and legal authority may lead to some uneasy partnerships. Any public labor-management relations program dealing with management rights or prerogatives must focus on maintaining proper balance between the incentive to manage and protecting the health, safety, and welfare of the employees. In the twenty-first century, the evolving realities in the general socioeconomic and work environment, as well as the changing nature of employer-employee relationship, will require significant adjustments in values, attitudes, and behaviors of both labor and management. The forces of change in these areas may lead to modification or replacement of such old administrative values as "sovereign authority," "efficiency," "economy," and "productivity," by mutual respect, cooperation, integrity, responsiveness, and administrative responsibility. In the future, it is expected that the discussion of issues relating to legal and ethical responsibilities of public administrators will dominate the public policy agenda covering management rights.

The survival of management in the collective bargaining process will depend on management's ability to retain its managerial authority through management rights clauses in the collective bargaining agreements, statutory laws, ordinances, and legal opinions. The management may still retain the right to hire, fire, discipline employees, and administer agency programs, but unions will continue to exert their demand to participate in policy-making and implementation aspects in these areas. Historically, the fundamental question in labor laws has been how to arrive at a balance between the rights of workers and employers that provide for "equity, economic stability, and procedural workability."^[10] What degree of control

^fThe STRS case study deals with the coverup of initial failure of university bureaucrats to carry out their responsibilities and the reluctance of STRS-Ohio to investigate and bring criminal charges against university officials to collect a large sum due to the retirement system. The STRS rules and the state law requires that a person who takes an early retirement cannot return to teaching within 2 months without forfeiting a month's pay for each month taught prior to the 2 months' limitation. The STRS has oversight responsibility for the entire program. The 1968 case highlights the failure of the university officials, STRS bureaucrats, and other public agencies responsible to execute state laws and regulatory policies established by the STRS to protect all retirees who have contributed to the system. The author of this case study raises an important issue about the failure of the bureaucrats running the system to carry out legal obligations and demonstrate the moral and ethical responsibilities to the public contributors and retirees of the STRS-Ohio. *Source*: V. Jerone Stephens, "The Smashed Model of Bureaucratic Behavior: The Way the Process Works to Aid Bureaucratic Irresponsibility," research paper prepared for delivery at the OAEPS, 2001.

has management retained, and what rights has it surrendered since the collective bargaining came to the public sector? There is no clearcut answer to this question because the situation significantly varies in the state and local government characterized by impressionistic statements.^[13]

Collective bargaining procedures require that administrators deal formally with their bureaucracies and negotiate on hosts of issues. Union's seek to insert contract provisions related to personnel policies, including job descriptions, position classification, work schedules, examination and promotion procedures, recruitment policies, and training programs. Many decisions long regarded by public employers as part of management's prerogative have become bargainable issues.

The issues related to organizational justice, workplace violence, acquired immune deficiency syndrome, smoke-free work environment, drug abuse testing, child/elder care, alternative lifestyles (same-sex relationships), flexible staffing, flexible location (flexplace), and romantic behavior/sexual harassment in the workplace will have significant impact on management's rights and legitimate authority to establish the required policies, procedures, and regulations. Certainly, unions also want to participate and share management's authority to regulate in these areas. Undoubtedly, collective bargaining has brought a new force into the decision-making process. As a partner in a bilateral decision-making process, management must be able to defend its decisions and be prepared for new challenges.^[16]

REFERENCES

1. Chamberlain, N.W.; Kuhn, J.W. *Collective Bargaining*; McGraw-Hill: New York, 1965; 80–90.
2. Killingsworth, N.C. Management Rights Revisited. In *Arbitration and Social Change, Proceedings of the Twenty-Second Meeting of the National Academy of*

- Arbitrators*; Somers, A.G., Ed.; Bureau of National Affairs: Washington, DC, 1969; 3–13.
3. Barnard, C.I. *The Functions of the Executive*; Harvard University Press: Cambridge, MA, 1938; Simon, H.A. *Administrative Behavior*; McMillan: N.Y., 1947.
4. Kochan, T.A.; Katz, H.C. *Collective Bargaining and Industrial Relations: From Theory to Policy and Practice*; Richard D. Irwin: Homewood, IL, 1988; 358–359.
5. Chauhan, D.S. *Prepared Class Lectures*; Department of Political Science, Bowling Green State University: Bowling Green, OH 43403, 1995; 1–3.
6. Huntley, G. Diminishing Reality of Management Rights. In *Reprinted in the Public Personnel World Readings on the Professional Practice*; Shafritz, J.M., Ed.; IPMA: Washington, DC, 1977; 201–205.
7. Drucker, P.F. *Management: Tasks, Responsibilities, Practice*; Cited in Ref.^[8].
8. Levine, M.J.; Hagburg, E.C. *Public Sector Labor Relations*; West Publishing Company: New York, 1979; 1–11.
9. March, J.G. *Handbook of Organizations*; Rand-McNally & Company: Chicago, IL, 1965; 69–70.
10. Rosenbloom, D.H.; Shafritz, J.M. *Essentials of Labor Relations*; Reston Publishing Company: Reston, VI, 1985; 17, 95–120.
11. Usery, W.J. *Summary of Public Sector Labor Relations Policies*; Labor-Management Services Administration, U.S. Department of Labor: Washington, DC, 1976.
12. Nigro, L.G.; Nigro, F.A. *The New Public Personnel Administration*; F.E. Peacock Publishers: Ithaca, IL, 2000; 233 pp.
13. Coleman, C.J. *Managing Labor Relations in the Public Sector*; Jossey-Bass Publishers: San Francisco, 1990; 145–186.
14. Kearny, R.C. *Labor Relations in the Public Sector*; Marcel-Dekker Inc.: New York, 1992; 215–234.
15. Stephens, V.J. The Smashed Model of Bureaucratic Behavior: The Way the Process Works to Aid Bureaucratic Irresponsibility. *A Research Paper Prepared for Delivery at the Ohio Association of Economists and Political Scientists*, Columbus, OH; 2001.
16. Moskow, M.H.; Lovenberg, J.J.; Koziara. *Collective Bargaining in the Public Employment*; University Press of America: Washington, DC, 1970; 216.

Lao People's Democratic Republic (PDR)

Clay Wescott

*Government and Regional Cooperation Division, Asian Development Bank (ADB),
Manila, Philippines*

INTRODUCTION

The governance system in Lao People's Democratic Republic (PDR) has evolved from key historical events, in the context of a number of geographical factors such as being landlocked, mountainous, and thinly populated, with ethnically diverse people living in small, isolated villages. Although the poorest Association of Southeast Asian Nations (ASEAN) country, Lao PDR has seen modest improvements since the early 1990s due to market-oriented reforms and improved institutions.^[1]

BACKGROUND

The fourteenth-century kingdom of Lan Xang broke up into three separate kingdoms in the late eighteenth century, and, some 100 years later, France took control and combined them under the rule of the royal house of Luang Prabang. With the exception of a brief period of Japanese control in World War II, French colonial rule continued until 1954. The Lao People's Party, later the Lao People's Revolutionary Party (LPRP), linked with the Pathet Lao resistance movement, established in 1955 under the leadership of Kaysone Phomvihane. In the 1960s, Laos, under the U.S.-supported government of Prince Souvanna Phouma, was subjected to heavy bombing as the United States sought to hinder the passage of North Vietnamese soldiers and supplies along the Ho Chi Minh Trail.

The Lao PDR was proclaimed in 1975. Prince Souphanouvong was named president, but the real power lay with Kaysone Phomvihane, secretary general of the LPRP. The new government faced severe challenges in the aftermath of 20 years of political struggle, civil war, and bombing; the emigration of skilled administrators and professionals; and a Thai economic blockade from 1975 to 1976. As many as 40,000 people were sent to reeducation camps, and anywhere from 30,000 to 160,000 were imprisoned for "political crimes." The former king, Savang Vatthana, had initially been assigned the position of supreme adviser to the president, but in 1977, Savang Vatthana was banished to a remote area of the country. Prince

Souvanna Phouma was allowed to live quietly in Vientiane until his death in January 1984.

Attempts to collectivize agriculture encountered strong opposition, and production stagnated. The LPRP took the first steps toward market-oriented reforms in 1979 by easing restrictions on private trade and encouraging joint ventures between the state and the private sector. It reduced agricultural taxes and increased government procurement prices for most crops; however, it made no move to dismantle central planning.

More far-reaching reforms began in 1986 following the adoption of the New Economic Mechanism (NEM). The government abandoned the collectivization of agriculture, eased many restrictions on private sector activity, and gave state enterprises more decision-making power. By 1989, most political prisoners had been released and the camps had been closed. The number of state-owned enterprises (SOEs) was reduced from more than 800 in the early 1990s to about 90.

THE PUBLIC SECTOR

Although the 1991 constitution contains elements of an earlier revolutionary orthodoxy, it emphasizes that the economy is market-oriented, permits private property and all forms of economic ownership, and protects religious, gender, and ethnic rights. The constitution legally establishes executive, legislative, and judicial branches of government, whereas the leading role of the party is assured.

With about 65,000 members, the LPRP's highest authority is the Party Congress, a gathering of party cadres who meet to ratify decisions already taken by the party leadership. Next in the party hierarchy is the Central Committee—59 party elites who fill key political positions and lead the party between congresses, which are held every 5 years. The Central Committee is headed by the nine-member Politburo.

The LPRP has shown itself to be remarkably resilient. Transitions of power have tended to be smooth, the new generation of leaders has proven to be more open to reform, the Politburo has some ethnic diversity, and organized opposition is weak. Some improvements

have been evident since the early 1990s in literacy, primary education for girls, and reduction of child mortality, malnutrition, and poverty.

The central government consists of 16 ministries and equivalent organizations, down from 32 after the 1986 NEM reforms. The National Assembly elects the president for a 5-year term. The Council of Ministers is the highest executive body and includes the prime minister and three deputy prime ministers. There is also a vice president. None of the key political and government leaders is a woman; however, there are four women members of the 59-member Central Party Committee: three female vice ministers and one woman governor. When assessing Lao politics, it is important to distinguish between the formal political structure and the informal networks through which much of the decision making takes place. Nepotism and patronage are endemic within the party and the bureaucratic system. The lack of transparency makes it difficult for the majority of the population or outside observers to understand the political decision-making process.

Since a border conflict with Thailand in 1987, Lao PDR has been at peace with its neighbors. However, the Lao People's Army still maintains a regular armed force of 29,100 and a local militia of around 100,000. In addition, the military remains politically powerful: only two of the nine-member Politburo have no military background. Although the military is regarded as politically conservative, it has exploited the nexus between its powerful position and the opportunities provided by economic reform. At the heart of the military's commercial empire is the Import-Export Company, whose diverse activities include agriculture and forestry, construction, light industry, trade, and tourism. Yet for most soldiers, army life is austere. Pay is low, and uniforms and equipment are in short supply. Consequently, most young Laotian men serve only the minimum term after being drafted.

From 1975 until the early 1990s, Lao PDR relied heavily on Vietnamese and Soviet military aids. Once the internal security threat from resistance groups had been reduced and relations with the country's neighbors had improved by the early 1990s, the leadership was able to reduce the defense budget and, as of mid-1994, Lao PDR had the smallest defense budget in Southeast Asia.

Lao PDR is, at best, at the very beginning of democracy. The National Assembly is elected every 5 years. It currently has 109 members (25 of them women), and generally meets twice a year. There is ongoing work with donor support to promote debate, use secret ballots, televise sessions, and improve public consultation, legislative drafting, and floor deliberation processes. Between sessions, the Standing Committee of the National Assembly, consisting of the president, vice president, and an unspecified number of other members,

prepares for future sessions and oversees activities of the administrative and judicial organizations.

Lao PDR is divided into 16 provinces, 112 districts, and more than 11,000 villages, yet there has been little real devolution of power. Most members of the government are members of the LPRP, and the party-controlled Lao Front for National Reconstruction must approve those who stand for election to the National Assembly. The present National Assembly includes only one non-party member.

Relations between Lao PDR and neighboring countries are cordial, in part the result of membership in regional groupings such as the Greater Mekong Subregion (GMS) and ASEAN. Despite its economic problems, Lao PDR has continued to take steps to meet tariff reduction targets to comply with ASEAN Free Trade Area commitments.

The civil service has about 90,000 employees (38% are women),^[2] along with roughly the same number of other workers paid by the government but not part of the civil service, including the military. Although the official size and cost of government are less than for regional comparators, these figures may understate the actual size and cost due to weak information systems. Over four-fifths of civil servants work in provincial governments. Inflation has eroded the value of wages to the point that they are well below the minimum needed for food and basic necessities, despite a 20% wage increase in 2002.^[3] Operating budgets are minimal. As a result, many civil servants lack the motivation to perform their jobs effectively. Low skill levels exacerbate this problem, along with recruitment and promotion based on factors other than merit, and frequent marginalization of talented staff through transfers and assignment to low-priority tasks. Job descriptions are only now being introduced and there is little clarification of roles and responsibilities. The civil service tradition is strongly top-down, and the staff are reluctant to make decisions or bring problems to their superiors' attention. Exchange of information is limited.

In 1991, the government established the National School of Administration and Management to train civil servants. It also created the Department of Public Administration and Civil Service in 1992, which is responsible for guidelines and policy issues. The Central Committee for Organization and Personnel appoints and manages staff at the director level and higher. Although there have been ongoing efforts supported by donors to address administrative weaknesses, progress remains slow.

Lao PDR is a member of the International Monetary Fund (IMF), which in 1998 laid down an international code for public financial management. The code covers four broad requirements: clarity of roles and responsibilities; public availability of information; open budget

preparation, execution, and reporting; and independent assurances of integrity. The government has made some progress in moving toward these requirements.

The constitution requires the National Assembly to approve both the budget and the development plan, but contrary to the code, the National Assembly does not receive detailed information on the budget. There is also no indication that the National Assembly approves or reviews extrabudgetary spending, quasi-fiscal activities, or government equity holdings.

Lao PDR has made some progress toward complying with the code's requirement for public availability of information. Although increasingly detailed budget information has been released to the public since 1997, there are serious delays in publication, and no coverage of contingent liabilities, tax exemptions, and other required elements.

As concerns independent assurances of integrity, government accounts are incomplete, inconsistent, and difficult to comprehend, and there is no independent audit authority. The National Audit Office is under the Office of the Prime Minister; thus it is not in compliance with the minimum standard of independence.

On the revenue side, the government has introduced a number of fiscal reforms intended to reduce the budget deficit and inflation. A key aim has been to shift the revenue base away from a dependence on transfers from state enterprises toward income, property, excise, and trade taxes. As a result, revenue has gone up as a percentage of gross domestic product (GDP): from 10.6% in 1999 to 14.5% in 2002.^[3,4]

The IMF's code also calls for the administrative application of tax laws to be subject to procedural safeguards. Various reforms have reportedly improved the filing rate from 30% to 80%, and have increased collections significantly. Customs procedures and clearance times have also improved, although customs revenues have not, probably because of low valuations of imported goods, smuggling, and tariff reductions required by ASEAN.

Weaknesses in public administration and financial management lead to problems in service delivery. For example, the country suffers from a shortage of schools and textbooks, poorly qualified teachers, and low enrollment and completion levels. Although educational expenditures have increased since the mid-1980s to 7.4% of the budget, with significant donor-funded investments, the amount would need to increase to around 15% to meet stated goals.^[4]

Health standards are also lower than desired due, in part, to a weak public health system that is inaccessible to much of the population. Many poor people rely on private pharmacies that have sprung up in recent years for treatment, but their personnel are often inadequately trained. A positive sign is that government

health expenditure increased from 2.5% in 1992 to 5.6% in 2001/2002, although this combines a sharp increase in the donor-funded capital budget with cutbacks in the recurrent budget. Overall, the budget remains heavily skewed toward capital rather than recurrent expenditure.

The government has initiated important reforms in the area of central-local government relationships. Fiscal and planning responsibilities were centralized in the 1990s, and then selectively transferred back to local authorities by decree in 2000, allegedly to speed up project implementation.^[5]

Past and present decentralization efforts face challenges because of unclear legal and regulatory framework and enforcement, lack of capacity at the provincial level, weak central institutions, and poor communication and transport links. The lack of accountability of subnational authorities and an effective and transparent system of fiscal decentralization has reduced the transfer of tax revenues to the center. The economic gap between Vientiane and the rest of the country has widened since the reforms began, as most investments have gone to the capital and its environs.

Legal and judicial reforms began in 1989. A draft criminal code established procedures for criminal cases, set up a court system, and established a law school. The Supreme People's Court is now responsible for judicial appointments. Laws are published in the Official Gazette, but only for limited distribution each quarter.^[2] The Ministry of Justice routinely publishes its decrees and those of other ministries it deems important. This leads to confusion in the application of unpublished decrees and arbitrary enforcement.

Access to justice is hindered by shortages of qualified judges and facilities, and administrative interference in judicial decisions. Outside of major urban centers, citizens rely on village-based conflict resolution committees to adjudicate disputes. Based on traditional systems, these parajudicial elected bodies relieve the underresourced civil courts of civil complaints.^[2]

A key governance constraint is corruption at all levels. New opportunities for corruption opened up as economic reforms started to take hold in the 1980s. For example, giving provinces the right to trade directly with neighboring countries has opened the way for trade-related graft; the opening up to foreign investment has introduced opportunities to collect money to facilitate required authorizations. The enhanced political and economic roles given to the army have provided new opportunities for smuggling. Corruption also spread to personnel management, leading to the rapid promotion of those close to powerful leaders. Donor-funded procurements are commonly abused for personal gain. There are also reportedly 4–10% kickbacks paid for government contracts.^[2]

The government has taken many actions to address the problem, adopting an Anti-Corruption Decree in 1999 and new directives at the Party Congress in 2002, placing new controls on illegal logging, publicly condemning lavish consumption, and strengthening the State Audit Authority, State Inspection Authority, and Inspection Department of the Ministry of Finance.

PRIVATE SECTOR, CIVIL SOCIETY, AND THE MEDIA

The private sector consists mainly of small family trading and food-related businesses, about 75% in rural areas. Although the government still views the emerging private sector—especially indigenous entrepreneurs—with suspicion, many initiatives are underway with donor support to improve the enabling and regulatory environment.

The National Chamber of Commerce and Industry represents 500 businesses, but still depends on party guidance. The party's mass organizations have extensive networks at the village level, and some accomplishments that point to, for example, the Lao Federation of Trade Unions have promoted the rights of textile workers and HIV/AIDS prevention as a right of all workers.

There are only a few indigenous organizations that might be viewed as the beginnings of an independent civil society. Nevertheless, there are increasing numbers of Laotian non-governmental organizations (NGOs) working at the village level (e.g., Champa works on health issues and Padek Lao works on community-based agricultural issues). Foreign NGOs have developed further than indigenous ones because the government finds them easier to control.

Information and communication have been tightly controlled since French colonial times. During the years of revolutionary struggle, the LPRP relied heavily on radio broadcasts to a largely illiterate, mountain-dwelling audience. In recent years, the number of newspapers, televisions, and radio stations has increased. The government still controls all domestic radio, television, and newspaper outlets, and reacts harshly to

expressions of political dissent. Despite government controls, the quality of reporting and scope of topics covered is increasing.^[2]

Given the proximity of Thai radios and televisions, Thailand remains both an open window to a different economic system and provides an alternative perspective on the news. Satellite and cable broadcasting and the Internet are legal and progressively reducing government monopoly control of the media.

CONCLUSION

The resilient party and government apparatus that inherited the historical legacy and geographical constraints of Lao PDR has come a long way from its initial socialist orthodoxy to the nascent constitutionalism and market mechanisms of today. Yet Lao PDR is still a long way from operating as a modern constitutional state based on a clear separation of powers, and further reform will be required for Laos to meet its ambitious goals of poverty reduction, social development, and economic development.

REFERENCES

1. This is a personal view, and may not represent the views of the ADB. It summarizes and updates "Lao PDR." In *Key Governance Issues in Cambodia, Lao PDR, Thailand and Viet Nam*; Wescott, C., Ed.; ADB: Manila, 2001.
2. Lao, P.D.R. *Priority Areas for Governance Reform: Public Service Reform, People's Participation, Rule of Law and Sound Financial Management—Policy Paper*; Government of Lao P.D.R.: Vientiane, 2003; 8, 39, 40, 47, 62–64.
3. ADB *Lao PDR—Country Strategy and Program Update*; ADB: Manila, 2004; 1, 2.
4. World Bank and ADB. *Lao PDR Public Expenditure Review Country Financial Accountability Assessment*. World Bank: Washington, 2002; Vol. 2, 5, 15.
5. Chagnon. *Looking Back to See Forward*. Swedish International Development Cooperation Agency: Vientiane, 2002; 22 pp.

Law and Ethics

John A. Rohr

Center for Public Administration and Policy, Virginia Polytechnic Institute,
Blacksburg, Virginia, U.S.A.

INTRODUCTION

This article has two major sections: 1) theory and 2) practice. The second section is subdivided into: a) law as command; b) law as empowerment; and c) comparative law.

THEORY

At the most general level, questions of law and ethics are matters of jurisprudence, or the philosophy of law. Here the most fundamental issue is the perennial debate between adherents to the natural law theory, on the one hand, and legal positivists on the other. The former hold that to be valid, a law must somehow conform to a moral order in the universe, or a “higher law,” whereas the latter consider law to be a duly enacted command of the state regardless of its moral value. Martin Luther King’s famous “Letter from a Birmingham Jail,” wherein he argued that segregation laws in the southern states were not laws at all because they were unjust, provides an excellent example of a modern author calling upon a legal tradition that reaches back through the High Middle Ages to the ancient Stoics of classical antiquity. The nineteenth-century British philosopher, John Austin, is usually considered the most important legal positivist, especially because of his important contributions to analytic jurisprudence, which treats a law as simply a fact regardless of its history or morality and seeks to clarify its meaning by a variety of rigorous analytical methods.

When questions of legal philosophy are applied to practical human affairs, both the natural law adherents and the positivists encounter certain long-standing difficulties. During the second half of the twentieth century, positivists were regularly embarrassed by the criticism that they have no principled basis for acknowledging the evil of Hitler’s anti-Semitic measures because they were, after all, commands of the German state and therefore entitled to obedience by the German people. The positivist knows no higher law to stand in moral judgment on duly enacted laws of the state.

The natural law school has difficulty explaining the content of the moral order of the universe to which the laws of the state should conform. Significantly, Saint Thomas Aquinas, the most prominent advocate of the natural law position, was always quite circumspect in expressing specific applications of such broad natural law principles as *bonum faciendum*, *malum vitandum*—good is to be done and evil to be avoided.

Writing in the thirteenth century, Aquinas, for all his brilliance, did not go unchallenged. The great medieval debate on the relation between law and morality is nicely summarized in the query, *Estne prohibitum quia malum aut malum quia prohibitum*? That is, is a particular form of behavior prohibited because it is evil or is it evil because it is prohibited? The medieval debate extended to strictly religious questions when learned theologians puzzled over the jurisprudential implications of certain aspects of the Decalogue. When the Lord commands that the Sabbath be kept holy, is He commanding in the same way as when He bids the faithful not to kill. Isn’t killing *inherently* evil regardless of whether the Lord has explicitly forbidden it? Is it not prohibited because it is evil—*prohibitum quia malum*? On the other hand, there seems to be nothing inherently sacred about the seventh day. Surely the Lord could have commanded that the fifth or the tenth day be kept holy as easily as the seventh. To violate the Sabbath is not evil in itself; it is evil because it is prohibited—*malum quia prohibitum*.

The current debate over abortion provides an interesting contemporary glimpse of the natural law tradition in action. Pro-choice advocates argue that extremely restrictive abortion laws violate the higher law that empowers a woman to use her body as she sees fit, especially in matters as intimate and personal as childbearing. Pro-life groups argue that permissive abortion laws violate the higher law that protects the life of innocent human beings, in this case the developing fetus.

The abortion debate is particularly illuminating because it spans the divide between two related but very different moral traditions—natural law and natural rights. As noted above, the natural-law tradition affirms a certain moral order in the universe that human beings ignore at their peril. Its origins are both

ancient and medieval. The natural rights tradition is modern, tracing its origin to the “state of nature” postulated by Hobbes and Locke, wherein all persons enjoy certain natural rights without any reference to an overarching universal moral order.

Not surprisingly, the abortion debate pays little heed to such fine academic distinctions. Pro-choice advocates and their pro-life adversaries are practical men and women, far more interested in seeing their cause prevail than in the pedigree of their arguments. Nevertheless, in making these very arguments, they unwittingly speak the lines that were written centuries ago. They cannot help doing so because the effort to define the relationship between law and morality is profoundly and abidingly human.

PRACTICE

Law as Command

Ordinarily, one thinks of laws as issuing commands—stop at the red light, pay your taxes, register for the draft, etc. In the domain of public service ethics, the most important laws commanding certain actions focus on avoiding conflicts of interest and filing accurate financial disclosure statements. Statutes, administrative regulations, and municipal ordinances addressing such matters can be found at all levels of government, but the most fully developed system is found at the federal level where a specific agency, the Office of Government Ethics (OGE), has been created to enforce laws and regulations pertaining to ethics in the executive branch.

The Ethics in Government Act of 1978 established OGE and tasked it with enforcing new ethical standards legislated by Congress in the aftermath of the Watergate scandal. When Jimmy Carter ran for president in 1976, Watergate was fresh in everyone’s memory. Not surprisingly, candidate Carter promised that, if elected, he would see to it that Congress passed a new ethics law. Congress honored this pledge with the Ethics in Government Act of 1978. Interestingly, however, the new legislation focused almost exclusively on avoiding or correcting financial irregularities, whereas the Watergate crimes were primarily concerned with abuses of power. It is difficult to imagine just how the Watergate events would have turned out differently if John Dean, Gordon Liddy, and John Ehrlichmann had been required to file financial disclosure statements. To the extent that law commands, one can surmise that Congress found it easier to legislate sweeping financial disclosure statements than to address the more subtle problems of abuse of power, a matter unlikely to be found amenable to statutory commands.

Many prohibitions against conflicts of interest had been in place long before 1978. For example, Congress had already prohibited government officials from participating in matters in which they had a financial interest and receiving supplements to their salaries from non-governmental sources. A striking innovation in the 1978 legislation, however, was the imposition of certain postemployment restrictions on officials about to leave government service. The duration and severity of these restrictions varied with the proximity of the connection between the work the official had done in government and the type of work he or she planned to take up in the private sector. These strict regulations were sanctioned with criminal penalties.

A major theme in the conflict of interest area is the importance of avoiding appearances of impropriety. Financial disclosure regulations are intended to detect and deter such appearances. The concern with appearance is intended to bolster the confidence of American citizens in the integrity of those who govern them. This is brought about by imposing a “Caesar’s wife” standard upon public servants.

Law as Empowerment

Not all laws command; some of them empower certain citizens to take certain actions on behalf of the state. For example, as far back as 1906, Congress empowered officials of the Bureau of Chemistry, the forerunner of today’s Food and Drug Administration, to regulate “substandard food” and “impure drugs.” Such vague directives conferred considerable discretionary authority upon administrators to govern substantial segments of American life. The same can be said for the congressional authorization to the Federal Trade Commission in 1914 to prevent “unfair competition.” Similarly, congressional enactments empowering the Federal Communications Commission to grant radio and television licenses as the “public interest, convenience, or necessity” will indicate and the Securities and Exchange Commission to maintain a “fair and orderly market” are steps down the same road. That road has led to a mature administrative state that is characterized by the salience of policy making by civil servants—both politically appointed and career professionals—in significant areas of American life. These administrative powers are grounded in statutes that do not command but rather empower administrators to share in governing the republic.

This empowerment of the bureaucracy creates an ethical problem for those so empowered. Conventional democratic theory requires that those who govern should be chosen by the people, but civil servants are not elected and, indeed, most of them are career personnel whose positions are secure regardless of

electoral results. The ethical problem for the conscientious civil servant is to devise a principled basis for exercising his or her discretionary power in a manner consistent with the exigencies of a constitutional republic.

Over a century ago, civil service reformers tried to avoid the political and ethical implications of a merit system of personnel selection by drawing a sharp distinction—indeed a dichotomy—between politics and administration, assigning politics to elected officials who make policy and administration to civil servants who simply do the bidding of their political masters. This effort to remove administrators from politics could not withstand serious analysis, and by the middle of the twentieth century, it was abundantly clear to all and sundry that statutory law empowers non-elected civil servants to play a significant role in governing the United States. This is an ethical problem of the highest order because non-elected officials share in governing a democratic society. For the past quarter century, scholars and practitioners of public administration have looked for suitable standards to guide civil servants in the exercise of their governing power without notable success.

Comparative Law

American administrators are not alone in exercising powers of governance without being subject to the discipline of the ballot box. The exercise of such powers is a hallmark of the modern administrative state. In the United States, one of the principal checks against administrators who use their discretionary powers unethically comes from the remarkably strong American judiciary, at both the state and federal levels. Not only can American judges nullify administrative actions that go beyond an agency's statutory powers, but they can also declare unconstitutional the very statutes that empowered administrators to act in the first place.

The ordinary courts in Great Britain and France have no such power, but these nations are by no means without institutional resources to control overly zealous administrators.

In the United Kingdom, the courts curb administrative excesses by relying on the principle of *ultra vires*, that is, the principle forbidding ministries and other executive agencies from taking actions beyond those authorized by the act of Parliament or the Order in Council grounding the particular administrative action in question. Ordinarily, British courts invoke this doctrine by measuring a challenged administrative action against specific statutory language. British courts have gone further, however, by presuming that Parliament never intends to authorize abuses, and, therefore, they

have held that safeguards against abuse must be *implied* in every statute they construe. In other words, British courts compensate for their lack of power to declare an act of Parliament unconstitutional by reading into statutes a principle safeguarding against abuse by administrators who use their statutory powers in an unethical or irresponsible manner. No statute is declared unconstitutional but the abusive administrative action is found to violate the implied meaning of the statute itself.

British courts achieve a similar result by invoking the closely related principle of “natural justice,” which they maintain is also implied in acts of Parliament. Its most important function is to require administrators to grant a fair hearing before imposing an adverse administrative action upon a British subject regardless of whether the statute authorizing the action required such a hearing. This principle has been applied in such routine matters as removing a company's name from a list of approved contractors or denying a petition for discharge from a mental institution.

When administrative abuses in such matters occur in the United States, the legal challenge ordinarily focuses on the Fifth and Fourteenth Amendments, as well as the specific statute authorizing the action in question. Since British courts cannot void a statute on constitutional grounds, they curb irresponsible bureaucratic behavior by reading into the statute or Order in Council legal principles that American judges would consider constitutional guarantees of due process of law.

The ordinary courts in France, like their British counterparts, have no power to declare statutes passed by the French Parliament unconstitutional. The French, however, follow a different path toward curtailing administrative abuse. Prior to the present French Constitution of 1958, there was no way that any act of Parliament could be declared unconstitutional regardless of how flagrantly unconstitutional it might be. The Constitution of the Fifth Republic changed all this by creating the Constitutional Council, which was empowered to void acts of Parliament. This council, however, is not a court; no litigants appear before it. Its jurisdiction is invoked only by officials displeased with an act of Parliament they consider unconstitutional. The jurisdiction of the Constitutional Council addresses parliamentary rather than administrative abuses.

Because of the extraordinary power of the French bureaucracy, a special jurisdiction developed within the Council of State—originally an advisory body to Napoleon—to adjudicate claims by French citizens that they have been mistreated by administrators. When councilors of state sit as an administrative court, they exercise a judicial power totally independent of the ordinary French courts where routine

civil and criminal cases are heard. There is no appeal to a higher court from a decision of the Council of State. Councilors of state are themselves civil servants, which might suggest a certain bias on their part against aggrieved citizens. Such, however, is not the case. For well over a century, the Council of State has developed a solid body of legal principles that protect French citizens against abusive and unethical behavior on the part of French administrators. Today the Council of State's reputation for fairness has made it one of the most prestigious institutions in France, thereby providing empirical evidence to support the proposition that one administrative institution is quite

capable of encouraging ethical behavior in other such institutions.

BIBLIOGRAPHY

1. Baumgarth, W.P., Regan, R.J., Eds.; *Saint Thomas Aquinas on Law, Morality and Politics*; Hackett Publishing Col.: Indianapolis, 1988.
2. Fuller, L.L. *The Morality of Law*, 2nd Ed.; Yale University Press: New Haven, 1964.
3. Hart, H.L.A. *The Concept of Law*; Clarendon Press: Oxford, 1961.

Leadership

Thomas W. Kent

*Department of Management and Entrepreneurship, College of Charleston,
Charleston, South Carolina, U.S.A.*

INTRODUCTION

There have been a number of studies conducted in the United States on behavioral factors associated with transformational leadership. These studies have attempted to identify the behaviors that comprise what Burns described as “transforming leadership.”^[1] Burns believed that the process of leadership changed both leader and follower. It affected, positively, the morals, ethics, and attitudes of both leader and follower. These personal changes occurring within individuals then instigate changes in organizations, according to Burns. This conceptualization of leading was, at the time, a far cry from the popular scholarly thought that leadership was made up of two types of activities. One type brought about the structure an organization or team needed to succeed (initiating structure); the second type exercised concern for the populace of the organization or team (consideration). Since Burns’ identification of this new concept of leadership, the question became what comprises this transforming leadership? What does a transforming leader look like in action? What does the transforming leader do to transform organizations and teams? This, then, led to the attempts to identify the behavioral components of transforming leadership.

The earliest attempts to identify these factors were made by Bass and his associates.^[2–5] Bass went on to develop a questionnaire (called the Multifactor Leadership Questionnaire or MLQ) that purported to measure and define transformational leadership (notice that Bass used the word transformational and not transformation or transforming leadership as Burns called it). In his studies, Bass found four factors related to transformational leadership. They were idealized influence—related to trust and faith in the leader, individualized consideration—related to the personal attention the leader gives an individual, intellectual stimulation—related to engendering creativity and innovation within followers, and inspirational leadership—related to creating vision. Later, Yammarino and Bass, in a study of military officers, changed idealized influence to charisma.^[6] Kouzes and Posner factor analyzed statements based on leaders’ experiences of their performance.^[7,8] They found five factors: challenging the process—related to experimenting and

taking risks; inspiring a shared vision—related to creating a vision and enlisting support of it; enabling others to act—related to fostering collaboration among others; modeling the way—related to setting an example; and encouraging the heart—related to recognizing performance of others.

Similarly, Kent, Crotts, and Aziz using SPSS, factor analyzed behavioral descriptions of leaders’ performance and developed four factors.^[9] A reanalysis of the data using LISREL confirmed the original four factors plus a fifth factor.^[10] These factors were visualizing greatness—which is related to creating an inspiring vision, empowering the “we”—which is related to strengthening, and identifying with, a unit (such as a team or an organization), communicating for meaning—which describes deep-level transactions between follower and leader, managing one’s self—pertaining to the leader’s skill at self-direction and control, and care and recognition—which is related to giving attention to followers’ effort and success. These ideas have been widely studied among North American managers and are fairly well accepted among both the scholarly and the professional communities. But do the ideas apply globally to all or most cultures? For example, the leadership function of “Empowering the ‘We’” is very much team oriented or focuses on the collective body. That notion may apply to team-oriented societies, but is it employed by leaders in more individualistic cultures? This chapter will explore that question by reviewing the findings of studies that sought to answer the question.

DEFINITION OF TRANSFORMATIONAL LEADERSHIP

Kent attempted to differentiate between the concepts of leadership and of management explaining that, at the time, there seemed to be confusion among both practitioners and scholars between the two concepts.^[11] Kent’s definition of transformational leadership included descriptions of the purpose, products, and processes of leading.

The purpose of leading according to him is to create direction and the unified will (among followers) to

pursue it through the development of peoples' thinking and valuing.

Leading's products include: 1) the establishment of thrust or purposeful energy toward a purpose, goal, or end; 2) the creation of social orderliness—as opposed to disarray or group chaos—to carry out that thrust; and 3) higher states of behavior in terms of principles, values, morality, and ethics—among both leader and followers.

The processes (or methods) of leadership—these would be categories or groupings of leadership behaviors—are: 1) empowering the “we”; 2) visualizing greatness; 3) managing ones' Self; 4) communicating for meaning; and 5) caring and recognition. These are the same five factors discovered by Kent, Crotts, and Aziz and Kent.^[9,10]

The question for this chapter is: Does this definition apply outside the boundaries of the United States and Canada? Berry and also Douglas and Nijssen warn against imposing culturally specific values, concepts, and language into the research process, by taking ideas and research tools developed in one culture and using them in the study of another culture, without insuring the reliability and validity of the concepts and tools within the second culture.^[12,13] Since most of the above research dedicated to discovering the behaviors that leaders employ was conducted on U.S. managers, this warning should apply when applying the concepts of transformational leadership to other cultures.

Berry proposes a number of steps to reduce the possibility of or the impact of a researcher imposing his or her concepts, instruments, or behaviors onto the subjects from another culture via his or her research techniques and tools—what he calls an “imposed etc.” He describes the importance of demonstrating *equivalence*—as far as the research goes—between the cultures under study. According to Berry, it is important to establish functional, conceptual, and metric equivalence. Functional equivalence is related to insuring the behaviors in question—the behaviors being studied serve the same purpose in the cultures in question. Conceptual equivalence involves taking steps to assure that the concepts and research tools possess similar meaning among the cultures involved in the study. Metric equivalence applies to the results of the study. Statistical and procedural steps should be taken to ensure that both the instruments and the results provide a valid basis for comparison between or among cultures.

Douglas and Nijssen describe the equivalence required for cross-cultural research as scalar, construct, and measurement unit equivalences. There are similarities between Berry's and Douglas and Nijssen's two sets of three equivalences. Scalar equivalence implies that steps be taken to insure that the responses to an instrument within a study have the same meaning

and interpretation. Construct equivalence is assured when concepts occur in, and have the same meaning across, cultures. Measurement unit equivalence requires that the units of measurement be translated to common units such as when comparing two economies and the local currencies are translated to a common currency.

This chapter will review a few of the studies that have attempted to measure the equivalence of transformational leadership concepts across cultures and national boundaries.

INTERNATIONAL STUDIES

Perhaps, the largest and most globally comprehensive study of the subject in recent years has been The Global Leadership and Organizational Behavior Effectiveness (GLOBE) Research Program. This program tested the hypothesis that transformational leadership was basically similar in 62 cultures.^[14] This massive study, which involved 170 investigators, attempted to investigate the relationships between societal cultures, organizational cultures, and transformational leadership. While they found many culturally specific leadership behaviors (behaviors that facilitated effective leadership in one culture but not in others), their primary finding was that behaviors normally associated with transformational behavior were strongly and universally endorsed across cultures.

In the GLOBE study, middle managers from each of the cultures were asked if specific leadership behaviors contributed to or inhibited effective leadership. The researchers' main hypothesis was that charismatic/transformational behaviors would be universally endorsed as contributing to effective leadership. This hypothesis was soundly supported by the study, indicating that transformational leadership behaviors are culturally endorsed in all cultures included in the study.

It is suggested that the GLOBE study established functional or scalar equivalence regarding the overall idea of transformational leadership among several international cultures, including the cultures involved in this study.

The GLOBE study did neither study actual behaviors of leaders nor the relationship of actual transformational leadership behaviors to organizational performance. Quesada, Gonsales, and Kent^[15] measured actual leader behaviors in Costa Rica and in Spain vs. the behaviors of leaders in the United States using instrument developed by Kent.^[10] The Quesada et al. study satisfied the requirements of both scalar and construct equivalence and measured actual behaviors as reported by subordinates (followers) of the leaders. This study found support for four of the five

leadership processes described in Kent's definition of transformational leadership.^[11] Care and recognition was the one factor that did not "translate" to both Spanish-speaking countries included in the study. Interestingly, that is the one factor that was not found in the original factor analysis by Kent, Crotts, and Aziz.^[9]

CONCLUSION

There is sufficient research on the international equivalence of transformational leadership to suggest that the idea is not bounded by cultural or national boundaries. The GLOBE study confirmed the international transference of the values associated with transformational leadership, and the Quesada, Gonsales, and Kent study supports the idea of an internationally equivalent set of behaviors that are common across countries. These studies measured leadership values and behaviors in a great variety of types of organizations—industrial, governmental, social, etc. Hence, it is indicated that there is a common set of transformational leadership behaviors that apply interculturally as well as in many different types of organizations. It is beginning to look as if leadership is leadership, no matter what type of organization you lead or where it is sited (in what country or culture).

Clearly, far more research is needed to replicate the Quesada, Gonsales, and Kent study in many more countries before we conclude that there is a single set of leadership behaviors that apply to all cultures. Additionally, we need to assess whether these behaviors are equally important to the performance of organizations across cultures.

REFERENCES

1. Burns, J.M. *Leadership*; Harper & Row: New York, 1978.
2. Bass, B.M. *Leadership and Performance Beyond Expectations*; Free Press: New York, 1985.
3. Bass, B.M.; Yammarino, F.J. *Multifactor Officer Questionnaire: MLQ Forms 11r and 11s*; Center for

Leadership Studies: State University of New York at Binghamton: Binghamton, NY, 1987.

4. Bass, B.M. Evolving perspectives of charismatic leadership. In *Charismatic Leadership: The Elusive Factor in Organizational Effectiveness*; Conger, J.A., Kanungo, R.N., Eds.; Jossey-Bass: San Francisco, 1988.
5. Bass, B.M.; Avolio, B.J.; Eds. *Improving Organizational Effectiveness Through Transformational Leadership*; Sage Publications: Thousand Oaks, CA, 1994.
6. Yammarino, F.J.; Bass, B.M. Long-term forecasting of transformational leadership and its effects among naval officers: Some preliminary findings. In *Measures of Leadership*; Clark, K.E., Clark, M.B., Eds.; Leadership Library of America: West Orange, NJ, 1990; 223–238.
7. Kouzes, J.M.; Posner, B.Z. *The Leadership Challenge: How to Get Extraordinary Things Done in Organizations*; Jossey-Bass, Inc.: San Francisco, 1987.
8. Kouzes, J.M.; Posner, B.Z. *Leadership Practices Inventory*; Pfeiffer & Co.: San Diego, 1993.
9. Kent, T.W.; Crotts, J.C.; Aziz, A. Four factors of transformational Leadership behavior. *Leadership Org. Dev. J.* **2001**, 22 (5/6), 221–229.
10. Kent, T.W. The five ingredients of leadership. *Proceedings of the 4th Annual Hawaii International Conference on Leadership*, 2004.
11. Kent, T.W. Leadership. In *Encyclopedia of Public Administration and Public Policy*; Rabin, J., Ed.; Marcel Dekker Publishers: New York, 2003.
12. Berry, J.W. Imposed ethics, emics, and derived etids: their conceptual and operational status in cross-cultural psychology. In *Emics and Etids: The Insider/outsider Debate*; Thomas, N. Headland, Kenneth, L. Pike, Marvin, Harris, Eds.; Sage Publishers: Newburg Park, 1990.
13. Douglas, S.P.; Nijssen, E.J. On the use of "borrowed" scales in cross-national research: a cautionary note. *International Marketing Review* **2003**, 20 (6), 621–642.
14. Den Hartog, D.N.; House, R.J.; Hanges, P.J.; Ruiz-Quintanilla, S.A.; Dorfman, P.W. Culture specific and cross-culturally generalizable implicit leadership theories: are attributes of charismatic/transformational leadership universally endorsed? *Leadership Q.* **1999**, 10 (2), 219–257.
15. Quesada, G.; Gonsales, M.; Kent, T.W. Leadership in Spanish speaking cultures. *Proceedings of the Annual Decision Sciences Conference*, San Antonio, TX, 2006.

Leadership in Organizations

Douglas Ihrke

*Department of Political Science, University of Wisconsin–Milwaukee,
Milwaukee, Wisconsin, U.S.A.*

INTRODUCTION

Leadership remains one of the least understood phenomena within the field of public administration, mainly because it has not been a focal point of public administration scholars until recently. Instead, scholars, for the most part, have approached the leadership of public organizations only indirectly under larger topics such as organization theory and behavior and public personnel administration. Thus, most theories within the field of public administration remain underdeveloped, and the potential exists for more scholarly research. Organization theory has contributed the most to what we know about leadership within the field of public administration.

One reason why the subject of leadership has not been addressed is because of the potential challenge it poses to the seminal theory within the field, which placed the public administrator in a subordinate rather than a leadership position. Early studies in public administration dealt primarily with finding a place for the public administrator under a new constitutional regime, where legislators were the primary actors in government asked to provide leadership. In fact, the guiding principle of public administration, that is the separation of policy development from policy implementation, places these actors into leader (legislator), as policy developer, and subordinate (public administrator), as policy implementor, roles. This guiding principle became known as the politics/administration dichotomy.

THE POLITICS/ADMINISTRATION DICHOTOMY AND BUREAUCRATIC LEADERSHIP

The politics/administration dichotomy was first articulated most notably by Woodrow Wilson^[1] and Frank Goodnow^[2] around the turn of the twentieth century, when like-minded progressive reformers came together with an interest in ridding government of corruption and wresting away power from political parties. Leadership in governmental organizations during the years leading up to the reforms implemented by progressives was largely based upon patronage, the quintessential form of what James McGregor Burns^[3] dubbed transactional leadership, a form of leadership based upon an exchange between leaders and followers.

The politics/administration dichotomy was the intellectual idea used by reformers to bring about change in government. The theory was non-threatening to those in power in that it left political decision making to politicians, while also elevated the public administrator to the level of formal actor in the process of governing. The theory was also protected from controversy by the cover of science. At the time, science, as an epistemology, was becoming popular in the United States and around the world due to its potential to advance human knowledge and raise the standard of living here and abroad. Science, the reformers argued, could be used as a means for discovering principles to manage organizations and thus make them more efficient and effective. Who could argue with such logic? Thus, reformers were a politically astute group of individuals interested in transforming government. Reformers were well aware of what they said and how it would be interpreted.

The politics/administration dichotomy as a theory of leadership was also popular during the early years of the field of public administration in part because of its emphasis on the use of business principles to manage government. In fact, the articulation of the theory in 1887 coincided with a period of rapid industrialization in the United States, and the predominant ethos in the United States at the time was social Darwinism. Thus, business principles emphasizing the bottom line and the quest to get ahead fit squarely with the promise of public administration to help advance the interest of Americans.

Leadership during the early years of the field was approached with caution. The prototypical leader was the technical manager, whose real value was ostensibly found in knowledge of scientifically derived business principles applied to organizations. Real leadership, that is handling tough value questions, was left up to elected officials under the politics/administration dichotomy. The technician would manage day-to-day governmental affairs in a rapidly changing society such that the politician could be left alone to deal with ethical and political questions that society brought about.

Scholarly works published in the field shortly after the articulation of the politics/administration dichotomy focused mainly on the structural characteristics of organizations. Authors such as Frederick Taylor,^[4] Chester Barnard,^[5] and Luther Gullick and Leonard

Urwick^[6] were associated with what the field now classifies as the Classical Period in public administration. Leadership, the Classicists maintained, consisted of figuring out the most efficient and effective form of organization. Thus, these scholars explored such questions as how to scientifically control the work of employees, how to use hierarchy to maintain authority and control information flow within the organization, and how many subordinates a manager could effectively supervise.

There remained, however, a thorny problem with regard to governance and leadership under the politics/administration dichotomy and the writings of the Classicists. The problem had to do with the fact that these theories, as implemented in practice, did not hold up to reality. That is, reality revealed that administrators were in fact engaging in value decisions, a violation of the pure form of the politics/administration dichotomy. This situation posed a challenge to democracy, because public administrators protected by civil service laws, implemented as part of the 1883 Act, could not be removed from office for political reasons. Thus, the public administrators were engaging in political decision making without being held accountable for those decisions via elections like their legislative counterparts.

LEADERSHIP AND DECISION MAKING

Herbert Simon, one of the most prominent scholars in the field of public administration helped to reveal this flaw in his writings, most notably in *Administrative Behavior*.^[7] Simon not only challenged the theoretical foundations of the field in terms of portraying the administrator as a fact-finding instrument of politicians, but also the basis for assuming full rationality in applying business techniques to the development of a science of administration. Through his efforts, Simon helped to change the focus of leadership studies in public administration from managing the structure of organizations to better understanding human decision making in organizations and emphasizing the needs and behaviors of the individual as they related to organizational needs.

Simon's new leader was faced with the challenge of trying to guide imperfect decision makers in a rapidly changing world. Simon argued that human beings are incapable of ever being fully rational, a desirable and assumed state under the politics/administration dichotomy, and are only capable of working within what he termed "bounded rationality." Decision makers, in Simon's view, were bounded in their decisions by their human limitations. No person could ever have all the facts, much less, the necessary time to accumulate all those facts, thus leaving the decision

maker in a position of "satisficing." To satisfice is to make the most satisfactory decision that will suffice in meeting the needs of the organization. The leader under this view was then left to put the decision maker in the best possible position in order to do the job. Among the alternative tools available to the leader to do this were hierarchy, as a means for controlling information flow and authority, and standard operating procedures (SOPs), used to provide insurance to the organization for consistency in decision making by subordinates.

Simon's work did much to challenge what had been written up to the time about leading public organizations. Simon did more than just criticize, however, as his work also spawned a new area of concern for the leader of organizations. Ideas about human relations became the new focus for many in the field of public administration due to Simon's emphasis on the role of the individual in organizations. In the late 1930s and well into the 1940s, Human Relationists called for leaders to consider individual needs and desires when attempting to manage public organizations. Organizations that neglect individuals could never fully tap the human resource that was potentially at the disposal of leaders. Human potential, the Human Relationists argued, is a resource more powerful than organizational structure and could be used effectively to maximize performance in organizations.

THE LEADER AND THE NEEDS OF THE INDIVIDUAL

A number of scholars are considered part of the Human Relations movement. Some of the more prominent include Abraham Maslow^[8] and Douglas McGregor.^[9] Maslow argued that all humans have certain basic needs (i.e., physiological, safety, social, self-esteem, and self-actualization), and these needs are arrayed in a hierarchy. Once a lower-level need is satisfied, it no longer serves as a motivator for the individual. For example, once a physiological need such as that for food and water is satisfied, the individual is no longer motivated to pursue it, and can thus move on to the next highest level of need, that is the need for safety, etc.

McGregor^[9] picked up on Maslow's need theory and developed a new leadership paradigm that challenged the basic assumptions of the management paradigm of the Classicists. McGregor argued that the way the Classicists viewed employees led to a particular leadership style. That is, by viewing employees as lazy and wanting to avoid responsibility, the leader was forced to engage in strict supervision, characterized by directing, coercing, and controlling employees through such means as hierarchy and standard

operating procedures. McGregor dubbed this type of leadership Theory X leadership, which he argued could only satisfy the lower-level needs of the individual. However, an alternative view of human nature was available to leaders, according to McGregor. That view thought of humans as being creative, motivated, and desiring responsibility. Based on this alternative view, leaders would be asked to engage in Theory Y practices. The crux of Theory Y management was the principle of integration, which asked the leader to design organizational structures and SOPs such that the needs of the individual could be fully integrated with the needs of the organization and thus achieve maximum performance. Theory Y would allow the individual to satisfy higher-order needs, those needs that released the potential of the individual.

In summary, early theories of leadership in public administration emphasized organizational structure and SOPs. Emphasis on these aspects of organizational life was the result of a particular view of human nature (Theory X). The work of Herbert Simon challenged these basic ideas and changed the focus of leadership research to that concerning the individual working within bounded rationality and that individual's needs and desires. This new focus was part of Theory Y management.

LEADERSHIP AND THE ENVIRONMENT

During the early period of public administration thought, two less dominant discussions within the field were taking place. One discussion explored the traits of individuals that make for good leaders. Early trait theories go back to the writings of the Greeks and Romans, and this discussion remained a part of public administration thought. The other more prominent discussion within the field dealt with the differences between closed and open systems of organizations. Closed organizational systems are characterized by repetitive tasks with consistent conversion of inputs to outputs and relatively minor environmental uncertainty. Open organization systems thinking views organizations as organic systems, akin to biological organisms, where environmental threats are a regular part of life, and organizations are asked to adapt to them or die.

The early literature on leadership in public administration essentially viewed public organizations as closed systems. By viewing public organizations as closed systems, scholars could limit their concerns to controlling organizational life through such means as hierarchy, rules, and rationality. This view also allowed for a clean separation of politics and administration as well as facts and values. Under closed systems thinking, administrative leaders were asked to deal

with factual issues, while political leaders dealt with values. The closed system perspective fit squarely with the politics/administration dichotomy.

The open systems view challenged the closed view that organizations are rational. Practice in public administration had revealed that administrators are not merely cogs in a machine and that they engaged in value decisions. Administrators were forced to make value decisions, because the environment they worked in was never fully closed. Instead, as Norton Long argued, administrative life was filled with politics, and the astute politician was one who understood that the "lifeblood of administration is power." The influence of the environment around public organizations was a constant threat, and the leader needed to play the game of politics in order to build and maintain the organization.

Thus, public administration remained a developing discipline with no clear articulation of what it took to successfully lead public organizations. Disagreement, more than anything else, seemed to characterize relations among scholars on the subject of leadership. The specific study of leadership had yet to be undertaken, at least not until after the debate over the closed and open aspects of organizational life had concluded.

James D. Thompson,^[10] in his seminal work, *Organizations in Action*, did much to end this debate. From roughly the 1930s through the 1960s, the closed system approach had come under attack, because it failed to meet the test of reality for most public organizations. Thompson saved the closed system view of organizations by placing it within his larger model of organizations. Thompson also put the open systems view in its place by collapsing it into the same model. He argued that both the closed and open systems views were correct, but only so far as within a certain domain in organizations.

Under Thompson's model, organizations consist of three basic levels, each with a particular set of responsibilities and challenges. The executive level sits atop the organization, in the middle, the managerial level, and at the bottom, the technical core.

Leaders are asked to do different things depending upon which level they work within in organizations. At the executive level, leaders are asked to deal with the internal and external environments of the organization, with primary emphasis on the external environment. The external environment includes the market the organization works within, its competitors, as well as such things as fluctuations in the workforce. Individuals at this level are asked to lead the organization in terms of the mission and vision the organization takes on as well as to strategically plan for the future. Leaders at this level work within an open systems environment, and the ability to deal with uncertainty

is at a premium in terms of the skills required of successful leaders.

At the lowest level of the organization is the technical core. At this level, rationality is at a premium due to the fact that it is here that the goods and services most identified with the organization are produced. Consistency in day-to-day operations is also at a premium, and employee tasks are structured so that consistency is maintained. Leaders at this level are asked to ensure that the right products are produced at the right time by virtue of getting the goods necessary to the technical core to produce those products. Environmental uncertainty is unwanted at this level, as it disrupts the operation of the technical. Thus, those in the technical core have a desire to work within a closed system.

In the middle of the organization rests the managerial level. Managers are asked to take what can be a nebulous vision and mission of the executive level and transform them into a set of achievable goals and objectives for the technical core. Managers are asked to buffer the technical core from the uncertainties of environmental fluctuations. Managers are also asked to implement changes desired from above in such a way as to minimize disruption in the day-to-day affairs of the technical core.

Thompson's work thus reestablished closed system theories to a level of respectability, while at the same time, provided a model of reality with regard to how organizations operate. Yet, Thompson's work does not provide much insight into what separates good leaders from bad leaders. Scholars in other fields such as business and psychology had developed theories of leadership, but they were largely left unnoticed until the 1970s.

CONTEMPORARY THEORIES OF LEADERSHIP

James McGregor Burns published one of the most popular books dealing with leadership in 1978. In his book, *Leadership*, Burns, among other things, distinguished between transactional and transformational leadership. Transactional leadership was based upon an exchange between the leader and the follower, where the leader would offer certain resources (e.g., authority, wages, etc.) in return for effort on the part of employees. Transformational leadership was something different in the mind of Burns. Transformational leaders went beyond transactions to fundamentally change organizations, such that they no longer resembled what they once were. Later, Bernard Bass^[11] would place these two types of leadership on a continuum with transactional leadership on one end, and transformational leadership on the other.

Works by Burns and Bass did not specifically address leadership in public organizations. Instead, these authors dealt primarily with what it would take for organizations to meet the challenges on the horizon, and those challenges were many.

The 1980s and 1990s brought a great deal of attention to leadership. Challenges from foreign competitors to American industry, such as auto manufacturers, pointed to the need for better leadership. Books such as *In Search of Excellence*, by Peters and Waterman, among others, provided some hope for the future of American organizations. An emphasis on quality was popularized by W. Edwards Deming^[12] during this period, as was an effort to rid organizations of inefficient processes, popularized by Hammer and Champy.^[13] Leaders would be asked to rid their organizations of outdated ways of producing products and services and implement new means for meeting customer needs.

Public organizations were facing challenges during this period as well. Developments such as tax revolts by citizens in the 1970s, along with efforts to privatize governmental services in the 1980s and 1990s, forced scholars to examine leadership in public organizations. Public organizations were faced with the dilemma of reducing taxes, increasing the quality and quantity of services, while at the same time, competing with private vendors.

The reinventing government initiative was one result of these changes. Osborne and Gaebler^[14] suggested that, among other things, that governmental organizations and their leaders need to "steer rather than row." At the federal level, reinventing government reached its zenith with President Bill Clinton's National Performance Review (NPR), a nationwide effort to reduce government of waste and inefficiencies. At the local level, leaders were asked to consider ways to reduce waste and inefficiency as well, through such means as privatization and implementing quality and process reengineering initiatives.

To date, relatively few public administration scholars have explored leadership in public organizations, with a few notable exceptions. These exceptions include, for example, works by Larry Terry^[15] and Robert Denhardt.^[16]

Terry develops a model of leadership known as the administrator as conservator. The role of the leader under this model is to protect the institutional integrity of the organization. The leader does this through conserving the mission, values, and support of the organization. The administrative conservator is guided by constitutional principles and is therefore not a threat to democracy. Terry equates conservatorship to that of statesmanship, which requires professional expertise, political skill, and a sophisticated understanding of what it means to be an active participant in governance.

Table 1 A sample of theories of leadership

Leadership theory	Authors	Elements
Behavioral	Yukl, 1994	The job of the leader is to build and maintain effective relationships with people whose cooperation and assistance is needed to accomplish task objectives. Examples of leadership behaviors include supporting, developing, recognizing, rewarding, team building and conflict management, and networking.
Charismatic	Bass, 1985; Conger and Kanungo, 1987; Trice and Beyer, 1993	Charisma is the result of follower perceptions and attributions influenced by actual leader qualities and behavior, by the context of the leadership situation, and by the individual and collective needs of followers.
Contingency (two examples)		These are theories that explain leadership effectiveness in terms of situational moderator variables.
1. Path-goal	1. Evans, 1970; House, 1971	1. Leader's motivational function consists of increasing personal payoffs to subordinates for work-goal attainment and making the path to these payoffs easier to travel by clarifying it, reducing roadblocks and pitfalls, and increasing the 'opportunities for personal satisfaction en route.
2. Least preferred coworker	2. Fiedler, 1964, 1967	2. Describes how the situation moderates the relationship between leader traits and effectiveness. Favorableness of the situation is influenced by leader-member relations, position power, and task structure.
Cultural	Trice and Beyer, 1991	This provides an emphasis on cultural innovation, where the leader either makes drastic changes in the existing culture or starts a new organization with a different culture.
Participative	Yukl, 1994	This involves power and behavior approaches to leadership. It includes aspects of power such as power sharing, empowerment, and reciprocal influence processes, and it includes aspects of leadership behavior such as the specific procedures used to consult with people to get their ideas and suggestions, and the specific behaviors used to delegate authority.
Power	Podsakoff and Schriesheim, 1985	Theory suggests there are different types of power (i.e., expert, referent, legitimate, reward, and coercive) at the disposal of leaders. Leaders use these types of power to guide their organizations.
Traits and skills	Stogdill, 1948	Certain traits and skills differentiate between leaders and non-leaders.
Transactional (one example)	Bass, 1985	Leaders exchange rewards for compliance on the part of followers.
1. Leader-member exchange	1. Dansereau, Graen and Haga, 1975	1. Leaders develop different exchange relationships over time with various subordinates. Basic premise of the theory is that leaders usually establish a special relationship with a small number of trusted subordinates (the "in group"). The exchange relationship established with the remaining subordinates (the "out group") is substantially different.
Transformational	Burns, 1978; Bass, 1985	The extent to which a leader is transformational is measured primarily in terms of the leader's effect on followers. Followers of a transformational leader feel trust, admiration, loyalty, and respect toward the leader, and they are motivated to do more than they are originally expected to do. The leader transforms and motivates followers by: 1) making them more aware of the importance of task outcomes; 2) inducing them to transcend their own self-interest for the sake of the organization or team; and 3) activating their higher-order needs.

Robert Denhardt,^[16] in his book, *The Pursuit of Significance*, explores stories of success in governmental leadership. Denhardt examines the ways in which leaders of public organizations have met the needs of customers, implemented quality, and reduced waste and inefficiency. Denhardt's leaders do this through a commitment to values, serving the public, empowering employees through shared leadership, using pragmatic incrementalism, and being dedicated to public service. Denhardt argued that these practices lead to the pursuit of significance on the part of leaders. These leaders have substituted the pursuit of significance for the pursuit of private interest.

CONCLUSION

Scholars in other fields have produced theories of leadership, but these theories have been nearly ignored by scholars in public administration. These theories and their basis, and the authors most identified with them, are listed in Table 1.

As the field of public administration evolves, more theories will be developed on what it takes to lead governmental organizations. The face of leadership in public organizations continues to change with the times, while at the same time, it is constrained by constitutional principles that place checks on what can and cannot be done.

REFERENCES

1. Wilson, W. The study of administration. *Polit. Sci. Q.* **1887**, *II* (1), 197–222.
2. Goodnow, F.J. *Politics and Administration*; Macmillan: New York, 1900.
3. Burns, J.M. *Leadership*; Harper & Row: New York, 1978.
4. Taylor, F.W. *The Principles of Scientific Management*; Harper: New York, 1911.
5. Barnard, C. *The Functions of the Executive*; Harvard University Press: Cambridge, MA, 1938.
6. Gullick, L., Urwick, L., Eds.; *Papers on the Science of Administration*; Institute of Public Administration: New York, 1937.
7. Simon, H. *Administrative Behavior*; Free Press: New York, 1945.
8. Maslow, A. *Motivation and Personality*; Harper & Brothers: New York, 1954.

9. McGregor, D. The human side of enterprise. In *The Management Review*; McGraw-Hill: New York, 1957.
10. Thompson, J.D. *Organizations in Action*; McGraw-Hill: New York, 1967.
11. Bass, B. *Handbook of Leadership: A Survey of Theory and Research*; Free Press: New York, 1990.
12. Deming, W.E. *Out of Crisis*; MIT Center for Advanced Engineering Study: Cambridge, MA, 1986.
13. Hammer, M.; Champy, J. *Reengineering the Corporation*; Harperbusiness: New York, 1993.
14. Osborne, D.; Gaebler, T. *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*; Penguin: New York, 1992.
15. Terry, L.D. *Leadership of Public Bureaucracies*; Sage Publications: Thousand Oaks, CA, 1995.
16. Denhardt, R.B. *The Pursuit of Significance*; Harcourt Brace Publishers: Fort Worth, TX, 1993.

BIBLIOGRAPHY

1. Bass, B.M. *Leadership and Performance Beyond Expectations*; Free Press: New York, 1985.
2. Conger, J.A.; Kanungo, R. Toward a behavioral theory of charismatic leadership in organizational settings. *Acad. Manage. Rev.* **1987**, *12*, 637–647.
3. Dansereau, F., Jr.; Graen, G.; Haga, W.J. A vertical dyad linkage approach to leadership within formal organizations: A longitudinal investigation of the role making process. *Organ. Behav. Hum. Perform.* **1975**, *13*, 46–78.
4. Evans, M.G. The effects of supervisory behavior on the path-goal relationship. *Organ. Behav. Hum. Perform.* **1970**, *5*, 277–298.
5. Fiedler, F.E. A contingency model of leadership effectiveness. In *Advances in Experimental Social Psychology*; Berkowitz, L., Ed.; Academic Press: New York, 1964.
6. Fiedler, F.E. *A Theory of Leadership Effectiveness*; McGraw-Hill: New York, 1967.
7. House, R.J. A path-goal theory of leader effectiveness. *Adm. Sci. Q.* **1971**, *16*, 321–339.
8. Peters, T.J.; Waterman, R.H. *In Search of Excellence*; Harper & Row: New York, 1982.
9. Podsakoff, P.M.; Schriesheim, C.A. Field studies of french and raven's bases of power: Critique, reanalysis, and suggestions for future research. *Psychol. Bull.* **1985**, *97*, 387–411.
10. Stogdill, R.M. Personal factors associated with leadership: A survey of the literature. *J. Psychol.* **1948**, *25*, 35–71.
11. Trice, H.M.; Beyer, J.M. Cultural leadership in organizations. *Organ. Sci.* **1991**, *2*, 149–169.
12. Trice, H.M.; Beyer, J.M. *The Cultures of Work Organizations*; Prentice Hall: Englewood Cliffs, NJ, 1993.
13. Yukl, G. *Leadership in Organizations*, 3rd Ed. Ed.; Prentice Hall: Englewood Cliffs, NJ, 1994.

Leadership: High-Integrity

Mark L. McConkie

*School of Public Affairs, University of Colorado at Colorado Springs,
Colorado Springs, Colorado, U.S.A.*

R. Wayne Boss

*Leeds School of Business, University of Colorado at Boulder,
Boulder, Colorado, U.S.A.*

INTRODUCTION

Within the wide range of differing views of what constitutes leadership, one view that has gained increasing momentum is the notion that integrity is at the heart of what makes leaders effective because it is the foundation upon which high-trust relationships are built. Therefore, it is the basis for effective interpersonal, managerial, and organizational interaction. In short, people more willingly follow those whom they trust.

REQUISITE CONDITIONS FOR INTEGRITY TO EXIST

Three conditions are required for integrity to exist: 1) the existence of moral law, or “fixed moral principles,” subordination to which produces consistent consequences; 2) knowledge of right and wrong, which is generally conferred by conscience; and 3) individuals must possess agency and choice, thus enabling them to become, by choice, agents of moral law or fixed principles.

Moral Law

We readily recognize the existence and impact of the laws of nature in the physical world. The laws of gravity, heat conduction, centrifugal force, thermodynamics, and mathematics, for instance, govern the machinery of everyday life. To influence others by the exertions of integrity assumes that moral laws operate in their sphere just as actively as do the laws of physics or biology in the physical world. Although often more difficult to define and describe, they are just as real. Thus, we say, for instance, in illustrating moral laws, that “we do not trust dishonest people,” that “when we violate conscience, we feel inner conflict,” or that “we reap as we sow.”

When people act in harmony with moral law, they produce predictable responses and consequences. This, in turn, means that when leaders act in harmony

with moral law, followers recognize such behavior as honorable and are therefore willing to follow it. In other words, by subordination to moral law, leaders learn and exhibit integrity, which inspires followership in others. People may choose to obey or disobey the law, but they cannot choose the consequences that will inevitably follow. The principles upon which the law operates are fixed, or unchanging; they transcend time, place, and cultural context. They are objective and operate external to the perception, bias, opinion, or viewpoint of those upon whom they have an impact. This does not mean that perception, bias, opinion, or viewpoint does not affect how we react to the laws, only that the laws represent “fixed principles.”

Knowledge of Right and Wrong

For people to demonstrate integrity, they must choose right over wrong. This necessitates that they know right from wrong and this implies the existence of opposites. All things, of course, are known by their opposite—we know hot because we know cold, sweet because we know sour, light because we know darkness, and good because we know evil. Similarly, we know integrity because we know self-betrayal, its opposite. The knowledge of right and wrong is lodged in individual conscience, or what Socrates called the “inner voice,” which is a unique human endowment. When people respond to and live in harmony with the promptings of conscience, they demonstrate integrity. However, when they violate or disregard the promptings of conscience, they self-betray. To lead with integrity, leaders must be aware of both right and wrong and must choose right.

Agency and Choice

Moral agency implies subordination to moral law, and choice is the process of selecting between opposites or alternatives. When fixed principles are in operation and one chooses to subordinate their behavior to them,

one can expect fixed consequences to follow. By subordinating self to moral law, one becomes an agent of that law, much the same as a representative is an agent of the person giving the original commission. For instance, when leaders tell the truth, they can expect to be treated and perceived as truth tellers or honest leaders. In the ancient world, Cato the Younger did this so well that his name became a synonym for integrity. Thus, the ancient Roman dictum in trials sought for multiple witnesses, and “not even a Cato” should be a lone witness, particularly where a story was so strange that it would be hard to believe “even though a Cato told it.”^[1]

Leading with integrity is often easier to illustrate than to define. In organizational contexts, it is often described in organizational folklore, organizational stories, anecdotes, parables, myths, and even jokes and homily. When Plutarch wrote *The Lives*, for instance, his motive was to illustrate virtue so that others might imitate it. “Admirable actions,” he wrote, “can produce, even in the minds of those who only read them, an eagerness which may lead them to imitate them The bare account of noble deeds can make us admire and long to follow the doers of them. Moral good is a practical stimulus; it is no sooner seen than it inspires an impulse to follow it.”^[1] By providing moral exemplars, Plutarch commended to his readers those same behaviors that had made his moral exemplars influential. However, he also showcased the negative effects that unfolded in the lives of those who lacked the virtues of which he wrote, thus teaching that in the absence of virtue, vice reigned.

INTEGRITY AND LEADERSHIP STYLE

Some leadership styles lend themselves more readily to the exercise of integrity than others. A participative style is more amenable to high integrity than a coercive one. Hitler, for example, wrote that coercion and brutal force were legitimate means of changing social behavior.^[2] He often concluded that he alone could make many of the decisions in his Reich. Because of his repressive and brutal actions toward those with whom he disagreed, many in his chain of command, fearful of offending him, refused to make many of the ordinary or mundane decisions associated with their command. Rather than thinking for themselves, they often sought to ingratiate themselves to Hitler, rather than doing what they believed was right. In short, they were untrue to themselves.^[3] In contrast, the more democratic style of interaction held by Lincoln allowed him to define objectives with his commanding generals and leave to them the task of finding ways to fulfill those objectives. General Ulysses S. Grant, for instance, said to his staff, “the President

is one of the few who has not attempted to extract from me a knowledge of my movements, although he is the only one who has right to know them.”^[4] On another occasion, Lincoln sat late at night signing military commissions and said to an aid, “I do not, as you see, pretend to read over these documents. I see that [Secretary of War, Edwin] Stanton has signed them, so I conclude they are all right.”^[4]

Although many descriptors might be used to describe different leadership styles, all styles tend to fit along a continuum that stretches from the high external control or authoritarian styles at one end to the low external and high internal control, democratic, or participative styles at the other end. The theory of high-integrity leadership assumes that high-integrity styles will gravitate toward the high internal or self-control end of this continuum and that there are many behavioral correlates to high integrity that are typically identified as virtues. These virtues usually include such traits as treating people fairly, being just, kind, civil, truthful, respectful or loving, consistent and reliable, expert, and dynamic. These traits inspire the trust of others and, therefore, correlate with the components of a highly trustworthy individual, meaning that high-integrity leadership is synonymous with leadership trustworthiness.^[5] High-integrity leadership further assumes that the leader’s focus is first on building his or her own character and then working through that character to exert an influence on others. This means that such intangible variables as credibility, reputation, perceived trustworthiness, and honor are central to the leadership equation. High-integrity leaders not only know it, but also use virtue to influence others for good. They can, therefore, be influential by appealing to the moral sentiment and conscience in others.

Lincoln, illustratively, was so well regarded for his integrity that he was given the sobriquet of “Honest Abe.” Aware of the power of being regarded as honest, Lincoln used his reputation to accomplish what he believed were noble ends. For example, deep in the Civil War, Lincoln’s wife stood accused of treasonously leaking information to the enemy. The Senate Committee on the Conduct of the War met to consider the charges. A secret meeting of the committee was set. “We had just been called to order by the Chairman,” said one committee member, “when the officer stationed at the committee room door came in with a half-frightened expression on his face.” The reason for the officer’s discomfort became immediately apparent, when, unannounced, committee members saw Lincoln, standing at the front of their committee room table. No one seemed to know what to say, and an awkward silence cloaked the room. Then Mr. Lincoln spoke: “I, Abraham Lincoln, President of the United States, appear of my own volition before this Committee of the Senate to say that I, of my own knowledge,

know that it is untrue that any of my family hold treasonable communication with the enemy.” Having so spoken, he left as silently as he had come. “We sat for some moments speechless. Then by tacit agreement, no word being spoken, the committee dropped all consideration of the rumors that the wife of the President was betraying the Union. We were all so greatly affected that the Committee adjourned for the day.”^[4]

Lincoln’s reputation as a man of integrity had saved himself, his family, and perhaps the nation from a significant anguish. In so doing, he demonstrated not only the power of a good reputation, but also one important dimension of high-integrity leadership—that living to have a claim on the conscience of others gives one power. This “reputational power,” or the power of a good reputation, is something resident within the character of an individual and is the outgrowth of having lived worthy of the confidence of others. It is, therefore, perceptual, meaning that high-integrity leaders must cultivate character predispositions and traits that enable them to be the same in public as in private. In this, the notion of high-integrity leadership differs from theories that say that leaders should reflect the values of followers (which arguably is a form of followership and not leadership), or from situational and contingency theories that argue that leaders are either the products of differing situations or simply take advantage of circumstances, or even from transactional theories that say leadership is “a bargain,” “a deal,” or a “transaction” that leaders make with followers. As Ghandi said, “man cannot do right in one department of life whilst he is occupied in doing wrong in any other department. Life is one indivisible whole.”^[6]

HIGH-INTEGRITY LEADERSHIP AT FOUR LEVELS

All organizations exist at four levels, and high-integrity leadership is manifest at each level. These levels include the personal level, the interpersonal level, the managerial level, and the organizational level. At each level, a governing principle dictates the exercise of integrity in all human interaction.

Personal Level

At the personal level, integrity is a matter of the individual acting in concert with conscience, and the governing principle is trustworthiness. Simply stated, one must be worthy of the trust of others before he or she will be trusted. Thus, influence exertions begin by leaders modeling behaviors that generate trust; they are first at one with self, and others see that oneness or consistency in the leaders’ behavior. Trustworthiness

grows out of two concepts: character and competence. Trustworthy leaders, thus, model trustworthiness (character) and have both the interpersonal and the technical skills (competence) necessary to lead in the particular realm in which they operate.

Interpersonal Level

Trust results in the perception on the part of an individual that a leader is worthy of confidence. It is the fundamental glue and lubricant of relationships; it holds relationships together and facilitates interaction within them. Trust is the governing principle at the interpersonal level: Where trust exists, people work more harmoniously and more productively, and production-related variables such as communication, decision making, coordination, cooperation, etc., are positively affected. Where trust is absent, suspicion, cynicism, doubt, and related negative emotions and perceptions result.^[7]

Managerial Level

By definition, leaders and managers accomplish things through other people. Thus, the governing principle at the managerial level is empowerment—the process of giving employees responsibility and resources, defining to them goals and objectives, giving them the freedom to define how they accomplish those goals, and then holding them accountable for results. One of the most powerful means for accomplishing such empowerment, or of facilitating trust at the managerial level, is the personal management interview (PMI)^[8] or what Covey has called “the win-win agreement.”^[9] The PMI or win-win agreement meeting is a meeting between two parties, generally a superior and a subordinate, in which past accomplishments are reviewed; goals, objectives, and/or desired results specified; resources agreed upon and strategies delineated; accountabilities specified; and consequences for accomplishment or failure determined. When such meetings occur on a regular basis, both trust and production levels are high; in their absence, both trust and production levels fall.^[8] Because PMIs or win-win agreement meetings hold both parties to the agreement accountable to one another and to the agreement itself, it gives leaders an opportunity to build and nurture trust at managerial levels.

Organizational Level

Trust exists at the organizational level, when an alignment exists between the strategy, structure, and systems of the organization and the vision, mission, values, and goals of the organization. The greater the congruity between these two grouping of variables,

the greater the level of trust or the more impactful the high-integrity leadership; the greater the gap between these two sets of variables, the lower the trust, and the less impactful will be the perceived integrity manifest by the leadership.

Strategy consists in the work processes people follow to get the job done; it is the “how to” or road map, followed by which accomplishments occur. Structure deals with reporting relationships—“who reports to whom.” Systems are those support networks that facilitate work flow by organizing labor, resources, and materials. When the ideas and values around which we organize—the mission, vision, values, and goals of the organization—are part of strategies we pursue, then alignment exists. By thus integrating the individual and the organization and its subparts, organizations demonstrate a high degree of alignment, which is the organizational manifestation of high integrity. Because organizational levels of interaction require more effort and coordination than activity at personal levels, and because it is often more difficult for people to think strategically, alignment is much more difficult to achieve than are the manifestations of trust at personal, interpersonal, and managerial levels. Leadership at this level, however, is the great essential to organizational effectiveness (for a detailed description of integrity leadership, which Covey calls “Principle Centered Leadership,” and from which much of the material in this article is distilled, see Ref.^[9]. For additional insight, see Ref.^[10].)

STRENGTHS AND WEAKNESSES OF HIGH-INTEGRITY LEADERSHIP AS AN APPROACH TO LEADERSHIP

As an approach to the study and understanding of leadership behavior, the focus on high-integrity has both strengths and weaknesses. Prominently, its weaknesses include the fact that it is based very deeply in the perceptions of those involved. Because perceptions among people differ so dramatically, particularly on issues relating to values and character, confusions can occur as to the meanings of perceived behaviors. At the same time, a given behavior can be subject

to multiple perceived interpretations, resulting in additional confusion. What one person might consider to be great and noble, another might view as weakness. Alexander the Great and George S. Patton, to illustrate, were seen by many as attack commanders who were decisive and determined; others saw in those same behaviors impetuosity and impulsiveness.

The perceptual nature of discerning high-integrity leadership, however, is also one of its strengths. People intuitively resonate with the idea that high-integrity people are to be trusted, followed, and encouraged by their followership. Yet, when the values of the followers align with those of the leader, it is likely that the leader will be perceived as possessing high integrity, which adds to his or her power base and thus the ability to get things done.

REFERENCES

1. Plutarch, Cato the Younger. In *Plutarch: Selected Lives and Essays*; Loomis, L.R., Ed.; Walter J. Black, Inc.: Roslyn, NY, 1951; Vol. xvii, No. 19, 264.
2. Hitler, A. *Mein Kampf*; Houghlin Mifflin: Boston, 1943; 169 pp.
3. Speer, A. *Inside the Third Reich: Memoirs*; The MacMillan Co.: New York, 1970; 61, 78, 96–97, 301, 363.
4. Sandburg, C. *Abraham Lincoln: The Prairie Years—The War Years*; Harcourt Brace: Jovanovich, New York, 1954; 385, 597, 681.
5. Giffin, K. The contribution of studies of source credibility to a theory of interpersonal trust in the communication process. *Psychol. Bull.* **1967**, *68*, 104–120.
6. Easwaran, E. *Ghandi the Man*; Glide Publications: San Francisco, 1972; 31 pp.
7. Golembiewski, R.T.; McConkie, M.L. The centrality of trust in group processes. In *Group Processes*; Cooper, C.L., Ed.; John Wiley & Sons: New York, 1975; 130–185.
8. Boss, R.W. Team building and the problem of regression: The personal management interview as an intervention. *J. Appl. Behav. Sci.* **1983**, *91* (1), 67–83.
9. Covey, S. *Principle-Centered Leadership*; Simon & Schuster: New York, 1990; 190–201.
10. Lee, B.N. *The Power Principle*; Simon & Schuster: New York, 1997.

Learning Organizations

Montgomery Van Wart

Department of Public Administration, California State University-San Bernardino,
San Bernardino, California, U.S.A.

INTRODUCTION

Learning organizations are characterized by a high number of both members and groups who adapt to job challenges by rapid learning, and by a culture that nurtures experimentation, innovation, and thoughtful change. Typically, learning affecting organizational practices is diffused throughout the organization (not just among executives and experts), and fundamental dialogs about best or innovative practices include both middle managers and line workers. Also, in contrast to traditional organizational structures that emphasize hierarchy and control, organizational structures tend to be flatter, more flexible, and more fast changing.

The concept was first broadly popularized by Senge in his extensively read book, *The Fifth Discipline*,^[1] which fully articulated the conditions and elements of such organizations. However, the idea and the term were in use much earlier. Argyris, in particular, had been working in this area for many decades, first regarding the full incorporation of lower-level individuals into corporate life^[2,3] and later regarding types of learning in organizations.^[4-7] More recent work focused on breaking down the defensive routines of individuals that impede individual and group learning. Others who were particularly important in the development of the concept were those in the human relations movement—Schön, who worked extensively with Argyris, and Schein, whose work focused on healthy organizational cultures and the role of leadership in achieving them.^[8] Learning organizations are an example of what French and Bell call “second-generation organization development (OD).” The first generation of OD spanned from the late 1940s to the mid-1980s, focusing on behavioral and management science interventions for organizational improvement by theorists such as Lewin, Argyris, McGregor, Blake and Mouton, Beckhard, Likert, Bennis, and Golembiewski. The range of organizational improvement strategies in OD certainly included organization-wide approaches. Although often largely unacknowledged in the rash of new management approaches that occurred in the 1990s, the contribution of first-generation ODs was substantial—see for example, Watkins and Golembiewski.^[9] The second generation—largely occurring in the 1990s—included total quality management (and a related version in the

public sector, reinvention), re-engineering, team-based organizations, learning organizations, and other reform strategies.

Senge’s overall definition of the learning organization is “where people continually expand their capacity to create the results they truly desire, where new and expansive patterns of thinking are nurtured, where collective aspiration is set free, and where people are continually learning how to learn together.”^[1] Garvin, however, has a much more popular overall definition: “an organization skilled in creating, acquiring, and transferring knowledge, and at modifying its behavior to reflect new knowledge and insights.”^[10]

THE LEARNING ORGANIZATION ENVIRONMENT

Although all the major writers of the learning organization speculate about the reasons for its ascendance as an idea (even if the reality has been quite rare), it is useful to look to a prominent public sector theorist for the specifics within that particular environment. In the first major essay in the public administration field that discusses learning organizations in light of the attention that Senge brought to the concept, Kettl provided four major reasons in support of the learning organization in the public sector.^[11] First, he speculated about the end of bureaucracy as we know it (which he more aptly called the “twilight of bureaucracy”) because of the need for faster and more responsive organizations with less expensive overhead (flatter). Second, he pointed out the increased government reliance of external expertise, causing organizations to become more like networks, and also causing a whole new set of learning dilemmas surrounding contracting and partnerships. Third, he highlighted the decline in the confidence of centralized science solutions, which means that there is a loss of confidence in expert staff teams managing all organizational change. Finally, the knowledge about problems has become highly decentralized because of the speed of change and the degree of customization more commonly required. Therefore, problem solving also needs to stay more localized when possible. In summary, learning

organizations should (theoretically) thrive in environments that are dynamic and in which knowledge is highly dispersed but necessary for successful adaptation. This can be compared with the environments in which traditional bureaucracies exist—regardless of whether they are hierarchical bureaucracies more typical of the public sector or rational bureaucracies found in the private sector—in which the ideal conditions are stable and knowledge is highly centralized.

Despite the widespread agreement about the enabling conditions for learning organizations, they have not flourished as they might because of the exceptional challenges confronting them.^[12] First, by definition, the structure is less clear. Typically learning organizations are flatter, have looser formal links, and look more like networks. The links are looser in terms of rules and regulations, but nonetheless tight informal links are required in terms of cooperative synergies. Because of the need for constant reorganization to adapt to new problems, they tend to be self-organizing and “messy.”^[13] Individuals have larger, but far less defined, roles. Although these features give rise to learning organizations when they are properly functioning, when improperly implemented they lead to confusion, lower productivity, in-fighting, and lack of organizational focus, among other dysfunctions. Second, this type of amorphous structure and environment gives rise to a high degree of stress for many individuals who prefer more “organized” environments, stable conditions and standards, sharper role clarity, and high job security. Those who like more structured environments expect higher incentives (e.g., pay, working conditions, and prerequisites such as training) for staying in the organization. Further, because of the loose management structures, the more creative types of people bred by the organization are easily alienated and more likely to move to other organizations if conditions are less than ideal. Third, learning organizations require highly trained individuals. This requires background education on the discipline, organizational knowledge about systems, processes, culture, etc., and training in the specific skills necessary for the job. It also requires advanced skills to make greater use of learning by sharing, comparing, systems thinking, competing, and suspending disbelief.^[14] That is, advanced learners not only learn about basic knowledge and skills, but also become adept at learning how to learn in order to solve entirely new problems that confront the organization. However, the cost of hiring and training employees rises dramatically as better-educated individuals demand better salaries and the cost of constant training, including self-training, increases exponentially. Thus, recruiting and retention problems for learning organizations are particularly challenging.

MACROLEVEL STRATEGIES IN ACHIEVING LEARNING ORGANIZATIONS

Senge's macrolevel strategies—essentially his operational definition—is the most well known. It includes systems thinking (the fifth and integrating discipline), personal mastery, mental models, shared vision, and team learning. These are discussed more fully in the section on microstrategies.

Another well-known set of strategies is that of Garvin who also has five elements to his schema, but focuses on a more developmental or sequential approach. The first element is a systematic problem-solving approach. Learning organizations understand that problems are part of conducting business and build them into business operations with the time, resources, and support to work on them. Further, problems are systematically sorted into minor and unusual ones, or significant and chronic ones for different types of analysis and attention. Second, learning organizations experiment with different possible solutions to see what works best. Such organizations may have pilot runs, multiple formats, or time-series experiments to find the best practice for their organization. Third, learning organizations systematically review their own experience. Although this may be the result of a pilot or test, more often it is simply the result of past performance. It may range from spending a few minutes at the end of each meeting to critique the meeting itself, to the rigorous analysis of budgetary data for long-term trends, to the systematic use of performance data by frontline workers to see what mistakes were made or what practices produce the best results. Without good performance measurement systems, an organization generally cannot achieve success in this element. The fourth element is learning from others. Although this includes learning from others in the organization, it often emphasizes such practices as benchmarking, which is studying other high-performing organizations to see what they do well and learn from the comparative experience. Finally, learning organizations systematically disseminate the information learned from solving problems, experimenting, reviewing experience, and learning from others. By disseminating information about the successes of one part of an organization with others, they can recognize and celebrate those achievements as well.

MICROLEVEL STRATEGIES IN ACHIEVING LEARNING ORGANIZATIONS

Inevitably, there are a host of methods, tools, and concrete strategies to implement the macrolevel strategies. In 1994, Senge with others produced a fieldbook to further explicate ways to put into practice each of the

strategies.^[15] For example, he identifies numerous useful tools that assist in this area, such as process mapping, culture analysis, and identifying dysfunctional archetypes (e.g., “tragedy of the commons,” where everyone uses common resources but no one is compelled to replace them). He also suggests a step model for breaking through organizational gridlock: 1) identify the original problem symptom; 2) map all “quick fixes”—superficial solutions that either ignore underlying causes or simply replace old problems with new ones; 3) analyze quick fixes; that is, identify undesirable affects (including affects on others) and the addictive side effects; 4) identify fundamental solutions based on a deeper understanding of the problem and systems connections; 5) find interconnections in the targeted solutions to fundamental process loops; and 6) identify the concrete high-leverage actions to take.

At one level, personal mastery is simply proficiency and skill at what one does. As Senge uses the term, however, it also includes a personal vision of overall professional competence and achievement. Such a vision must have a highly realistic assessment of current strengths and weaknesses, as well as a notion of what can be improved or accomplished. It is acceptable to have “creative tension” between the vision embedded in personal mastery and the shared vision of an organization because the synergies of the two will enhance each other as strategic opportunities occur to achieve both, even though they are not identical. Personal mastery begins by articulating one’s goals (both professional and personal), values, and strengths/weaknesses. It is enhanced by actively seeking feedback about one’s mastery through others and through performance, as well as an active self-review and updating of one’s vision.

Because of the welter of experience that bombards each of us hourly and daily, we typically use mental models to frame this experience and to make sense of it. At a mechanical level, it assists us to filter the important information for tasks, such as driving a car, from the more judgmental need to value the dependability, intelligence, skills, cooperativeness, etc., of others based on the data we have. Mental models are often useful because we need to selectively screen sensory experiences through the use of shorthand assumptions if we are to function effectively and act efficiently. The problems with mental models are just as important, however, because they often lead to systematic biases about the information that is screened (leading to false assumptions) and because old assumptions are often not reviewed for accuracy. Those in learning organizations learn to test and improve mental models that often inhibit creativity, enhance insulation, and encourage negatively objectifying others.

Shared vision embodies the destiny of the organization, the sense of purpose or mission, the vision of

where the organization will be in the future, the values that people believe are necessary for work, and the goals or milestones used to achieve the vision. Senge asserts that the most important aspect of shared vision is that it stems from an ongoing discussion of what the organization is about, so that it can be as equally well described by line workers as senior managers. Having broad opportunity to be involved in direction setting has two implications for chief executive officers (CEOs): first, they must be willing to have an active level of involvement by others, and second, those included must be knowledgeable about the overall organization and committed to system-wide objectives. Senge points out that the levels of inclusion can range from telling (the lowest level), selling, testing, consulting, and cocreating (the ideal level).

Senge’s final discipline is team learning, because so much work cannot be done alone and individuals alone cannot solve the complex, interdependent problems that face them. Senge goes to some pains to discriminate between dialog and discussion. Discussion is when ideas are bounced off others in a “percussive” way; debate is an extreme form of discussion. It is good for quick information exchange and surfacing technical issues. However, dialog, which involves deep listening, suspension of past assumptions, and reflective thinking, is needed for the robust analysis of significant problems and to learn how to find new solutions. A specific technique to “unclog” teams is to have carefully facilitated sessions to discuss “undiscussables,” which gets negative mental models in the open so that they can be ameliorated by feedback and changed behavior. Senge also encourages the organizational skill of harnessing people’s uniqueness, and making diversity a virtue and strength.

CONCLUSION

The concept of the learning organization, regardless of the exact definition, brings together many of the ideal practices for our age of abundant information, rapid change, and greater organizational democracy. The practices of these high-performing organizations include greater levels of problem solving, experimentation, rigorous analysis, comparative learning, and systematic information dissemination. We can tell when such organizations exist because frontline workers will be as aware and excited about the shared vision as the CEO; individual competency and problem solving will be uniformly high across the organization; most important problems will be handled and solved by teams for fundamental solutions; and there is noticeable lack of stereotyping, cliques, and intergroup hostility.^[16]

Yet, we also know that such organizations are rare, despite our rather clear vision of what they should be like. Employees have to be well trained, well rewarded, committed, and empowered. Managers must be able to create communities of dialog to facilitate team learning and shared vision, facilitate the analysis and overcoming of negative mental models, and encourage personal learning—difficult talents to achieve indeed. Everyone must strive to embrace a practice of systems thinking rather than conventional, and more self-centered, specialist perspectives. In other words, the challenges are steep in trying to create the culture necessary for a learning organization, but it is certainly a worthy goal.

REFERENCES

1. Senge, P. *The Fifth Discipline*; Doubleday: New York, 1990; 3 pp.
2. Argyris, C. *Personality and Organizations*; Harper Collins: New York, 1957.
3. Argyris, C. *Integrating the Individual and the Organization*; Wiley: New York, 1964.
4. Argyris, C.; Schön, D.A. *Theory in Practice*; Jossey-Bass: San Francisco, 1974.
5. Argyris, C.; Schön, D.A. *Organizational Learning*; Addison-Wesley: Reading, MA, 1978.
6. Argyris, C. *Reasoning, Learning and Action: Individual and Organizational*; Jossey-Bass: San Francisco, 1982.
7. Argyris, C. *Knowledge for Action: A Guide to Overcoming Barriers to Organizational Change*; Jossey-Bass: San Francisco, 1993.
8. Schein, E.H. *Organizational Culture and Leadership: A Dynamic View*; Jossey-Bass: San Francisco, 1985.
9. Watkins, K.E.; Golembiewski, R.T. Rethinking organization development for the learning organization. *Int. J. Org. Anal.* **1995**, 3 (1), 86–101 (January).
10. Garvin, D.A. Building a learning organization. *Harvard Bus. Rev.* **1993**, 80, 78–91.
11. Kettl, D.F. Managing on the frontiers of knowledge: the learning organization. In *New Paradigms for Government: Issues for the Changing Public Service*; Ingraham, P.W., Romzek, B.S., Eds.; Jossey-Bass: San Francisco, 1994; 19–40.
12. Kanter, R.M.; Stein, B.A.; Jick, T.D. *The Challenge of Organizational Change*; Free Press: New York, 1993.
13. Wheatley, M.J. *Leadership and the New Science: Learning About Organizations from an Orderly Universe*; Berrett-Koehler: San Francisco, 1992.
14. Van Wart, M. Organizational investment in employee development. In *Handbook of Human Resource Management in Government*; Condrey, S.E., Ed.; Jossey-Bass: San Francisco, 1998; 276–297.
15. Senge, P. *The Fifth Discipline Fieldbook*; Currency Doubleday: New York, 1994.
16. Carnevale, D.G. *Organizational Development in the Public Sector*; Westview Press: Boulder, CO, 2003.

Lebanon: Corruption

Dimitri R. Hanna

*Department of International Affairs and Diplomacy, Notre Dame University,
Zouk Mosbeh, Lebanon*

Ramzi N. Nasser

Notre Dame University, Zouk Mosbeh, Lebanon

INTRODUCTION

Corruption benefits the individual or group at the expense of others or society at large. Corruption within public administration has a detrimental effect on the development of a viable democracy and a national economy and is a detriment to political stability, social cohesion, and economic growth. This paper conceptualizes constructs for corruption in the Lebanese situation. It analyzes the contexts for which types and ways corruption transpires and functions in public administration.

CORRUPTION AND PUBLIC ADMINISTRATION

Corruption in general can be defined as the misuse of public authority to gain private profit or to promote private interests. Classic and modern theories of public administration reduce corruptive acts by constructing internal control mechanisms and human resource incentives that deal primarily with, education, institutional design, and equal opportunity.

The application of Weberian concept applied to the field is prevalent in the public administration sector in the Middle East. It stresses external and internal structural control mechanisms, such as hierarchies, the role of the supervisor, a well-defined set of rules, standardized procedures, and strict implementations of programs. On the other hand, modern public management theories governed by human resource theory, stress on the individual sense of need, motivation, ability to make decisions, and social inculcations.^[1] Modern theories of public administration have not yet entered into the scheme of governance within public institutions in Lebanon or the Arab Middle East.

Since Woodrow Wilson's debate over the dichotomy between politics and administration, the field of public administration has become more attentive to administrative, managerial, and legal techniques that limit the exposure of individuals, institutions,

and societies to corruption. Despite the differences over value priorities between efficiency and effectiveness, the Weberian model subsumes important theoretical underpinnings on ethical conduct. Adopting the Utilitarian school of ethics, the Weberian model, stresses not only the development of internal structural control mechanisms in an organization, such as direct supervision, narrow span of control, and long hierarchies, but also external structural ones, such as a system of checks and balances consisting of anticorruption rules, procedures, judicial or legislative oversight, and organizational arrangements.^[2]

There are several causes of corruption in public administration in the Middle East and particularly in Lebanon. Cultural and social behaviors contradict many Weberian principles and the rule of law and becomes more flagrant in program implementation. With a largely incapacitated workforce in the public sector and strong interference from the political cadre, the effectiveness–efficiency dichotomy between political and public administrative functions have been blurred, leading to mismanagement, malpractice, inadequate control and corruption.

Weberian administrative systems with its various configurations and administrative political conditions dominate most Middle Eastern public sectors. When public administration organizations lack the necessary resources to run public programs along the Weberian values of specialization, efficiency, and political neutrality, the Weberian organization in its structural and social context creates an environment in which acts of corruption proliferate.^[3] Authority in tall structures become centralized and unchecked irrespective of organizational needs. In the absence of strong institutional leadership with clear objectives and standards, political and administrative conflation and prioritized relations open the means to all types of illegal activities at the individual and group levels such as drug businesses, unchecked border smuggling, “outlaw” enterprises, and money laundering by individuals and organized groups in cooperation with government officials and politicians.^[4]

Corruption within public administration can be conceptualized along two theoretical dimensions: structural (institutional) and cultural. The structural type within the Weberian model suggests that in administrative organizations, services meet the standard rules, regulations, and procedures. However, within ill-developed governmental administrative units, the context is set for bureaucrats to manipulate regulations, laws, and procedures, to gain in favors, monies, and political power. For example, officials as individuals might create delays, reject an application, ignore due process, or introduce costly fees that lead to some sort of kickback. The Weberian concept of administrative control is often “massaged” to serve particular individuals, networks, or even political groups.

The cultural and confessional mosaic of Lebanese society arranges the distribution of power in government asymmetrically among different groups and players. It is not unusual when a group uses state power to impose policies on another through public officials. Even the selective use of discretionary power in implementing programs might serve the confessional group, a narrow political agenda, or a multifarious confessional community; while, a great majority of citizens are left with no administrative recourse to correct governmental abuses. The political realism in Lebanon draws on political or confessional adversaries readily coalescing around corrupt arrangements where “supra-individualistic” behavior dominates over legality, political loyalty, and social affiliations.

STRUCTURAL DIMENSION TO CORRUPTION: THE POLITICAL STIGMA

Lebanon’s modern political system is organized around the division of power between the legislative and the executive branches represented respectively by parliament and the cabinet at large. The emerging political forces shaped along sectarian balances and formulations determine the balance of power between the different government branches. Competition for political position takes place within and across sectarian lines where prospective politicians compete first within their confessional group to occupy their allocated government seats and then between each other. The constitution draws the map of political interaction and the lines of legality among the different confessional actors in the political system. Despite what is written in the constitution, and in subsequent agreements, the power among different government branches varies with leadership. Thus, a strong leader of a new political group or coalition forming a majority in government tends to abuse power at the price of their adversaries. Such practice remains politically legitimate within government so long as it has

the consent of the majority and it is played within the democratic political system through the legislative, executive, and legal branches.

Nation states in the Middle East are either constitutional democratic republics—by statute, or constitutional monarchies or regimes. With the exception of Lebanon, Israel, and Turkey; invariably, all Middle Eastern countries have a strong and dominant executive authority, headed by a president or monarchy. Lebanon is politically structured around sectarian franchises where power is spread between the executive and the legislative branches. In the absence of political consensus and the rule of law, the promotion of equality, equity, and democratic practice in government is substantially hampered. With a public sector accommodating multiple sectarian groups, based on political alliance, Weberian neutrality and efficiency are downgraded substantially and sacrificed for the interest of a minority or group. The problem is further heightened when quota of confessional representation, in government hiring is protected and prioritized over Weberian administrative principles of merit, neutrality, or procedural regularity.

In the more adverse cases in the Middle East, such as in oil-producing countries, the Weberian administrative system is brought to a complete halt where autocratic, monarchistic, tribal, and ideological values and norms supersede national and international legal stature. Emptied from its organizational logic, the use of the Weberian model is still attractive and widely popular in the Middle East for its simple structure and concentration on authority. The attractiveness of the Weberian system by Middle Eastern regimes is found in its political form than in its productive outcome.

Leaders of political parties and multiconfessional groups provide variegated perspectives on governance and interest in public programs, leading to a divergent consensus about how public institutions should be run. The greater the cultural heterogeneity and political divergence, the harder it is for the state and public sector to accommodate the Weberian bureaucratic principles of positive neutrality, common values, merit, efficiency, and ethical standards.^[5] Overall, the Weberian model applied to a political environment is filled with contradictions, complexities, and confusions, rendering it susceptible to individual or group acts of corruption.

SOCIOCULTURAL DIMENSION TO CORRUPTION

Society at large has an inexorably strong role in shaping citizens’ sense of fairness and legitimacy, and in determining their level of tolerance and permissive attitudes toward corruption. Society in the Middle East plays an important role in the proliferation of

corruption.^[3] The patriarchal conditions in strong kinship society often replace the rule of law by an economy of affection centered on the “Patra.” These conditions distort the workings of public sector institutions in the traditional Weberian concept.

Public administration practice under confessionalism is one of the primary causes for contradictory administrative practices. For example, in Lebanon, public sector appointments are divided across confessional lines and are meant to maintain a delicate confessional balance. The efficiency value is compromised at all levels of the workplace: in hiring, development, and firing. Organizational needs and workforce specialization are always subject to compromise not in the best of government interests. For instance, the positioning of nonspecialized jobs of the first and second ranks in government are based on the strength of the connections between the patra and the individual combined with confessional quota. Public administrators pay loyalty to the patra more than public organizations or local governance. The bureaucrat, manager, or clerk, is apt to break the law, prioritize workflow, take bribes, and serve only those who placed them in the position.

In a strong patriarchal society, as in the Middle East and Lebanon in particular, identity formation and behaviors are shaped by kinship affiliation.^[6] The patriarchal belief system permeates the social, economic, and political institutions, and not only legitimizes but also elevates the role of the patra over the legal system.^[7,8] In collective spirit, individual commitments to the kinship group or confession supersede legal frameworks and concrete rules. In addition, in many areas of the Middle East, traditional behavior outpaces legal standards and rules, and the protection of tribal or confessional values displace national and international legal ones. In almost all cases, state or government plays a role in the proliferation of tribal and patriarchal interests through promotion of patron–client relations based on the exchanges of goods where the patra receives benefits in exchange for bribing government officials, whether *Baksheesh* or personal favors. Public employees may be assisted by politicians and people of influence to perform acts of corruption only to benefit in the form of monies and favors. When the issue is less politically salient, bribery can prove to be the most efficient mean for citizens to cross all types of sectional, sectarian, institutional, cultural, and organizational barriers and receive the best public service on time.

THE EFFECTS OF CORRUPTION: ECONOMIC DETERIORATION

Money made through corruption by individuals or groups is rarely invested back into the economy. It

rather distorts economic growth and development and eliminates the poor share in the greater wealth made from it.^[9] The impact on the allocation of funds from the productive sectors to consumption shifts the nation’s developmental course toward a consumer-dependent type of economy. The severity of corruption can be calculated through government collusion with private or individual interests where money and resources are diverted to nonproductive sectors like the military or security forces.^[10] For instance, most Middle Eastern countries have an exceptionally high level of military spending. A country like Lebanon whose military prowess is limited in comparison to regional powers, and experiencing a national debt of almost 40 billion dollars at the end of 2005 (seven times its yearly revenue), had a military budget in 2003 roughly equal to 10% of its budget expenditures.^[11] Despite the high spending on the Lebanese armed forces, the rise and continuation of militias supported and financed by regional forces continue to develop into political institutions independent of state control.

In essence, the huge economic damage produced by corruption in Lebanon is felt within the service sectors whether health, agriculture, energy, communications, finance, education, welfare, environment, or transportation, and is reflected in government debt. Officials may receive large payments to sign off on incomplete, ill-finished, below standards, or even non-existing projects, as in waste management or to monopolize markets such as the case of the two major cell-phone companies and their control of the cell-phone industry. In addition, a large black economy still thrives through the Lebanese–Syrian border: costing the government unrealized revenues in the form of tariffs, adding to market asymmetry, the development of monopolies, and the increase in negative externalities.

CONCLUDING REMARKS

Political and administrative reform is at the heart of political discourse in Lebanon in the wake of large economic scandals unfolding in many governmental and non-governmental institutions. The Syrian military and intelligence withdrawal from the Lebanese political and social scene and thriving political competition make reforms easily attainable with international actors showing renewed interests in the country’s public sector. A balanced use of discretion, ingrained public administrative constraints, citizen activism, and political restructuring, can possibly contribute to create a modern public administrative program with an entrepreneurial sense of public management.

In addition, demand for access to information has shaped public opinion and caused the formation of a national consensus on the anticorruption discourse.

The media, sectarian and political in its composition, has often promoted ahead of grass root movements and government, transparency through exposures, and investigative reporting.

The national consensus rallying behind Syrian withdrawal from Lebanon opened opportunities for a new era based on the clean conduct of politics, transparency, and accountability. However, external forces continue to impact the political process and governance in Lebanon. Despite the departure of Syrian military and intelligence forces from the country, their impact is still felt in many political, economic, and public security organizations. In addition, the Iranian funded Hezbollah continued to make decisions independent of the Lebanese state, which ignited in July 2006 an open Israeli war on Lebanon and the more recent, summer 2007 showdown between the Lebanese army and Al-Qaeda inspired "Fath Al-Islam" led to the destruction of many of its roads, communication, and economic infrastructures. In the wake of external forces impinging on Lebanese society, international efforts (EU and UN) have intensified to impact governance in Lebanon.

In addition to these external political factors, a dysfunctional Weberian organizational structure makes administrative reforms, whether incremental or radical, difficult to achieve in the short run. Prospective grass roots movement and civil society at large are laboring for genuine anticorruption educational programs perpetuated by a general political will and aimed at societal transformation and administrative reforms.^[3,12]

REFERENCES

1. Vigoda, E. New public management. In *Encyclopedia of Public Administration and Public Policy*, 1st Ed.;

- Rabin, J., Ed.; Marcel Dekker, Inc.: New York, 2003; Vol. 2, 812–816.
2. Denhardt, R.; Gubbs, J. *Public Administration*, 7th Ed.; Harper Collins: New York, 2003.
3. Kpundeh, S.; Hors, I. Corruption and integrity improvement initiatives in the context of developing economies. In *Corruption and Integrity Improvement Initiatives in the Context of Developing Economies*, Paris, France, Oct. 24–25, UNDP: Paris, 1997; 1–12.
4. Tamesis, P. Different perspectives of international development organizations in the fight against corruption. In *Corruption and Integrity Improvement Initiatives in the Context of Developing Economies*, Paris, France, Oct. 24–25, UNDP: Paris, 1997; 129–139.
5. Mills, A. Strengthening domestic institutions against corruption: a public ethics checklist. In *Corruption and Integrity Improvement Initiatives in the Context of Developing Economies*, Paris, France, Oct. 24–25, UNDP: Paris, 1997; 141–150.
6. Khalaf, S. *Cultural Resistance: Global and Local Encounters in the Middle East*; Saqi Books: London, 2001.
7. Sharabi, H. *Mukadmat Li-Drasat Dl-Mujtam'a Al-Arabi* (Introduction to studies of the Arab World); Dar Altali'a Liltiba'a Wa'anasher: Beirut, 1975 [In Arabic].
8. Sharabi, H. *Neopatriarchy: A Theory of Distorted Change in Arab Society*; Oxford University Press: New York, 1988.
9. Gould, D.J.; Amaro-Reyes, J. *The Effects of Corruption on Administrative Performance: Illustrations from Developing Countries*; The World Bank: Washington, DC, 1983. World Bank Working Paper No. 580.
10. Hodes, R. *Global Corruption Report: Countries Reports*; Transparency International: Berlin, Germany, 2003.
11. European Neighborhood Policy. Country Report: Lebanon. Europa Country Report. Commission of the European Communities: Brussels, Belgium, 2005.
12. Stapenhurst, F.; Langseth, P. The fighting role of public administration in corruption. *Int. J. Public Sector Manage.* **1997**, *10*, 311–330.

Legal Ethics

Christina M. Bellon

*Department of Philosophy, California State University, Sacramento,
Sacramento, California, U.S.A.*

INTRODUCTION

Legal ethics refers to the body of codes and, in some cases, statutes that define and regulate the professional practice of law. Legal ethics also has a normative meaning expressed in the branch of applied or professional ethics that studies the connection between legal practice and the principles that *should* govern such practice. These two aspects of legal ethics are not independent of one another. This is most clearly evident when existing codes of conduct are revised or modified to better express the legal profession's or society's views about what those ethical standards ought to be.

Legal ethics is about trying to ensure that the practice of law is consistent with the ends of law; that is, to ensure ethical practice in the attainment of just outcomes. Its public policy import is best expressed by David Luban, "[O]ur nation is so dependent on its lawyers that their ethical problems transform themselves into public difficulties. Put simply: the ethical problems of lawyers are social and political problems for the rest of us."^[1]

CODES OF CONDUCT FOR LAWYERS

The organized regulation of legal practice has only a recent history in the United States. Prior to 1908, when the American Bar Association (ABA) issued its 32 Canons of Professional Ethics, no formal rules of conduct were available to guide lawyers in the ethics of their practice. These canons served this function until they were replaced in 1969, when the ABA issued its Model Code of Professional Responsibility. The Model Code was fairly rapidly adopted by most states as the binding standard for ethical conduct by lawyers practicing within their jurisdiction. Shortly thereafter, in 1983, the ABA issued its most recent set of ethical rules, the Model Rules of Professional Conduct. By the mid-1990s, most states had adopted the Model Rules;^[2] however, these are only models. Any proposed code of conduct becomes binding on practicing lawyers only after a state's highest court has adopted it, usually

after considerable modification. Of course, a particular state's bar association and its highest court may opt to create a distinct code of ethical conduct for lawyers practicing within its jurisdiction, as is the case with California. Or it may create a code from a combination of those available to it, as is the case with New York's blend of the Model Code and Model Rules. Once a state has adopted a code, the authority to hear complaints against specific legal professionals is typically delegated to the ethics committee of the state bar association. In this way, the legal profession is self-regulating. Those persons who practice as professional lawyers are required to adhere to the code of ethics as adopted by the state in which they practice.

Adherence to the ethical requirements of the profession is mandatory for each practicing lawyer. When legal professionals fail to adhere to the requirements of the code or rules governing the practice of law, they may receive significant penalties ranging from private or public reprimand to suspension or disbarment. The most severe penalty, disbarment, not only deprives practitioners of their livelihood, but more important, from the point of view of a self-regulating profession, they have been barred from the legitimacy the profession conveys upon its members. In the disciplinary setting, the connection between ethics and the professional (i.e., legitimate) practice of law is most clear.

For lawyers, ethical conduct is not simply wishful thinking nor is it optional. Rather, it contributes to the very understanding of legitimate, authoritative, and professional practice. According to the Preamble to the New York Code of Professional Responsibility,

But, in the last analysis it is the desire for the respect and confidence of the members of the profession and of the society which the lawyer serves that should provide to a lawyer the incentive for the highest possible degree of ethical conduct. The possible loss of that respect and confidence is the ultimate sanction.^[3]

Adherence to the requirements of the ethical code, from the practicing lawyer's point of view, serves as a mark of professional conduct. From the client's point of view, it is a surety to instill confidence in the professional conduct of the legal advisor.

Additional Regulation of Lawyers' Conduct

What other means are available for the ethical regulation of lawyers? The ABA Model Rules is not the only document that governs the conduct of legal professionals. More specialized codes are occasionally issued by professional legal associations, some of which have been adopted by state and federal courts. Among these are specialized standards and rules governing practice in securities law, and governing the behavior of prosecutors and attorneys general. State governments may enact statutes governing legal malpractice, rules of evidence, disclosure and other trial procedures, antiracketeering, and conflicts of interest, just to name a few. The Constitution of the United States is also a basis for justifying certain prohibitions or limitations on legitimate legal practice. State and federal courts have interpreted the meaning and significance of many of the rules of the Model Code and Model Rules. Further, and perhaps most directly, individual courts can find and impose penalties for unethical behavior. For example, lawyers who act contemptuously in the courtroom may be found in violation of the state's rules of professional conduct and immediately be sanctioned or otherwise punished by the attending judge.^[4]

LEGAL ETHICS AND THE PUBLIC GOOD

To understand the public policy implications of legal ethics, one must first understand the relation between lawyers and the larger public they serve, and between the ethical requirements of legal practice and the broader ethical concerns of the public. For example, the New York Code claims no less than that the proper and just functioning of a free and democratic society depends on the right conduct of legal professionals. "The continued existence of a free and democratic society depends on recognition of the concept that justice is based upon the rule of law Lawyers, as guardians of the law, play a vital role in the preservation of society."^[3]

As public policy considerations suggest, understanding the ethical regulation of lawyers requires an appreciation of the context in which law is practiced. The institutional setting for the practice of Anglo-American law and common law in general is the adversarial system. In addition to the adversarial system, an appreciation of the function and purpose of lawyers within that institutional setting is also helpful. The combination of the adversarial system and the lawyer's role within it contribute to a better understanding of the ethical dimensions of the relationship between the practice of law and the attainment of justice.

Adversarial System and Individual Rights

Among the central tenets that frame the practice of much of the law in the United States and that therefore provide the context against which ethical considerations must be evaluated, is the adversarial administration of law. According to Luban, the adversarial system is distinguished from other systems of law by the following features: "an impartial tribunal of defined jurisdiction, formal procedural rules, and . . . assignment to the parties of the responsibility to present their own cases and to challenge their opponent's."^[1] For the purposes of a discussion of legal ethics, the latter feature is most significant. The responsibility of parties to present their own cases and to challenge the opposing party's case imposes on lawyers for each party the following duties: the duty of partiality to the client's perspective and interests, the duty of confidentiality, and the duty of zealous advocacy of the client's legal interest. These are the three fundamental duties of lawyers in an adversarial system of law.

The adversarial administration of law is typical of most common law legal systems, although the alternative inquisitorial model is dominant in countries with histories tied more closely to the civil law tradition. Most nations whose modern administrative and legal systems have been inherited largely from the British through colonial expansion and settlement employ some variation of an adversarial legal system. Many other nations, for example, France and Germany (and those nations that have inherited the French or German administrative systems through colonial or other means), have legal systems based on the inquisitorial model. The inquisitorial system affords lawyers a lesser role, the judges a greater role, and requires the disputing parties to cooperate in the impartial search to uncover the truth of the matter by neutral investigation.

The adversarial model, however, assumes that truth will come through the rational, rule-governed, presentation and challenge of facts by partisan parties. According to Lawrence Friedman, the adversarial system is characterized as the sort of system in which

the parties (and their lawyers) control the case. They plan the strategy; they dig up the evidence; they present it in court. The two sides battle it out mainly by putting witnesses on the stand and asking questions The judge sits on the bench, more or less in the role of an umpire. He or she sees to it that both sides obey the rules of the game.^[5]

This characterization is most appropriate to the operation of the criminal legal system, but is less descriptive of the practice of family law and of the juvenile justice

system. Both of these specialized areas of law are less adversarial than the criminal law. The former results from recent modifications to increase the settlement of disputes via mediation; the latter, by design, is a means to ensure the protection of vulnerable and inexperienced children. However, the juvenile justice system has become increasingly adversarial as the parties involved seek to protect the expanding rights of individual children accused of criminal or otherwise deviant behavior.

The justification for the adversarial pursuit of justice is that it is a more fair and equitable means to attain justice for the individual parties to the dispute. Indeed, Monroe H. Freedman, one of the strongest proponents of the adversarial system, characterizes it with the following claim, “Let justice be done—that is, for my client let justice be done—though the heavens fall. That is the kind of advocacy that I would want as a client and that I feel bound to provide as an advocate.”^[6] Indeed, he continues, the inherent fairness of the system comes from the claim that “[t]he adversary system not only takes the fact of disagreement into account, but takes advantage of it to produce a judgment informed by contrasting points of view. In doing so, the adversary system serves the public interest in a unique way.”^[6]

To underscore the close connection between the duty to zealously represent one’s client, the centrality of the adversarial process, and the larger public good the New York Code presents the matter thus,

An adversary presentation counters the natural human tendency to judge too swiftly in terms of the familiar that which is not yet fully known; the advocate, by zealous preparation and presentation of facts and law, enables the tribunal to come to the hearing with an open and neutral mind and to render impartial judgments. The duty of a lawyer to a client and the lawyer’s duty to the legal system are the same: to represent the client zealously within the bounds of the law.^[3]

In this way, the New York Bar envisions the relation between securing rights for individuals as parallel to securing the public good. The adversarial system combined with the duty of zealous representation ensures, so the reasoning goes, the best interest of the public as a whole because it preserves the proper functioning of the legal system. To reiterate the point that Monroe Freedman makes, this is the sort of representation we each are entitled to expect from our lawyer and it is what each would want “if we should come to need our own champion against a hostile world.”^[6] In this way, the strongest justification of the adversarial system is political. The struggle is between the individual and other possibly hostile, but at least contrary individuals. The struggle may take place between even

less equal parties as, for example, between the individual and the state. In this view, the preservation of both the individual client’s rights and the public good requires a strong and zealous defense by lawyers whose duty is to their client alone.

Indeed, it is because of the constitutional weight given to the protection of individual rights and to the duty of lawyers to defend their clients’ rights that the adversarial system is dominant in the United States. Wherever it is found, the adversarial system demands that lawyers secure the individual rights of their clients through the zealous defense of their clients’ claim against the possible claims of others. As a lawyer, then, the right thing to do is to serve one’s client to the limits of the law, to the limits of professional ethics, and if necessary past the limits of public morality. It is at the juncture of risking or threatening to pass the limits of public morality that the limits of formal legal ethics appear. Critics of the adversarial system are troubled that it necessitates circumstances in which a lawyer must choose between being a good person and being a good lawyer.

Role Morality and Moral Responsibility

After considering what lawyers’ duties to their clients are in an adversarial system, we should consider whether any ethical considerations might limit the zealotry of their advocacy. According to the ABA Model Rules,

[a] lawyer’s responsibilities as a representative of clients, an officer of the legal system and a public citizen are usually harmonious. Thus, when an opposing party is well represented, a lawyer can be a zealous advocate on behalf of a client and at the same time assume that justice is being done.^[7]

However, the actual experience of practicing law in an adversarial setting does not always lead to this anticipated harmony of the client’s interest in justice in their own case, the lawyer’s interest in succeeding as an ethical professional, and the public interest in justice and fairness according to law. Indeed, the ABA Model Rules are intended as a guide for those occasions when this harmony is undermined. The Model Rules are intended to guide scrupulous lawyers to identify their duties and thereby to determine what course of action they ought to take to resolve ethical conflicts. The concern about the conflict between role morality and public morality exposes the limits of the ABA Model Rules and of any other ethical codes governing the practice of law in an adversarial system. Here, we should ask, “Can a good lawyer also be a good person?”

Several scholars of law and ethics have attempted to answer exactly this question. Luban argues that, in practice, being a good lawyer often is antithetical to being a good person. The ethical (and constitutional) requirement to render a zealous defense of a client's rights often requires that a lawyer forego the requirements of conventional (i.e., public) morality. Why is this so?

An important feature in discussions of legal ethics is role morality. Role morality refers to the moral conduct required of individuals as they perform specialized tasks typically associated with a defined social role. This approach finds its roots in the work of Niccolo Machiavelli, a sixteenth-century Italian philosopher of politics.^[8] The virtuous ruler (Prince), he argues, must forego the ordinary virtues of the good man and do what is necessary to his function as ruler. The job, it would seem, is not for the morally squeamish. Similarly, discussions of legal ethics often turn to a consideration of the role morality of the legal practitioner. For practicing lawyers, the requirements of the role are such that one must be willing and able to do what under other circumstances would be considered immoral or lacking virtue. Consider the lawyer defending her client in child custody hearings, but whose client is also a convicted child sex offender. Acting as a good lawyer, she owes her client the fullest and best advocacy of his interests. But as a good person, the lawyer may well find the thought of placing children in her client's custody to be morally reprehensible. She may rightfully worry that she would share responsibility for any harm that comes to the children once they are returned to their abusive father's custody, if she is successful in acting as a good lawyer should. This is the kind of dilemma raised by considerations of role morality in legal practice.

There are two broad approaches that can be employed in an attempt to reconcile the role morality of lawyers with the public morality of the society in which they practice. One is to argue that the role morality of legal professionals be brought into line with a universal, enlightened, and rational morality that places social justice at its center. The second is to argue that the role morality of legal professionals be made flexible so as to be consistent with the individual conscience of legal practitioners and the values of the community in which they work. The first position is reflected in the work of Richard Wasserstrom,^[9] the second in that of David Luban.^[1]

Wasserstrom suggests that the ethics of legal practice ought to be made compatible with a universal justified morality. He suggests that the public good, and individuals who comprise the public, would be better served by lawyers who "see themselves less as subject to role-differentiated behavior and more as subject to the demands of the moral point of view."^[9] If some

behavior otherwise required in the defense or advocacy of a client's rights violates a justified universally applicable moral principle, then that behavior ought not to be undertaken by the lawyer. If a lawyer cannot advocate or defend a client's interests without such violations of morality, then such defense of the client's interest must be contrary to the requirements of morality. An immoral interest cannot generate morally defensible rights and is inconsistent with justice. Successful advocacy of the client's interests would, therefore, be contrary to the requirements of justice. Here, the criticism is not that lawyers fail to follow any justified moral principles. Rather, the concern is that even *ethical* legal practice is itself unjustifiable in larger moral (rather than role-related) terms. If the practice of law is itself justified by its relation to the pursuit of justice, then a corrective would lie in making ethical legal practice consistent with universal morality.

This line of criticism raises concern about the morality or justice of ethical legal practice. The problem, as Luban and others see it, is that the practice of law is too far removed from the professed ends of justice so long as lawyers (as practitioners of the law) are understood to act according to a separately justifiable and independently derived morality that is separate and independent of the morality to which ordinary persons are expected to adhere. This criticism is posed as a problem not only for the legal profession, but also for public policy. If lawyers are held to a different moral standard in the attainment of justice for their clients and if that standard is morally questionable *but not in violation of ethical rules of conduct*, then how can the public be certain that justice will be secured? The problem is a perfect example of the moral disparity between identified ends (justice) and the means by which those ends are attained (zealous adversarial advocacy). Can a just end be attained by unjust or immoral means?

Monroe Freedman addresses the problem by resisting the move toward requiring lawyers to adhere to either the publicly accepted morality or some enlightened universal morality. His reasoning hinges on the centrality of the value of individual client autonomy. For lawyers to act according to any morality other than the ethical standards of the profession and the client's own moral values is to subject the client to paternalism. The ethical problem presented by paternalism is its violation of individual autonomy. To behave in this way is to encourage lawyers to supplant their own moral values or point of view for their client's.^[10]

The virtuous or good lawyer is not, on Freedman's account, so distinct from the virtuous person. Indeed, the role difference between lawyers and non-lawyers is significant. Roles can make all the ethical difference. Often individuals are held to differing standards of

conduct according to the socially determined role each occupies. Consider the following analogy. As a friend, we have an obligation to console in times of loss, which a stranger does not; a friend has duties to another friend by virtue of the special relationship of the friendship. We would neither expect nor feel justified to cajole a stranger to listen to our troubles, but we do so of friends. Freedman argues that lawyers, not unlike friends, have special duties to their clients because of the role lawyers occupy in the legal system. Their special training, their privileges of education and professionalism, and their special charge as proponents and advocates of the law, generate for lawyers duties to act in ways we would not, indeed could not expect of non-lawyers.^[10] For lawyers to do otherwise would be for them to fail in their appointed responsibilities not only to their clients, but also to the larger public. Freedman, among others, argues that zealous advocacy in the defense of the rights of individual clients serves the larger public interest because each one of us would want and is justified to expect our own lawyer to do just this in defense of our own rights, should the need for such defense arise.

Special Duties of Prosecutors and Government Lawyers

However, even so staunch a proponent of zealous adversarial advocacy as Freedman recognizes that there may be circumstances in which a lawyer has a larger duty to justice than only to his particular case or client. This is routinely the situation for public prosecutors and other government lawyers. Here, the public policy concern that the ethics governing the means of legal practice correspond closely with the larger conception of social justice is clearest. We may not all agree that individual defense attorneys or divorce lawyers have an ethical obligation to act so as to attain the greater justice of the community when this conflicts with their clients' interests. But, when the means pursued by the public prosecutor fail to serve the larger public good, then they fail to meet the requirements of social justice. In such circumstances, all members of the community have cause for concern.

Of course, public prosecutors must adhere to the ethical rules incumbent upon all practicing attorneys. But, by virtue of the prosecutor's role and relative power as representative of the state's interest, prosecutors have a special duty to ensure that the effect of their actions is to attain justice in so far as the law permits. The duty of prosecutors is not simply to win the conviction, but to ensure that the conviction once won is also a just conviction. Depending on the particular situation, this duty to ensure that the broader public good is served through each prosecutorial action

results in special duties incumbent upon prosecuting attorneys from which non-prosecutorial and non-governmental lawyers are exempt.

The special duties of prosecutors derive principally from the several discriminatory powers typically the preserve of clients in non-criminal matters. First, prosecutors have the power to direct criminal investigations. Second, they have the power to define the crime to be charged. Third, they have the power to affect punishment upon conviction by making recommendations to the court, by arguing for one sort of punishment or another, or persuading the court toward more lenient or more harsh penalties. Finally, and perhaps most important for the attainment of social justice, is the power of governmental lawyers to decide whether to prosecute a particular crime or individual at all.

Each of these powers can be misused or abused perhaps innocently, as in the zealous attempt to secure the public from the ill effects of crime, or more maliciously, as in the selective prosecution of politically unpopular individuals. In either case, the ethical concerns at issue for government lawyers are substantially greater than for non-government lawyers and this is due to the additional discretionary powers with which they are vested. The requirements of the ABA Model Code are clear, and there is near consensus on the matter in the ethics literature.

The discretionary powers of prosecutors ought to be used to effect the public good. Prosecutors who use this power well can enhance public confidence in the prosecutor's office, as well as improve the public's confidence in the legal profession as a whole. It is important to understand that these special discretionary powers reflect prosecutors' positions as advocates on behalf of the public as a whole and of the state, and whose duty it is to uphold the public interest in health, security, and especially justice. In civil law, parallel discretionary powers are vested in the client.

LIMITS OF LEGAL ETHICS

What appears from the above discussion of the relation between legal ethics and the public good are two important limitations on the effectiveness of any ethical regulation of legal practice. These include intention and knowledge.

Intention and Motivation

According to self-imposed standards for prosecutorial conduct, prosecuting attorneys are to pursue only those charges they reasonably believe are substantiated by available evidence and they are not to use inflated charges as leverage to coerce testimony.^[11] But how

is it possible to ascertain the purpose for which a prosecutor decides on one charge instead of another? This problem is a problem of intention, and is well known in moral and ethical discourse generally. The problem stems from the related metaphysical and epistemological problem posed by “other minds.” In philosophical discourse, the problem of other minds refers to the difficulty we each have in ascertaining 1) that other humans have minds “like I do,” and 2) what the content of another human’s mind might be. The content of a human’s mind refers to the beliefs and values that inform moral thinking, which in turn informs moral action.

As for many other areas of ethical inquiry, assessments of intention are often reduced to assessments of action and its consequences. Individuals are often unclear as to their own intentions when engaged in complex activities; this is no less true for practicing attorneys, whether they are prosecutors or not. Generally, only two avenues are available to determine what a person’s intentions are: 1) the person’s own declaration about what was intended; and 2) the act in which the person is engaged. Both these avenues are indirect and somewhat unreliable. A prosecutor brought before a disciplinary hearing may declare that the intentions were noble and moral, and were truly in accordance with the requirements of ethical conduct. But, the effect of what was done may indicate otherwise. In this way, the two available means by which to ascertain the lawyer’s intent may contradict. The personal declaration points to lacking intention (say, to introduce a perjured testimony), but the actions point to intention (perjured testimony was introduced at trial).

Attorneys may be motivated to do what is right and clearly intend to act accordingly, but nonetheless they may engage in actions that have negative legal consequences. One way to view the requirement that lawyers act according to ethics codes and rules is as an exercise in deontological ethics. According to Kant, the moral worth of an action does not lie in “the purpose to be attained by it, but in the maxim in accordance with which it is decided upon.”^[12] In other words, the goodness of my actions should not be determined by the consequences that result, but by the motivation I follow in acting as I do. If a lawyer acts according to the ethical rules yet those actions result in detrimental legal effects (e.g., perjured testimony is entered as evidence leading to an innocent person’s conviction), it is the lawyer’s motivation to follow ethical rules that should determine the moral quality of the resulting actions. However, motivation and intention are difficult to ascertain.

The problem of intention and its correlation with deontological ethics points to one of the weaknesses of deontological approaches to professional ethics. Following the rules, acting from duty, and acting with right motivation do not ensure right conduct. Instead

of focusing on the following of rules and relying on the good motives of lawyers (especially when their actions are detrimental to the legal system), lawyers ought to determine right actions by their good consequences regardless of the motivations for acting. This teleological or consequentialist alternative allows more concrete determinations of good conduct by focusing ethical attention on the effects of actions. On this account, lawyers can act ethically and bring about good consequences, even if their motivations for doing so are less than virtuous (in terms of either role, universal, or public morality). It is at this point that a discussion of legal ethics becomes involved in problems of ethics more generally, and these are problems of a more theoretical rather than practical nature. It also leads the discussion of legal ethics full circle, in another form, to revisit concerns about the ends of the practice of law, that is, about justice.

Problem of Limited Knowledge

The problem of limited knowledge also has philosophical roots. One of the most obvious ethical claims, almost itself a truism, is that a person cannot be responsible for what she does not know. Individuals can only be guided in their actions by what they know or have good reason to believe is true. If the bridge is out, but I do not know this and direct you to take the bridge to return to the highway, I cannot be responsible for the harm that befalls you when you drive off the bridge. Or can I? This problem is well known to ethicists and to lawyers alike. In law, culpability (responsibility for the ill effects of one’s actions) is often contingent upon the degree of knowledge an individual has about the relevant circumstances. Similarly, the rules governing ethical legal practice often contain an epistemological qualification. For example, lawyers should not allow testimony *they know is false* to be entered as evidence. If they do not know the testimony is false, have they violated the ethical prohibition by allowing the testimony?

The ethical worry seems to be that attorneys, aware of the reduced ethical responsibility that comes from ignorance, may make themselves willfully ignorant of some facts of their cases to ensure that they have not behaved unethically. The willfully ignorant lawyer is one who does not want to know and so does not pursue the truth of the matter regarding a witness’ testimony, for example. If the lawyer knows the testimony is false, then it should not be used in pursuit of the client’s interests. If the lawyer does so nonetheless, disciplinary proceedings and punishment might result. If the lawyer does not use it, the client’s case might be lost. The willfully ignorant lawyer is free, it seems, to use the otherwise suspect testimony as part

of a zealous representation of the client's interest. But, is the lawyer responsible for this willful ignorance and the possible detrimental effects of acting on it?

The distinction between innocent ignorance and willful ignorance, similar to the determination of intent, is not always evident and its moral status is somewhat controversial. This problem raises questions as to the criteria for knowledge. One familiar criterion is the reasonable person standard: a lawyer is responsible for acting when information to guide that action would be known by a reasonable person. A slight modification of the reasonable person criterion is reasonable expectation. A lawyer, *with* special training, experience, and access to information, is reasonably expected to know more than a reasonable person *without* such training and experience. Willful ignorance is less appealing as an ethical loophole if additional criteria of knowledge are employed.

CONCLUSION

The practice of law holds a privileged place in the social trust. The law is often the practical determinant of justice. The practice of law, then, is reasonably associated in the social conscience with the attempt to ensure some means for securing justice within a given society or community. At the very least, individuals tend to retain the services of legal professionals in the narrower pursuit of justice in their own case. For the individual members of the public, the law is often most personally and concretely represented by individual lawyers. For the legal profession, then, the ethical requirements of practice are closely associated with the attainment of justice in some form. Legal ethics is about ensuring this connection both through the regulation of professional legal practice, and through the normative discourse that critically evaluates actual practice and proposes to remedy the practical and theoretical limitations of such regulation.

REFERENCES

1. Luban, D. *Lawyers and Justice: An Ethical Study*; Princeton University Press: Princeton, 1988; pp. xviii, 57, 104–147.
2. Zitrin, R.A.; Langford, C.M. *Legal Ethics in the Practice of Law*; The Michie Company: Charlottesville, VA, 1995; 4–7.
3. *New York Code of Professional Responsibility*, 2002 Ed.; Preamble. Legal Information Institute, Cornell University Law School; <http://www2.law.cornell.edu/cgi-bin/foliocgi.exe/ny-code> (accessed June, 2002), EC 7-19.
4. Some of the following links may be useful in providing samples of the variety of codes, statutes, and interpretations of these which supplement the ABA Model Rules guiding lawyers in the ethical regulation of their practice. Many of these websites are regularly updated and reflect the current opinion on legal ethics and practice. Cornell University Law School's Legal Information Institute, <http://www.law.cornell.edu/>; Department of Justice's Office of Professional Responsibility, <http://www.usdoj.gov/opr/>; ABA's Center for Professional Responsibility, <http://www.abanet.org/cpr/>; National Organization of Bar Counsel (US, Canada, Australia), <http://www.nobc.org/>; Internet Legal Services' ethics opinions database, <http://www.legalethics.com/>; Association of Professional Responsibility Lawyers, <http://www.aprl.net/>; Tribal Law and Policy Institute's Tribal Court Clearinghouse, Tribal Codes and Constitutions, <http://www.tribal-institute.org/lists/codes.htm>.
5. Friedman, L.M. *American Law: An Introduction*, 2nd Ed.; W.W. Norton & Co.: New York, 1998; 87 pp.
6. Freedman, M.H. *Lawyers' Ethics in an Adversary System*; Bobbs-Merrill Co.: New York, 1975; pp. 9, 14, 24.
7. American Bar Association. *Model Rules of Professional Conduct*, 2002 Ed.; Preamble, Para. 8; http://www.abanet.org/cpr/mrpc/mrpc_toc.html (accessed June, 2002).
8. Machiavelli, N. *The Prince*; Adams, R.M., Ed.; W.W. Norton & Co.: New York, 1977; Trans.
9. Wasserstrom, R. Lawyers as professionals: some moral issues. *ABA Hum. Rights* 1975, 5, p. 12
10. Freedman, M.H. *Understanding Lawyers' Ethics*; Matthew Bender: New York, 1990; pp. 47–48, 57.
11. *National Prosecution Standards*, 2nd Ed.; National District Attorneys Association 1991; Standards 43.3 and 43.4, which read: **43.3** The prosecutor should file only those charges which he reasonably believes can be substantiated by admissible evidence at trial; **43.4** The prosecutor should not attempt to utilize the charging decision only as a leverage device in obtaining guilty pleas to lesser charges.
12. Kant, I. *Grundlegung zur Metaphysik der Sitten*, 2nd Ed.; (Trans. *Foundations of the Metaphysics of Morals*); as cited in Walsh, W.H.; Kant, I. *Macmillan Compendium to Philosophy and Ethics*; Borchert, D.M., Ed; Macmillan: New York, 1999; 531 pp.

Life Cycle Costing

Greg Chen
Daniel Williams
Troels Pind Adrian

*School of Public Affairs, Baruch College, City University of New York,
New York, New York, U.S.A.*

INTRODUCTION

Governments usually purchase goods and services through competitive bidding processes. However, government contracts are often awarded solely on lowest purchase price basis. The initial acquisition price does not reflect the recurring cost of operations and maintenance. Nor does it include the potentially large disposal liability, if environmentally hazardous materials are involved. Life cycle costing (LCC) is a technique used by government and business to determine the total cost of ownership. The results of an LCC analysis can be used to assist management in procurement decision-making.

This article introduces the LCC concept and its application. It first outlines the principles and processes of LCC. It then reviews the history and the evolution of the concept. It puts the technique in context by listing its problems, solutions, and current applications. It concludes by suggesting extended application of LCC as a decision-making tool in public purchasing and beyond.

WHAT IS LIFE CYCLE COSTING?

Life cycle costing is an application of present value concepts principally to purchasing decisions, although it may be applicable in other settings. The core idea is that minimizing the purchase price can be dysfunctional because other necessary costs may negate the advantage of the lowest purchase price of a good. For example, in anticipation that a good might be purchased at the lowest price, a manufacturer might use substandard materials or inadequate design, reducing manufacturing costs and allowing a low selling price. However, these shortcuts might result in such auxiliary costs as high maintenance or operating costs, short product life, low residual value, or high disposal cost. Once these additional characteristics are considered, the decision to select on lowest purchase price alone is revealed to be an ineffective use of resources.

Life cycle costing may sound exotic, but it is actually quite common. The concept can be used at home

and in institutions, in the private sector and in the public sector, in consumption and in production.^[1] For example, the energy efficiency ratings marked on major electronic appliances sold in the United States are simply a clever way of communicating a component of LCC information to the appliance market. The selection of building materials based on initial and operating cost in school construction is an example of LCC in public sector capital budgeting.^[2] The evaluation among various product lines based on total costs is an example of the application of LCC in private enterprise investment decisions.^[3]

HOW DO I PERFORM LIFE CYCLE COSTING?

Life cycle costing requires two basic steps. First, one must determine all costs associated with the purchase, ownership, operation, maintenance, disposal of wastes, and disposal of a product. In general, LCC is applied as a microeconomic concept, not as a public policy concept, so only direct costs are considered. Each cost must be estimated both in amount and in time flow, unless the time period is quite short. Second, the time flow costs must be discounted to present value.

As an illustration, a simplified life cycle cost of a project can be calculated using the formula:

$$\text{Life cycle costs} = A + O + M + D$$

where A is the acquisition cost of a project, which includes the initial capital expenditure for facilities and equipment, the system design, engineering, and installation. These costs typically incur at the beginning of the project. O is the operating cost, which is total operation labor cost, energy cost, material cost, overhead cost, etc. discounted over the life of the project. M is the maintenance cost, which is the total maintenance and repair costs discounted over the life of the project. D is the disposal cost, which incurs when the project involves toxic material that impact on the environment. Disposal can also be a benefit when the project has a terminal value, which is its present

value of the project in the final year of the life cycle period.

The present value of a cash flow in a future time must be discounted for comparison. The discounting process follows the time value of money principle. The project with the lowest life cycle costs is most cost effective, and should be considered among alternatives in purchasing decisions.

Sometimes, alternative projects produce goods or services of uneven quality and/or characteristics. In this case, it may be necessary to consider benefits in addition to cost; otherwise hidden costs in the form of unequal benefits may not be exposed. There can be significant unequal benefits when maintenance requires that a product be out of service or when there is a difference in the product capacity, output, life, or residual value/disposal cost. As with any benefit–cost analysis, these values should be monetized, placed in time, discounted to present value, given the appropriate sign, and added to the costs. To avoid misleading results, the costs and benefits should be separately revealed before they are netted together.

The recommended decision rule should be to purchase the product with the highest net present value, except where such nets are small in proportion to total flows. In that circumstance, other product characteristics may be more substantially considered.

A BRIEF HISTORY OF LIFE CYCLE COSTING

Life cycle costing is such a common sense approach that it has likely been around long before it gained a foothold in academic literature. That foothold is in the 1960s in the United States and the 1970s in Britain.^[4–7] Its formal development is at the Department of Defense (DOD) as an operations research method. In 1967, DOD solicited an operations research analyst in the Wall Street Journal, listing knowledge of LCC as an expected qualification. The implication of this job notice is that the term had substantial currency by 1967. However, the major business, finance, economics, political science, statistics, and history of science journals, as participating in JSTOR, list nothing before the 1970s. *Public Administration Review*, a non-JSTOR journal, has one brief mention of LCC in 1968 and associates it with total package procurement (TPP), which itself is not substantially older having one *PAR* citation from 1966 and one JSTOR citation in the *American Economic Review* from 1968.^[4,6,7] Interestingly, Neiburg's 1968 article, which is the first identifiable mainstream article that goes beyond merely mentioning these concepts, is primarily concerned about difficulties that arise with TPP in the defense environment.

Over the past 40 years, the basic character of LCC has not changed. What has changed is its application and devices for its communication. Life cycle costing began its existence as a technique in contract procurement to avoid perverse side effects of minimum bid purchase decisions. It has expanded into a consumer device to make better purchasing decisions and may sometimes be used as a device to require manufactures or other sellers to internalize costs that they typically try to evade by leaving as externalities. It has, thus, proved to be a useful and flexible device that extends beyond the public finance environment.

PROBLEMS AND SOLUTIONS

Life cycle costing is relative of net present value and benefit–cost analysis. It is, thus, subject to some of the same kinds of abuses that these might suffer. These include deliberate manipulation of discount rates, inaccurate estimation of some costs or benefits, the omission of sources of cost or benefit, and the inability to reach agreement on monetary value of some costs or benefits.

The solution to most of these technical difficulties can come in three forms: rules, sensitivity analysis, and audits. Typically, the federal government sets very clear guidelines on what discount rates may be used in present value calculations based on Circular A-94, http://www.whitehouse.gov/omb/circulars/a094/a94_appx-c.html. As LCC is not used in considering the tradeoff between the use of money for public or private purposes, the Circular A-94 discount rates may not be appropriate. The local government may be more interested in a discount rate that reflects its own cost of borrowed money. However, the discount rate should be set in advance of LCC at a fixed amount by decision-makers whose choice is unaffected in the outcome of the LCC analysis. Decision-makers should be particularly cautious of implausible discount rates when alternate product selections lead to sharply differing cash flows.

Uncertainty about costs can be addressed through sensitivity analysis. Sensitivity analysis consists of making estimates of several plausible levels: expected, somewhat lower than expected, and somewhat higher than expected. All estimates should be realistic; however, they may reflect the uncertainty surrounding unknown future costs and benefits. In sensitivity analysis, the product is fully costed at its expected cost and benefit and again at its lowest cost/highest benefit and highest cost/lowest benefit. With these three estimates in hand, the decision-maker is able to see to what degree the conclusion is robust. Sensitivity analysis

can also be applied to uncertainty concerning the discount rate.

Uncertainty about sources of costs and benefits can be addressed through independent review, auditing in some contexts. To be effective, this review should be genuinely independent. Independent review should also be considered for issues of discount rates and cost estimates.

The inability to monetize costs or benefits is more difficult to resolve. Empirical economists model and monetize many decisions that others find abhorrent to monetize, such as the value of human life. There are both moral and epistemological questions that render these results uncertain. The practice nevertheless persists. It is not proposed that this issue can be resolved in this article.

CURRENT APPLICATIONS

Although having a long history, the application of LCC is limited in public purchase decision-making. In addition to the theoretical, technical, and philosophical concerns depicted above, data quality and quantity limit its wide application. Researcher and analyst often find that relevant and timely data are not always available. The acquisition of the data can be costly and time consuming. Given that LCC is data intensive, lack of quality cost and performance data impede its application.

Moreover, the bureaucratic and legal constraints of governments could prevent the use of LCC. In many instances, the local government may require the selection of the least expensive options measured only in acquisition cost. Budget constraints also limit the use of life cycle cost, which, at a minimum, includes the cost of acquisition, operation, maintenance, and disposition.^[8]

Finally, like cost-benefit analysis, an LCC analysis is one piece of information that feed into the public decision-making. Many other considerations, political and bureaucratic, may enter the equation and determine the selection outcome. In the present highly politicized public service organizations, political considerations often overrun economic and efficiency concerns.

CONCLUSIONS

Life cycle costing is a powerful and yet simple device that can improve procurement decisions, aid consumers, and be applied in the public policy arena. It appears to be underused at this time. A number of approaches can be used to improve its utility in public

procurement decision-making. Education is often the key, as not all stakeholders are fully aware of the benefit of the technique, and not all government agencies are capable to apply it. Second, the regular collection of relevant, reliable, and timely cost data is prerequisite to its wider application. Without access to data, the application of this relatively data intensive techniques cannot be implemented. Third, the change of government purchasing by-laws and regulations from least bidding price criterion to life cycle cost will promote the use of the technique. Finally, public demands for government economy, efficiency and effectiveness promote the use of rational analytical techniques, such as LCC, that reduces long term costs for government services.

Like many management tools, LCC is alive and evolving. New and hybrid procedures, such as activity-based LCC, a combination of activity-based costing and LCC,^[9] have emerged in recent years. Many computer models have been developed to make these techniques accessible to government in-house application. As governments increasingly face tough choices under severe fiscal and economic pressures, and as the general public becomes more and more concerned of energy costs and environment issues, the future of LCC and its newly arrived derivatives, as a technique in public decision-making, seems to be very promising.

REFERENCES

1. Cole, R.; Sterner, E. Reconciling theory and practice of life-cycle costing. *Build. Res. Inf.* **2000**, 28 (5/6), 368–375.
2. Kreuze, J.G.; Newell, G.E. ABC and life-cycle costing for environmental expenditures. *Manage. Acc.* **1994**, 75 (8), 38.
3. Moussatche, H.; Languell, J. Flooring materials life-cycle costing for educational facilities. *Facilities* **2001**, 19 (10), 333.
4. Nieburg, H.L. Social control of innovation. *Am. Econ. Rev.* **1968**, 58 (2), 666–677 (Papers and Proceedings of the 8th Annual Meeting of the American Economic Association). Available: <http://links.jstor.org/>.
5. Sizer, J. Some implications for operational research in the area of inflation as it affects a company's profitability and liquidity. *Oper. Res. Q.* **1977**, 28 (1, Part 2), 125–137. Available: <http://links.jstor.org/>.
6. Staats, E.B. The GAO: present and future. *Public Adm. Rev.* **1968**, 28 (5), 461–465. Available: <http://links.jstor.org/>.
7. Webb, J.E.; Cleveland, H.; Gross, B.M.; Staats, E.B.; Ahmed, L.N. Currents and soundings: from the professional stream. *Public Adm. Rev.* **1966**, 26 (3), 205–216. Available: <http://links.jstor.org/>.
8. Lee, R.C. Life-cycle costing. In *Tools for Decision-Making, A Practical Guide for Local Government*; Ammons, D.N., Ed.; CQ Press: Washington, DC, 2002.

9. Emblemavag, J. Activity-based life-cycle costing. *Manage. Audit. J.* **2001**, *16* (1/2), 17.

BIBLIOGRAPHY

1. Fabrycky, W.J. *Life-Cycle Cost and Economic Analysis*; Prentice Hall, 1991.
2. Fuller, S.K. *Life-Cycle Costing Manual for the Federal Energy Management Program*, 1995th Ed.; U.S. Department of Commerce, Technology Administration, National Institute of Standards and Technology: Gaithersburg, MD, 1996. Available: <http://purl.access.gpo.gov/GPO/LPS16832>.
3. National Association of State Purchasing Officials. In *State and Local Government Purchasing*, 4th Ed.; National Association of State Purchasing Officials: Lexington, KY, 1994.
4. Seldon, R.M. *Life Cycle Costing: A Better Method of Government Procurement*; Westview Press, Inc.: Boulder, CO, 1979.

Linear Programming

Ronald John Hy

*College of Arts and Sciences, Texas A&M University-Kingsville,
Kingsville, Texas, U.S.A.*

INTRODUCTION

Linear programming (LP) is a set of mathematical optimization equations that are used to solve resource allocation problems; LP provides a narrow-range estimation when a course of action definitely leads to a given result, even though the range of that result is unknown before computation of the equations. The most important aspect of LP is that it provides a framework for making projections. That is, LP forces users to determine the objectives and constraints (restrictions that affect the possible solution) associated with estimations. Thus, LP helps users improve their planning skills by expanding their analytical capability and forcing them to state assumptions (objective and constraints) explicitly. LP, moreover, enables users to improve their estimation technique by requiring precise specifications of objectives, key variables, and crucial constraints. Such precise specifications help users understand the important ramifications and effects of each alternative course of action.^[1]

KEY CHARACTERISTICS OF LINEAR PROGRAMMING

Although LP is quite beneficial to the analytical process, users must remember that LP is foremost a quantitative technique and, as such, incorporates the following distinct characteristics:^[1]

- The problem must be formulated in quantitative terms.
- The problem must include at least two realistic choices among the proposed solutions. (No decision needs to be made when there is only one way to achieve an objective.)
- The problem must include at least one crucial constraint (any restriction that limits choice).
- The problem must have at least one optimal solution. (The solution must maximize an efficient course of action or minimize an inefficient course of action.)
- The key elements (the objective, variables, and constraints) must be measured in non-negative

numbers. (It makes absolutely no sense to talk about a negative solution.)

- The relationships among key variables must be linear. (Variables have to have a constant and reciprocal ratio.)
- The relationships among key variables must be additive, meaning that the total effect of the variables is equal to the sum of the effect of each variable.

CONCEPTUALIZING THE PROBLEM

The following steps are used to build LP equations:

- The objective of the problem (either to minimize inefficiency or maximize efficiency) is clearly defined.
- Key variables are specified.
- Constraints (the restrictions placed on the solution) are unambiguously developed.

LP is a set of iterative procedures that first satisfies all the constraint equations. This solution, then, is modified to satisfy the objective equation. The process is repeated until no further improvement is possible. Each iteration comes closer to the optimal solution until the optimal solution is reached.

Mathematical Notation

To use LP, the objective, key variables, and crucial constraints of a problem must be stated as mathematical equations so that the optimal solution can be computed. The first equation (the objective equation) states the relationship between the objective of the problem (minimizing inefficiency or maximizing efficiency) and the key variables. The other equations (constraint equations) specify the limitations imposed on the solutions. One constraint equation must be developed for each constraint.^[1]

Using mathematical notation, a set of LP equations is written as follows:

To minimize an objective:

$$Z(\min) = a_1X_1 + a_2X_2 + \cdots + a_nX_n \text{ (objective equation)}$$

Subject to the following constraints:

$$b_1X_1 + b_2X_2 + \cdots + b_nX_n > k_1 \text{ (constraint equation)}$$

$$c_1X_1 + c_2X_2 + \cdots + c_nX_n > k_2 \text{ (constraint equation) and so forth.}$$

To maximize an objective:

$$Z(\max) = a_1X_1 + a_2X_2 + \cdots + a_nX_n \text{ (objective equation)}$$

Subject to the following constraints:

$$b_1X_1 + b_2X_2 + \cdots + b_nX_n < k_1 \text{ (constraint equation)}$$

$$c_1X_1 + c_2X_2 + \cdots + c_nX_n < k_2 \text{ (constraint equation) and so forth.}$$

Where

a_1 = the objective value associated with the first key variable (X_1)

X_1 = the first key variable

a_2 = the objective value associated with the second key variable (X_2)

X_2 = the second key variable

a_n = the objective value associated with the last key variable (X_n)

X_n = the last key variable

b_1 = the value associated with the first key variable of the first constraint equation

b_2 = the value associated with the second key variable of the first constraint equation.

b_n = the value associated with the last key variable of the first constraint equation

k_1 = the minimal or maximal value of the first constraint equation

k_2 = the minimal or maximal value of the second constraint equation

c_1 = the value associated with the first key variable of the second constraint equation

c_2 = the value associated with the second key variable of the second constraint equation

c_n = the value associated with the last key variable of the second constraint equation.

Because LP problems often include hundreds of variables and constraints, a computer solution is a virtual necessity. Because of the iterative nature of LP, a variety of standard easy-to-use LP computer packages have been written. These LP packages usually can calculate problems involving more than 1000 variables and constraints.

Sensitivity Analysis

Sensitivity analysis answers the question: How much can the values associated with the key variables vary without drastically affecting the optimal objective? Because the values of the optimal solution often represent estimates, sensitivity analysis helps determine the range of the optimal solution. (Most LP computer solutions automatically yield sensitivity analysis information.) For instance, assume that the value associated with supplement *I* is three pounds. The optimal objective of \$560.00 can be maintained if supplement *I* is decreased by 0.25 pounds or increased by 0.21 pounds.^[1] That is, the optimal amount of supplement *I* actually can vary between 2.75 and 3.21 pounds without affecting the optimal cost (\$560.00).

Sensitivity analysis provides an acceptable range to the optimal solution. Such an analysis also facilitates determining the absolute maximum or minimum values that can be used without violating the constraints or objective. Obviously, the smaller the decrease and increase, the sounder the optimal solution.

LIMITATIONS OF LINEAR PROGRAMMING

LP is a powerful technique that can be used to make narrow range projections. Users, however, must be aware of the major limitations of LP if they are to fully understand and interpret the technique.^[1]

First, LP can only be applied to projections that have a single objective. Unfortunately, some projections frequently have multiple objectives. If general rather than specific objectives are sought, this limitation can be overcome. One way of dealing with multiple objectives is to develop a set of LP equations for each objective and determine via sensitivity analysis the range of each of the optimal solutions. The various optimal solutions then can be used as ballpark figures for making projections. Another way of dealing with multiple objectives is to use a more sophisticated LP format called goal programming. Goal programming lets users set upper and lower limits for each objective and then rank them in ordinal sequence. Such a format allows users to specify the extent to which each objective is expected to be achieved within the prescribed resource constraints.

Second, LP assumes the relationships among variables are linear. (This restriction exists for maximization and minimization problems.) Linearity assumes that all statistical relationships among variables are proportional; that is, the coordinates fall on a straight plane. If the linear relationships among variables cannot be assumed to exist, the LP solution is not useful. There are, however, various non-linear programming techniques that are similar to LP and can be used when

the relationships among variables are assumed to be non-linear.

Users must constantly remember that the optimal solution, though mathematically feasible, may not be socially desirable or politically viable. Moreover, it is often impossible to reconcile conflicting objectives. Consequently, an LP solution must be perceived as the starting point in an attempt to make projections. Armed with the optimal solution, users can then consider the myriad political and social factors before finally deciding the “best” solution.

LP is normally exceedingly complex, involving an enormous number of variables and constraints. The value of understanding LP lies in being aware of the

utility, applicability, and limitations of the technique so it can be used correctly. The procedures, concepts, and calculations employed to solve a problem can be applied to any projection, regardless of size. The only constraint on the size of an LP problem is the computer’s computational capability. LP is limited only by one’s imagination.

REFERENCE

1. Hy, R. *Financial Management for Health Care Administrator*; Quorum Books: New York, 1989; 145–147, 157.

Local Governments: State Supervision of Budgets and Finance

Julia Beckett

*Department of Public Administration and Urban Studies, University of Akron,
Akron, Ohio, U.S.A.*

INTRODUCTION

States supervise local government budgets by requiring specific actions and procedures, by establishing and limiting revenue sources, and by monitoring and reviewing budget actions of local governments. States fulfill their responsibility to supervise in five ways: states enable, authorize, and empower local governments; states set budget controls and fiscal limits; states impose requirements and mandates; states employ administrative supervision; and states intervene in fiscal crises.

The state supervisory role depends on state laws and practices, as well as on the type of local government involved. The patterns and types of state supervision can only be described in general terms because there is great variety within the 50 states and more than 85,000 local governments in the United States. The types of local governments include counties, cities, towns, townships, school districts, and special districts. The jurisdiction of local governments may be based on geography, services, or population. Counties and cities are general-purpose governments, and they provide a wide range of services. In contrast, special districts often provide one service or function, and they are the most numerous and varied type of local governments in the United States. Within this diversity, states exert control over local government budgets.

One of the most important government functions is budgeting. Budgeting links the ability to tax and spend with managerial planning. Other governmental powers linked to budgeting include the ability to contract and to define and undertake activities. By enabling and defining local governments, states both empower and limit local government budget authority. Few states directly charter local governments and define their authority in detail; it is more common for states to have general laws that express broadly the extent of local governments' powers and functions. The techniques of state supervision range from general to specific, and the amount of control states exert over local governments ranges between loose and tight.

STATES EMPOWER LOCAL GOVERNMENTS

The intergovernmental authority relationship between state and local governments is first defined as that of

superordinate and subsidiary. This is often poetically stated as "local governments are the creatures of the states." This formal structure of power, also called Dillon's rule, is that local governments can be created or abolished by the state, and so they are subject to limits imposed by the state. States enable local government actions, including budgeting, spending, taxing, and borrowing. Local governments may exercise only those powers expressly granted by the state constitution or by state statutes. There is no common law right of local self-government, and local government powers to tax and spend are subject to state constraints.

In practice, states empower local governments through broad grants of powers; the specific type and manner of public activities undertaken are determined by local governments. The independent authority of local governments has been asserted through the use of municipal or county home rule charters, but this autonomy is limited. Some local government powers remain subject to preemption by the state. Quite often, even with home rule, states impose limits on local fiscal actions.

Although the hierarchical powers and derived authority description of local governments as being subsidiary and subservient to the states, the relationship between the state and local governments regarding budgets supervision is quite varied in practice. The manner that supervision takes is through intergovernmental techniques. State statutory authority of intergovernmental control is described in six major categories: 1) forms of government and home rule; 2) altering boundaries and responsibilities; 3) local elections; 4) administrative operations and management; 5) financial management; and 6) personnel management.

Commonly, supervision is defined as "directing the work or actions of another." The typical supervisory relationship is a manager and worker in an organization. This simple model of supervision within a hierarchical organization with centralized authority is, however, not consistent with American intergovernmental practices. In practice, state legislative and administrative review, oversight, and regulation of local governments are complex. States and local governments take cooperative, coordinated actions. State authority is often felt indirectly, rather than exerted directly. In practice, states and local governments work

as a partnership, addressing shared public problems with mutually developed policies.

States establish rules, regulations, and conditions within the political and administrative systems that affect local government budgets. State statutes spell out detailed and stringent requirements on local governments' budget and revenues. However, the intergovernmental supervision has indirect means of review or enforcement, which often relies on self-implementation of requirements by local governments.

States enable local government to try management techniques to reduce budget costs. States encourage coordination and networking of governments in risk or purchasing pools. States enable contracting out and other methods to reduce costs of service production or provision.

States supervise indirectly by their grants of authority, by revenue sharing, by requiring actions and services, and by requiring standard accounting and auditing procedures. States control local government budgets by setting specific limits on revenue sources and amounts. In addition, states engage in monitoring local governments by requiring them to file budget and fiscal documents. States possess the authority (although they seldom exercise it) to intervene directly and circumscribe local government budget actions.

STATES SET BUDGET CONTROLS AND FISCAL LIMITS

States exert control over four areas of local government financial management: 1) budgeting; 2) raising revenue; 3) issuing debt; and 4) financial reporting. These four inter-related categories are essential to local governments' budgeting and management decisions; yet, few states require a comprehensive approach to financial management. Most states' legislation requires local governments to enact an annual budget and sets requirements for that budget. The required budget information may range from a simple summary of revenue and expenditures to a detailed line item budget format. Local governments are commonly required to follow a three-step procedure—first submitting the proposed budget to the local governing body, then providing for a public hearing, and finally formally adopting the budget by motion and vote of the governing body. The enacted budget is often required to be filed with the state.

State-required budget actions include prescribing budget content, level of detail, timing, and processes. Some states encourage using program or performance budget formats. Other states may specify using a fiscal year or impose a calendar for considering and

enacting a budget. Often, the amount of information included in a budget varies by the size and type of local government, with more sophisticated approaches expected of larger city and county governments.

Some states institute more stringent requirements, using highly specified and detailed budget forms. Another stringent control prohibits local governments from making transfers between expenditure categories once a budget is enacted. Some states specify which local official, whether elected or appointed, is responsible for the form and management of the budget. These officers may have required qualifications such as certain training or experience.

State supervision over raising revenue takes various forms. It is common for state legislatures to institute stringent limits and controls on taxes and revenue sources for local governments. Types of statutory controls include limiting the types of revenues that can be raised, exempting some properties from the tax base, and placing caps on tax rates. States specify revenue types and sources available to local governments. Historically, property taxes were the predominant source of local government revenue and they continue to be the primary revenue source for school districts. State controls applied to property tax include methods of assessing property value, creating or denying exemptions, and limiting tax rates.

For local governments, other than school districts, revenue sources include sales, income, and excise taxes; non-tax revenues include user fees for services and enterprise funds. States exert budget spending control by limiting tax base or tax rates and, thus, limiting tax revenues collected. One stringent practice is indexing or limiting increases from previous years. States often earmark revenues by limiting the purposes that certain revenue can be used; for example, states may limit the revenue from gasoline taxes to be used for road repairs. States impose conditions for use of different types of taxes; an example is the centralized state collection of sales taxes.

An area of particular concern and attention is local government borrowing and use of debt to pay for expenditures. Many state provisions regulate local government borrowing based on the purpose, amount, and term of debt. Detailed conditions exist for both long- and short-term debts. Many states encourage local governments to institute capital budgets and to link long-range capital planning to managing debt. The particular processes and steps that states establish for local governments to enact budgets or borrow funds are often based on administrative and good government reform movements or are instituted by states in response to problems and scandals.

State constitutions limit local government budgeting, taxing, spending, and borrowing. The local

government-balanced budget requirement is the most well-known constitutional budget restriction; it requires that the local governments' annual operating budget to balance which is enumerated as planned expenditures must not exceed projected revenues and reserves. Another common constitutional constraint requires voter referenda to approve general obligation debts, and this serves to limit borrowing. Since the 1980s, in response to citizen tax revolts, many states have added constitutional provisions that require voter referenda before local governments can increase taxes or user fees.

The purposes for state controls on local government budgets and revenues are to impose fiscal restraint and provide public accountability. Many states instituted limits and prohibitions after public outcry. For instance, state limits on borrowing were adopted after numerous local governments defaulted on development bonds in the 19th century. State-defined budget controls and spending limits are intended to impose fiscal restraint; however, local governments often respond by circumventing the regulations and limits through innovations and creative financing. For example, the innovation of revenue bond financing removed the requirement of public acceptance and authorization of debt. Local governments issue revenue bonds to avoid exceeding limits on permissible levels of general obligation debt. The use of special district governments is also seen as a way for local governments to circumvent state-imposed limits on debt and budgets. Other innovations include increased use of fees for services, public-private partnerships, and shifting service responsibilities.

State fiscal reporting requirements also affect local government budgeting. To ensure uniform practices and to improve fiscal management of local governments, states require general fiscal reporting, book-keeping, accounting, and auditing standards. Historically, states required itemized budgets and detailed information to improve local governments' administration and operations. Revenue and borrowing provisions were enacted as constraints on local government spending. Financial reporting provisions also serve administrative functions for local governments and to provide information to the public.

State enforcement of budgeting requirements varies. Sometimes, there are statutory requirements for local governments, but no statutory enforcement mechanisms. Sometimes provisions are expected to be implemented by local government officials, with public pressure for compliance. Even when states have similar budgeting requirements, they may differ widely in approaches to ensure compliance. Whether minimal, advisory, or mandatory, these fiscal standards and practices affect the budgeting decisions and management of local governments.

STATES IMPOSE REQUIREMENTS AND MANDATES

One of the more contentious areas for local government budgeting in recent years has been the issue of mandates. State and federal government mandates have imposed increasing responsibilities on local governments, by requiring specific services, certain levels of public activity, or administrative procedures. For example, mandates concerning clean water impose new reporting and training activities, or mandates regarding public access require expenses of retrofitting buildings and sidewalks. State-required activities, services, or programs may be imposed through legislation, executive order, administrative rules, or court decree. The broad reach of mandates affects local governments by reducing discretion in service delivery and because the expenses entailed in satisfying mandates limit the personnel and resources available.

The controversy over mandates has frequently focused on whether the government that requires certain actions of local government should fund the required actions. Often, it is difficult to determine the costs and administrative burdens of mandates. For example, if a new state law requires local governments to install mechanical signs to warn of approaching trains for all streets that have railroad crossings at grade, then the question arises as to whether the state should pay the expenses for these signs. When states policies and citizen revolts have constrained revenue sources, mandates impose great pressures on local government budgets, because they require a greater number of expenditures.

State fiscal transfers and grants also affect local budgets. A major source of revenue for local governments has been state aid, and the timing and amount of this funding affects budgets. Grants-in-aid are awarded to local governments from state and federal programs. The direct conditions of grants may impose additional financial burdens in the form of requiring services, matching funds, or establishing additional procedures. Thus, in addition to general mandates, specific grants of aid have strings attached that serve to supervise local government practices.

STATES EMPLOY ADMINISTRATIVE SUPERVISION

States have administrative agencies that affect, advise, and regulate local governments both indirectly and directly. State administrative supervision of local government may range from informal advice to training, to periodic review, and to direct prior approval for local government action. Many states have administrative departments or divisions of local government

that provide general training, advice, and support for activities such as economic development, budget preparation, or database management. Some state agencies provide advice in response to local government requests. Oversight, in these cases, may be informal and non-directive.

Other state agencies may have direct regulatory authority, including setting standards and requiring actions of local governments. For example, state departments of education set standards and requirements for school districts. States may have regulatory agencies that can approve or deny local government fiscal activities; for example, some states have agencies that authorize local governments to issue securities such as bonds.

States also require local governments to file copies of the budgets and annual financial reports with an agency, often the state auditor or state treasurer. The requirement to file these budgets and reports serves three purposes: they impose managerial practices; they provide the general public access to information; and they enable state agencies to review for fiscal responsibility. The offices of state auditor and treasurer often serve as regulatory agencies in regards to the filing, form, and substance of local governments' budgets and financial statements.

STATES INTERVENE IN FISCAL CRISES

A fiscal crisis of insolvency, default, or bankruptcy raises grave concerns about the stability of a local government. In the event of a fiscal crisis, states take a more direct role in the management of the local government's budget. When audits reveal that local governments have troubled fiscal conditions of ongoing deficits or insolvency, states intervene in the day-to-day operations to achieve budget stability and fiscal solvency. Some states impose increasingly stringent reporting and filing requirements regarding the fiscally troubled local government's financial obligations, contract and purchasing actions, fund management decisions, and general operational decisions. Other states provide for receivers to take over the financial management and governance of insolvent local governments, and some states allow local governments to reorganize debts in bankruptcy litigation.

It is in fiscal crisis situations that states take the most directive role in a local government's budget and operations. This stringent and direct control is

temporary. Once the crisis is resolved, states return budget control to local officials.

CONCLUSION

State supervision of local government budgeting is complex and changing. In the federal system, states are expected to supervise local governments, and they employ various oversight techniques. In regards to the collection of revenue and the distribution of services, states affect local government budgets. States mandate functional and policy requirements that have serious budget implications. State supervision of local government taxing, spending, borrowing, and reporting includes both authorizing and setting limits on these activities. Through other means, such as giving advice, states exert control over local governments. The tensions between local governments asserting independence and states taking supervisory action raise continuing issues in intergovernmental oversight.

ARTICLES OF FURTHER INTEREST

Financial Resource Management in Local Government, p. 766.
Fiscal Federalism, p. 783.
Forms of Government, p. 800.

BIBLIOGRAPHY

1. Advisory Commission on Intergovernmental Relations. *State Laws Governing Local Government Structure and Administration*; Government Printing Office: Washington, DC, 1993.
2. Leighland, J. External controls on public authorities and other special purpose governments. In *Public Authorities and Public Policy: The Business of Government*; Mitchell, J., Ed.; Greenwood Press: New York, 1992; 31–48.
3. Rubin, I.S. *Class, Tax and Power: Municipal Budgeting in the Untied States*; Chatham House: Chatham, NJ, 1998; 173–230.
4. Sbragia, A.L. *Debt Wish: Entrepreneurial Cities, U.S. Federalism and Economic Development*; University of Pittsburgh Press: Pittsburgh, PA, 1996; 80–162.
5. Wright, D.S. *Understanding Intergovernmental Relations*; Brooks Cole: Belmont, CA, 1988.
6. Zimmerman, J.F. *State-Local Relations: A Partnership Approach*; Praeger: Westport, C, 1995.

Local Services, Provision and Production

Robert M. Stein

Department of Political Science, Rice University, Houston, Texas, U.S.A.

INTRODUCTION

What governments do is significantly related to how they arrange for the provision and production of their functional responsibilities. There is evidence that links how governments arrange for the provision and production of their service responsibilities with the scope and content of their functional responsibilities and the efficiency, effectiveness, and citizen satisfaction associated with the performance of these functional responsibilities. In this article, I review the theoretical and empirical literature on the relationship between municipal functional responsibility and the means by which municipalities organize for the provision and production of their service responsibilities. My review of the literature is augmented with new analysis of functional assignment patterns and service arrangement data for a sample U.S. municipal government surveyed in 1997. The article concludes with a discussion of emerging research questions.

PREVIOUS RESEARCH

Savas^[3] and other researchers^[4–7] have argued that private vendors and not governments are invariably more efficient and effective producers of the goods and services provided by municipal governments. These researchers have marshaled empirical evidence that shows that private sector service producers are superior (i.e., more efficient) to public sector producers of the same goods and services (cf. Ref.^[6]). Others^[1,2,8,9] have suggested that the service producer and its sector location may be less significant than the mode of service delivery used by the city. Different means of service arrangements, both public and private, provide for greater efficiency in the provision of different goods and services. Given the content of the service/good and the nature of the market for service provision, the relevant issue in the debate about efficiency of municipal services is not whether government entities or private entities are more widely used, but whether governments have matched their service

responsibilities with the appropriate method of service arrangement.

Tiebout's^[10] model of local public expenditures has provided the theoretical foundation for most empirical research about the functional assignment patterns among municipal governments. Tiebout maintains that individuals select communities in accordance with their preferences for publicly provided goods and services, that is, "voting with their feet." The content and level of municipal service bundles is defined by the preferences of the median voter in each community. For Tiebout, an efficient market mechanism for the production of goods and services exists by virtue of: 1) the number and diversity of states and communities and 2) the mobility of voters to choose from among these communities the one that best approximates their preferences for community goods and services.

Tiebout's model of local expenditures is neutral regarding the content of governmental functional responsibility and similarly agnostic about what functions state and national governments should assume. Buchanan,^[11] Peterson,^[12] and Miller^[13] each introduce an important qualification to Tiebout's theory of local expenditures that significantly informed the debate on the scope of and content of municipal public policy. Each author notes that the mobility of consumer voters in a metropolitan area produces a significant bias for the retention of wealthier residents. Buchanan demonstrates that collective decision making under majority rule with equality of consumption and taxation will advantage low-income citizens at the expense of high-income residents. The result is income redistribution, a significant cost for middle- and high-income persons that can be avoided by the out-migration to another jurisdiction. Peterson further explores this bias toward the retention of wealthy residents. His thesis is "local politics is not like national politics," consequently "there are crucial kinds of public policies that local governments simply cannot execute."^[12] By inference, Peterson claims there are policies only the national government can efficiently perform.

Peterson maintains that the scope and level of a community's functional responsibilities are constrained by the inability of cities, and to a lesser extent states, to control the mobility of capital and labor across their respective borders. Though national governments can control the flow of capital and labor

This article draws upon the author's previously published work (see Refs.^[1,2]).

across their borders, states and cities are less effective at preventing their citizens, businesses, and capital base from migrating to other locales where tax-to-benefit ratios are superior. Consequently, governments seek to maximize their economic well-being by pursuing policies that enhance the economic base of the community, while avoiding those policies and activities that threaten the community with the loss of productive capital and labor. For cities, the pursuit of developmental policies (roads and highways) generates economic resources (i.e., tax base) that can be used for the community's collective welfare. Conversely, cities are expected to avoid redistributive public policies that benefit dependent and non-productive persons (e.g., welfare, housing, and health care), while drawing resources away from productive citizens without providing these persons commensurate benefits. Again, by implication, the responsibility for redistributive policies is less harmful to states and national governments because of the significant transaction costs associated with migration between states and nations.

The evidence to support Peterson's empirical hypotheses is mixed. Municipalities spend considerably more on developmental and allocation functions (road, street, water, and other infrastructure) than on redistributive goods and services (health, welfare, and housing). Peterson further shows that measures of demand (e.g., poverty, income, and the percent of municipal population that is African American) are unrelated to spending for redistributive social services. Stein^[1] reports, however, considerable variation in the range of municipal spending on redistributive social services. Many more communities than would be expected by Peterson and market theorists have an active role in the provision of social policies that are potentially harmful to interests of the city. Why?

As suggested above, there are several important dimensions to the debate over what functional responsibilities government should provide for their citizenry and how these governments should arrange for the provision of these goods and services. Different goods and services present cities with different challenges. Consequently, different strategies are needed to overcome these obstacles to the delivery of different goods and services.

The empirical question is whether there is an observable relationship between the attributes of different municipal service responsibilities and the method cities employ for the delivery of these service responsibilities. My thesis is that what governments do is fashioned by the means they employ for providing these goods and services. Efficiency and program effectiveness are maximized when functional responsibilities are properly matched with the optimal mode of service provision. This condition mitigates the income redistribution associated with municipal

service responsibility for many social services, enabling municipal governments to provide these services without suffering the significant flight of productive capital and labor. When functional responsibilities are matched with appropriate mode of service arrangement, many of the problems associated with externalities and other inefficiencies are eliminated.

THE ATTRIBUTES OF GOODS AND SERVICES

The failure of vendors, both private and public, to optimally produce specific goods and services derives from two traits associated with all goods and services: exclusion and jointness of consumption. The interaction of these two traits identifies four different types of goods and services, each requiring different institutional arrangements for their efficient service delivery.^a

Exclusion occurs when producers at a relatively low cost can deny buyers of a good or service consumption or use of the product. The absence of exclusion can lead to the production of externalities: "When the actions of one individual confers benefits or imposes costs on another individual for which no fee can be charged or no recompense collected."^b According to Peterson, it is the existence of externalities created by spending decisions of city governments that results in the flight of productive capital and labor. Specifically, cities actively pursue developmental policies (e.g., roads and highways) that generate economic resources (i.e., tax base) that benefit the entire community. Conversely, cities avoid redistributive policies that benefit dependent and non-productive persons (i.e., the poor) while drawing resources away from productive citizens without providing these individuals with commensurate benefits. These types of policies (welfare, housing, health, and hospitals) operate to encourage in-migration of unproductive and dependent persons to the city, while driving out more productive non-dependent persons to communities where they can receive a more favorable ratio of benefits to taxes paid.

Jointness of consumption exists when one person's consumption of a good or service does not diminish the consumption of that particular good by another person(s). Goods that diminish with their consumption lack jointness of use and possess the trait of subtractability. Most goods and services possess some degree of subtractability: At some level of consumption, one additional person's consumption of a good subtracts from any other person's consumption of that good. The level or threshold at which subtractability increases is defined as congestion and can drastically

^aThis discussion draws upon Ostrom and Ostrom (see Ref.^[14]).

^bJohn Heilburn (see Ref.^[15]).

affect the provision of goods and services. As cities grow, public roadways that bestow equal access to all residents can no longer provide adequate service to all citizens without significantly impairing service levels of other citizens (e.g., traffic delays).

Goods and services can be initially characterized by the two extreme conditions of jointness and exclusion. Goods for which exclusion is feasible and that possess a significant degree of subtractability are defined as private goods (e.g., food, health care, housing) and can be expected to be produced by private markets. The other extreme, where goods are characterized by joint/non-subtractable usage and non-exclusion, defines a collective good (e.g., national defense, pollution control). Between the extremes of private and collective goods are toll goods and common property resources. Toll goods have joint use but feasible means of exclusion. Movie theaters, libraries, cable television, and emergency medical services are common examples of these goods. Private vendors can arrange for the delivery of these goods (e.g., entrance fees to movie theaters), and further regulation (e.g., no drinking or smoking in the theater) can ensure that subtractability does not diminish usage and enjoyment of the good or service for some users. For common pool resources, exclusion is difficult, but consumption by one user will diminish consumption by other users. The issue of service arrangement for common property resources is how to manage and conserve a resource for which private markets are inadequate.^[16] This is achieved by employing institutional arrangements that price access to the common property resource and thus convert the flow or product of the common property resource into a private good (e.g., charging fees for ambulatory and emergency medical services).

Service delivery of each type of good presents a different set of problems for a governmental unit, particularly for municipalities. The externalities produced

by municipal service responsibility for non-excludable goods and services vary with the specific institutional arrangement employed to fulfill this type of service responsibility. Municipal service delivery of non-collective goods and services through collective taxation produces externalities for those individuals (e.g., wealthier citizens) who may significantly under-value the good or service. The key to municipal service arrangement is the proper match between benefits received and costs paid by each individual consumer, particularly for the public provision of private and toll goods. Identifying the location in the service delivery process where inefficiencies occur can help to identify alternative institutional mechanisms for efficient and equitable service arrangements.

For collective goods and services, inefficiencies are likely to arise from the absence of the appropriate scale of production. A solution for this problem requires more than an alternative arrangement for financing service delivery. Production agreements with vendors who can vary the size of their production plant is one alternative means of service production that municipal governments use to remedy the inefficiencies that result from an inadequately sized production plant. The appropriate scale of production varies with the attributes of the good or service.

The Modes of Service Arrangement

Table 1 identifies nine alternative institutional arrangements for the delivery of municipal service responsibilities. The columns identify the components of service delivery. Each row in Table 1 identifies a unique mode of service arrangement, defined by the mix of municipal responsibility for planning, financing, producing, and distributing a specific good or service. Cell entries identify the role of the municipal government in each

Table 1 Alternative institutional arrangements for municipal service delivery

Service mode	Planning	Financing	Producing	Distributing
<i>Traditional</i>				
Direct	+	+	+	+
Contract	+	+	O	O
Joint contract	+	+	+/O	+/O
<i>Regulatory</i>				
Voucher	+	+	O	O
Subsidy	+	+	O	O
Tax incentives	+	+	O	O
Franchise	+/O	O	O	O
Volunteers	+	O	O	O
Self-help	+	O	O	O

Note: + Government has an active role; O government does not have an active role; +/O government has a limited role.

stage of service delivery. This role takes on two basic forms: 1) the government substitutes itself for another vendor and directly produces and distributes the desired good or service or 2) the municipal government alters the relationship between a vendor (private or public) and the consumer. This latter policy mechanism is identified in the policy literature as the regulatory mode and is directed at altering or controlling the behavior of individuals in order to achieve a specific policy outcome. The major distinction between modes of service arrangement is whether the municipal government takes a regulatory or traditional role in arranging for the service. Empirically, the distinction between regulatory and traditional service modes is a function of the scope and content of municipal responsibility for different phases of service delivery. Traditional service modes have the government assuming responsibility for the planning and financing of the service activity. Only responsibility for production is shared with or assigned to another entity. When governments adopt regulatory service modes, they assume only partial responsibility for planning and financing and assign responsibility for production and distribution to other governments or private vendors.

Service Arrangement and Policy Attributes

The 10 alternative modes of service arrangement vary in their suitability for each type of functional activity. The delivery of collective goods is most likely to be dominated by direct municipal arrangement. Their traits of non-exclusion and non-subtractability make the production of collective goods quite problematic for any private vendor. Even contractual arrangements for the production of collective goods are problematic because of the difficulty of unambiguously pricing the good or service. Common property resources are likely candidates for direct municipal service delivery. The subtractability of these goods and services enables governments to convert common property resources into private goods, once exclusion is achieved by the government. This is achieved by appropriating the flow of the common property (e.g., emergency medical services) and setting a price for the sale of the resource. Vouchers, franchises, and some contractual arrangements are common service modes for toll goods. The municipal government awards an exclusive or limited license to a vendor(s) to sell the regulated good to individuals residing in the municipality. This enables those who undervalue the toll good to avoid its consumption and any tax contribution to the municipal delivery of the good.

Private goods are unlikely candidates for a traditional mode of service delivery. The exclusiveness and subtractability of private goods and services make

Table 2 Service responsibility and the direct mode of arrangement by policy type: 1997

	Private	Toll	CPR	Collective
% responsible	50.3	55.8	84.2	83.8
% directly arranged ^a	44.0	64.7	66.3	76.0

^aArranged entirely with the municipality's own employees.

Source: From Ref.^[17].

a non-direct service mode a more efficient and equitable service arrangement. In the case of private and toll goods, there are ample opportunities for consumers to pay individually for the delivery of these goods or services. Moreover, these financing mechanisms (e.g., user fees, vouchers, subsidies) can be finely tuned to a desired or tolerable level of income redistribution. Substituting a collective tax with a subsidy, voucher, franchise, or user fee allows municipal governments to closely match individual preferences for these goods or services with the recipient's willingness and ability to pay for the services.

I expect that municipal governments that choose to provide their citizens private goods (e.g., health, hospital, and welfare services) will choose a non-direct service mode to fulfill this responsibility. A non-direct mode for these redistributive goods and services will significantly curtail the municipality's responsibility for direct financing. This should lessen the non-equivalence produced by a collective tax for the delivery of a good that bestows concentrated and exclusionary benefits. Conversely, a direct mode of service arrangement will be modal for collective service responsibilities, specifically those with significant jointness of consumption and limited opportunities for exclusion.

Empirical Evidence

Elsewhere, I have validated these hypotheses for a sample of U.S. municipal governments in 1982 and 1987.^[1,2] Table 2 reports the pattern of functional responsibility and mode of service arrangement (i.e., direct and non-direct) for a sample (N=1220) municipalities surveyed in 1997.^c Again, the pattern is consistent with previous findings and the hypotheses outlined above. First, municipalities have limited functional responsibility for private and toll goods and services. Functional responsibility is greatest for collective goods and services and common property resources. These findings are consistent with the

^cThese data were collected by the International City Management Association (see Ref.^[17]). The response rate for the survey conducted between the fall of 1997 and spring of 1998 is 32%.

Peterson's expectation and previous research. Cities eschew responsibility for private goods and services.

As expected, however, the modal means of service arrangement for private goods and services is some vendor or vendors other than the municipality. Another governmental unit and/or a variety of private sector for-profit and non-profit agencies are the most frequent source of service provision for municipal government's private service responsibilities.

To test the hypothesis that the scope of municipal functional responsibility for private goods and services increases with a non-direct mode of service provision, I have estimated a model where the dependent variable is the proportion of private goods provided and the independent regressors are the proportion of private functional responsibilities arranged by an entity other than the municipality, the population size of the municipality, and dummy measures for the region in which the municipality is located.^d The estimate for the proportion of private goods and services arranged by non-direct means is statistically significant and positively signed (t-value=7.6, $P < 0.000$). A 1% increase in the proportion of private sector goods and services arranged by non-direct means is associated with a 0.26% increase in the scope of functional responsibility for private goods and services. The positive and significant effect of non-direct modes of service arrangement on the scope of functional responsibility for private goods and services is modest. This finding, however, supports the argument that a municipality's aversion to responsibility for these potentially redistributive goods and services can be mitigated by non-direct provision of these potential troublesome functional responsibilities.

CONCLUSION

The positive relationship between the mode of service arrangement and the scope of functional responsibility, especially for private goods and services has been confirmed since 1982. This raises several interesting questions for further study. What are the political consequences of this relationship for continued public support for the municipal provision of private goods and services? Buchanan, Miller, and Peterson argue that cities that fail to avoid responsibility for redistributive social services suffer both economic distress and political fallout, i.e., defeat of incumbent officeholders. Is the latter true? What are the political consequences of the municipal responsibility of redistributive social services, especially under different modes of service arrangement? There is nearly a quarter century of

municipal experience with alternative modes of service arrangement, specifically for redistributive social services. This time frame offers researchers sufficient opportunities with which to study the political consequences of alternative modes of service arrangements.

REFERENCES

1. Stein, R.M. *Urban Alternatives: Public and Private Markets in the Provision of Local Services*; University of Pittsburgh Press: Pittsburgh, PA, 1990.
2. Stein, R. Arranging city services. *J. Public Adm. Res. Theory* **1993**, 3, 66–93. (Spring, 1993).
3. Savas, E.S. *Privatization: The Key to Better Government*; Chatham House: Chatham, 1987.
4. Stevens, B. *Delivering Municipal Services Efficiently: A Comparison of Municipal and Private Service Delivery*; 1984. A report prepared for the U.S. Department of Housing and Urban Development, Washington, DC.
5. Butler, S. *Privatizing Federal Spending: A Strategy to Eliminate the Deficit*; Universe Books: New York, 1985.
6. Borchering, T.; Burnaby, B.; Pommerehne, W.; Schneider, F. Comparing the efficiency of private and public production: Evidence from five countries. *J. Econ.* **1982**, 2, 127–156.
7. Hodge, G.A. *Privatization: An International Review of Performance*; Westview Press: Boulder, 2000.
8. Ferris, J. The public spending and employment effects of local service contracting. *Natl. Tax J.* **1988**, 41, 207–217.
9. Ferris, J.; Graddy, E. Contracting out: For what? With whom?. *Public Adm. Rev.* **1986**, 46, 332–344.
10. Tiebout, C. A pure theory of local expenditures. *J. Polit. Econ.* **1956**, 65, 416–424.
11. Buchanan, J. Principles of urban fiscal strategy. *Public Choice* **1971**, 4, 1–16.
12. Peterson, P. *City Limits*; University of Chicago Press: Chicago, IL, 1981; 4 pp.
13. Miller, G. *Cities by Contract*; MIT Press: Cambridge, MA, 1981.
14. Ostrom, E.; Ostrom, V. Public goods and public choices. In *Alternatives for Delivering Public Services*; Savas, E.S., Ed.; Westview: Boulder, 1977; 7–14.
15. Heilburn, J. *Urban Economics and Public Policy*; St. Martins: New York, 1987.
16. Ostrom, E. *Governing the Commons*; Cambridge University Press: New York, NY, 1995.
17. ICMA. *Profile in Local Government Service Delivery*, 1997; 1998. <http://icma.org/download/cat15/grp120/sgp224/asd97web.pdf>.

BIBLIOGRAPHY

1. ICMA. Alternative approaches for the delivery of public services. *Urban Data Serv. Rep.* **1982**, 14, 10–22.

^dThe northcentral region is the excluded category.

Logical Positivism and Public Administration

Laurent Dobuzinkis

Department of Political Science, Simon Fraser University,
Burnaby, British Columbia, Canada

INTRODUCTION

Public administration is both an art and a science. In both senses of the term, it deals with the management of public affairs. But, as a social-scientific discipline, public administration asks, “*why* are public policies formulated and implemented in the way they are?” or “*why* are government agencies operating the way they do?” Practitioners, in contrast, are more concerned with the issue of *how* to run government agencies efficiently, yet in a manner consistent with legal standards and norms of democratic accountability. This distinction must be kept in mind when assessing the influence of positivism on, and its relevance to, public administration.

Positivism offers a certain view of how research outcomes and procedures can be deemed to be scientifically valid. Thus, it appears to be more relevant to the study of public administration than to its practice. But public administration as a practice often relies on scientific and technical expertise. Positivism used to be the authoritative definition of the scientific methods until the mid-twentieth century. Today, however, it has become a contested notion; not surprisingly, therefore, the same debates that have divided or continue to divide other disciplines are replicated in public administration.

THE RISE AND FALL OF LOGICAL POSITIVISM

Logical positivism in a strict sense consists of certain positions adopted by twentieth-century philosophers of science (e.g., Rudolf Carnap and Carl Hempel) in their attempt to establish the best way to define the goals and methods of science. In that sense, it can be contrasted with other schools of thought (e.g., pragmatism and relativism). It refers more specifically to the ideas developed and discussed by a group of influential Austrian philosophers and scientists in the 1920s (i.e., the Vienna Circle). But many of their ideas can be traced back much farther. The British empiricists of the seventeenth and eighteenth century (from John Locke to David Hume) insisted on the primacy of empirical observations over idle speculation and began to articulate the essential principles of the inductive

method.^a In so doing, they laid the groundwork for the future development of logical positivism. Indeed, logical empiricism is sometimes used as a synonym for logical positivism.

The other sources of logical positivism are developments in formal logic, which took place in the late 1800s and early 1900s. These efforts were aimed at clearly demarcating logical relationships from empirical ones to clarify what one is testing when subjecting a scientific theory to the double tests of experimental verification and logical consistency. The stated purpose of logical positivism was to achieve a rigorous synthesis of logic and the experimental method of science.

Logical positivism can be characterized in terms of three basic principles: 1) facts must be collected by impartial observers in an objective manner (the facts-values dichotomy); 2) theoretical hypotheses must be validated experimentally (the verification principle, which, in its most radical form, states that statements that cannot be experimentally validated are meaningless); and 3) the goal of scientific research is to identify law-like regularities that can be used to explain and/or predict both natural and social phenomena (the covering laws principle). Or, to summarize, science is derived from facts and consists of testable hypotheses that can be applied to an ever-widening range of phenomena.^b

Although logical positivism was for some time regarded as the most sophisticated analysis of the scientific method, it was not long before critics began to question the validity of its fundamental principles. Contemporary philosophers and sociologists of science do not agree amongst themselves about what alternative view or views are preferable. But there is a general agreement that pure facts do not exist. Factual observations are almost always colored by prior commitments

^aDavid Hume, however, stressed that generalizations arrived at inductively are never absolutely true.

^bFor a more subtle discussion of the positions defended by the Austrian logical positivists than can be offered here, see Friedman, M. *Reconsidering Logical Positivism*; Cambridge University Press: New York, 1999. On the place of positivism within the social sciences, see Hollis, M. *The Philosophy of Social Science: An Introduction*; Cambridge University Press: New York, 1994; chapter 3.

to certain theoretical views, if not by certain sociological beliefs that scientists share as members of the institutions to which they belong (the importance of the latter is more hotly debated than that of the former). However, recent scientific developments strongly suggest that many phenomena, especially in the domain of the social sciences, are not characterized by stable and universalizable relationships. Such phenomena consist of dynamic, sometimes even chaotic, patterns that are difficult and probably impossible to predict accurately or even explain unequivocally.^c

Positivism is not dead, however. In the philosophy of science, the excesses of postmodern constructivism—the idea that not only are facts colored by theoretical preconceptions, but also that science is merely another way of constructing reality out of the values and beliefs scientists bring to their work—have led the way back, if not to strict positivism, at least to “critical realism.” In the social sciences, quantitative methods of analysis continue to be justified in terms of a more or less positivist epistemology. Indeed, public administration scholars often use analytical instruments (e.g., survey research) that pose little or no challenge to positivist methodology.^[1] These methods are generally regarded as constituting only one of several tools from which researchers can choose. Public administration scholars have learned to appreciate the strengths and weaknesses of various approaches to the study of government and public policy. Eclecticism, however, is not necessarily as commonly found among the many experts in applied sciences and engineering—including experts who hold managerial positions in various government agencies—who continue to adhere to uncompromisingly positivist conceptions of science. Technically trained experts have been known to resist attempts by stakeholders or officials in other agencies to question the validity of their assumptions about the policy issues to which their expertise is relevant.

PUBLIC ADMINISTRATION: POSITIVIST BEGINNINGS

In both Western Europe and North America, public administration as a profession defined by strict standards of technical competence, and an academic discipline, came into existence roughly between the 1870s and the early 1900s. This was precisely the era

when positivism gained ascendancy among scientists and philosophers of science, leading eventually (see above) to the logical positivist synthesis. (Positivist conceptions of law as an instrument of statecraft divorced from morality also emerged at the same time.) During that time, various administrative reforms were implemented to deal with a host of new demands placed on public institutions by rapidly industrializing societies. Modern bureaucracies staffed by competent experts replaced ad hoc administrative structures whose principal function had often been the allocation of patronage appointments. The reigning positivist orthodoxy provided strong justifications for this transfer of control and power to a new class of bureaucratic specialists, even if many of them were not actually trained in the natural sciences but in disciplines such as accounting, law, or economics.

It has become customary to cite Wilson and Max Weber as the authoritative contemporaneous sources on these reforms.^[2] The conventional wisdom is that they defended a strict separation between the domains of politics and administration. Administration became synonymous with objective knowledge and technical expertise, whereas politics was understood in terms of partisan conflicts and subjective values. Administration was all about policy implementation and politics about policy formulation. This is an oversimplification. To begin with, neither of these two scholars can be deemed to have had much influence on the reforms they were describing or even on the crystallization of the politics–administration dichotomy into the dogma of their times. Max Weber’s works were not discovered in translation by American sociologists of administration until the 1940s, and the creation of the administrative state in the Europe of his time owed nothing to his influence. In fact, his role in that respect was more that of a critic than of a reformer. As for Wilson, the politics–administration dichotomy, which he himself became rather critical of later in his life, was not picked up by his contemporaries as having been his major contribution to the field. It is true that the fledgling academic discipline of public administration in the early 1900s often emphasized the conceptual differences between politics and administration, but pioneers such as Frank Goodnow, Leonard D. White, or W.F. Willoughby never insisted on the need to build walls between these two domains.

All the same, it is clear that between the 1880s and the 1920s, the first steps toward the creation of an administrative state were taken, and the study of public administration found its place within political science departments in American universities. (One must also mention the creation of New York City’s Bureau of Municipal Research in 1907.) Positivist beliefs about the capacity of science to produce objective knowledge of both natural and social phenomena strongly

^cThis second criticism does not directly pose a challenge to science itself; chaos theory proposes testable hypotheses about systemic patterns. But it shows that observed facts cannot always be used to make reliable predictions about specific events whose uncertainty is, therefore, irreducible.

influenced these initiatives. It was believed that a more scientific approach to the management of both public and private organization would be a marked improvement over the age of the amateurish entrepreneur or narrowly partisan politicians (as the saying goes, “there is no Republican way to build a road”).

THEORY AND RESEARCH: EXORCIZING THE GHOST OF POSITIVISM

Throughout the second-half of the twentieth century, public administration moved away from positivism. However, depending on whether one considers the academic or practical side of public administration or whether one looks at macrotheoretical developments of microlevel analyses and models, the aptness of that remark varies somewhat.

Since the 1940s, public administration theorist, organization theorists, political scientists, and sociologists have attacked the Weberian bureaucratic model for being exaggeratedly mechanistic and rigid, dehumanizing, and intellectually bankrupt. Some of these attacks are directed at straw men whose actual goals had also been reformist. Nevertheless, the linkage between these critiques and the decline of positivism is indirect but rather obvious. Insofar as these critiques emphasize the interdependence of political values and administrative dilemmas, the importance of horizontal, non-hierarchical communications, the messiness of most policy problems, the contingent and value-laden nature of the information available to decision maker, and the need to empower both street-level civil servants and their clients, among other things, they run counter to the positivist commitment to objectivity, as well as the certainty inherent in technocratic methods of posing and resolving problems. There has been, therefore, a general trend toward a postpositivist, if not truly postmodern, understanding of governance issues.

Although this has been the case at the macrolevel, research agendas at the microlevel have moved in another direction. When Herbert Simon wrote *Administrative behavior* in 1945, he demolished the scientific pretensions of earlier attempts by Luther Gulick and Lyndall Urwick to construct a science of administration. His intention was to fashion more truly scientific analytical tools for the study of decision making within organizations. In typically positivist fashion, he insisted on the need to empirically test whatever statements public administration scholars make about their subject. Simon's recommendation has not always been followed in practice, but remained unchallenged for several decades. To this day, the empirical analysis of decision making, performance, and other aspects of organizational life (e.g., the demographics

of the public service; who believes in what in an agency and to what effect; and levels of job satisfaction) has become increasingly sophisticated, although perhaps not as sophisticated or as common as in other disciplines.^d The way in which these models are designed and tested is consistent not only with logical positivism, but also with its reduction to linear modeling (i.e., proportionality of causes and effects). In recent years, some original attempts at applying the mathematics of non-linear dynamics to administrative behavior have brought to light the unpredictability inherent in the outcomes of even some of the most basic administrative routines,^[3] but these remain rather atypical research projects.

More radical challenges to positivist tenets have come from studies that emphasize the extent to which policy analysts and managers interpret and reinterpret the information they receive from their clients, and communicate to their colleagues or hierarchical supervisors; the same is true of the latter in their relations with the former.^[4]

THE REAL WORLD OF PUBLIC ADMINISTRATION

Many of the new ideas about governance alluded to above have been implemented under the banner of the “new public management” in the United States, Canada, and several other countries around the world. One of the tenets of the new public management is that the traditional bureaucratic emphasis on procedures must give way to an emphasis on outcomes. In some respects, this trend creates a greater need and demand for objective methods of performance measurement and of evaluating the actual impact of government programs on the economy. Computers make it possible to analyze an increasing amount of data on internal managerial issues, as well as on relations with clients, other agencies, and so on. (At this level, there might be an opportunity for greater congruence between empirically oriented public administration scholars and practitioners who, as a rule, were not always particularly interested in the quantitative side of public administration research.)

However, the tendency to encourage innovation, flexibility, and responsiveness point in the direction of a more intuitive management style. Coupled with political and societal trends, such as deregulation, the

^dSeveral authors have lamented the relative lack of statistical sophistication that characterize much public administration research in comparison to sister disciplines such as political science; see, for example, Gill, J.; Kenneth, J.M. Public administration research and practice: a methodological manifesto. *J. Public Admin. Res. Theory* 2000, 10 (1), 157–199.

increasingly pluralistic character of civil society, and the deepening chiasm between the world views of various policy communities (e.g., environmentalists vs. energy producers and regulators), the shift toward the new public management poses an increasingly severe threat to the ideal of objective knowledge as a foundation for administrative expertise. Agencies whose organizational culture rests on that notion (e.g., the U.S. Forest Service) are more likely to find such changes in their political and social environment disturbing. This background accounts in large measure for the growing interest in non- or anti-positivist approaches.^[5,6] It would be a mistake, however, to conclude that these approaches have become dominant.

CONCLUSION

While Charles Lindblom did not have the field of public administration itself in mind when he coined the phrase “muddling through”—he was referring to what policy makers do—it seems an apt description of how public administration deals with methodological issues. Public

administration scholars and practitioners continue to be divided among themselves but perhaps even also within themselves about the merits and weaknesses of positivist methods.

REFERENCES

1. Miller, G.J.; Whicker, M.L.; Eds. *Handbook of Research Methods in Public Administration*; Dekker: New York 1999.
2. Wilson, W. The study of administration. *Polit. Sci. Quart.* **1887**, 2, 197–222; Reprinted in *Pol. Sci. Quart.* **1941**, 56, 481–506.
3. Kiel, L.D. *Managing Chaos and Complexity in Government*; Jossey-Bass: San Francisco, 1994.
4. Majone, G. *Evidence, Argument and Persuasion in the Policy Process*; Yale University Press: New Haven, 1989.
5. Spicer, M.W. Determinism, social science, and public administration: lessons from Isaiah Berlin. *Am. Rev. Publ. Admin.* **2005**, 35 (3), 256–269.
6. Miller, H.T.; Fox, C.J. *Postmodern Public Administration*, Revised Edition; M.E. Sharpe: Armonk, NY, 2006.

Logistics and Transportation

Linda L. Stanley

School of Business, Our Lady of the Lake University, San Antonio, Texas, U.S.A.

Darin Matthews

Multnomah County Schools, Portland, Oregon, U.S.A.

INTRODUCTION

The management of transportation services is frequently an overlooked opportunity to contribute to the success of governmental operations that depend on purchasing's ability to evaluate transportation options and carriers, determine the means to reduce transportation and inventory carrying costs, expedite and trace shipments, and process claims. The selection of a transportation system is a specialized form of procurement in which services are purchased to provide a link among manufacturing sites, warehouses, distribution centers, and/or office sites. Components of the transportation service buy include 1) evaluating the prices of various transportation methods, 2) determining the effect of transportation service on operations, and 3) evaluating the effect of transit times, carrier dependability, and safe delivery on inventory levels.^[1]

Traditionally, the procurement and management of transportation services were relegated to the supplier who provided a "laid-down price"—the cost of transportation rolled into the price of purchased goods. This manner of operation was particularly popular prior through the late 1970s because rates were regulated by federal and state governments and were not negotiable. However, since deregulation, flexible pricing plans and more service options are available. Carriers are now logistics service companies and have integrated their expertise with that of their customers.^[2] Thus savings opportunities are possible by paying closer attention to the costs of transportation, whether directly arranged by the procurement officer or managed by a supplier. Second, businesses and government agencies have begun to realize the value of evaluating the *total cost* of procuring goods, which may include transportation costs, carrying costs, disposal costs, etc.

This article is designed to provide governmental purchasers an introduction to transportation management, to offer strategies to effectively manage the procurement of these services, and to suggest some future trends. We will start with some definitions common to the transportation industry.

DEFINITIONS

Transportation is "the physical movement of goods and people between points."^[3] Goods may include new materials brought into a production process or finished goods shipped to the customer. *Carriers* are those organizations that provide transportation services, and the primary *modes* include railroads, motor carriers, airlines, water carriers, and pipelines. Procurement officers may also hire *third parties*, or intermediaries, who serve to connect the various modes of transportation. Some examples include *transportation brokers*, *freight forwarders*, and *shippers associations*.

Transportation is part of both the inbound and outbound movement of goods. The inbound side is often referred to as *materials management*, which combines the activities of procurement, warehousing, transportation, inventory management, quality control, and scrap/disposal with the goal of improved customer service.^[4,5] *Physical distribution* focuses on the outbound side, dealing with the flow of the finished goods to the final customer.^[4] There may be several channel members in physical distribution of goods, including wholesalers and distributors.

Logistics management or *integrated logistics* management combines both inbound and outbound sides and can be defined as "the art and science of obtaining and distributing materials and products."^[6] Factors that contributed to the move toward an integrated logistics function and beyond are provided in the next section.

EVOLUTION OF LOGISTICS

The term *logistics* was derived from military activities related to the deployment and the support of armed forces during the times of war.^[7] In fact, logistics was referred to as early as 500 BC in Sun Tzu's *The Art of War*,^[8] a book about logistical activities and their relationship to war tactics and strategies. During World War II, logistical activities contributed to victory for the United States and its allies. More recently, logistical activities were identified as a critical

component of success for the United States during the Gulf War in 1990–1991. Involved was a massive movement of 122 million meals, 1.3 billion gal of fuel, and 31,800 t of mail.^[9] Since the 1960s, logistics has come to the forefront of best business practices.

During the 1960s and 1970s, the impact of global competition on U.S. firms was significant. Companies were losing market share and revenues to firms from other countries, particularly Japan and Germany. Competitive pressures increased in the mid-1970s as trade barriers were reduced, forcing U.S. firms to improve the quality of their products and services, to focus on higher levels of customer service and satisfaction, and to reduce costs. The amount spent on logistics was identified as the final frontier for cost reduction and improved customer service.^[10] Businesses turned their attention to the physical distribution of goods and materials for improved customer service and cost savings. They moved toward logistics management, the integration of transportation, distribution, warehousing, finished goods, inventory management, packaging, and materials handling. Deregulation of transportation played a large part as increased competition among carriers resulted in lower transportation rates and better management of inbound transportation.

Beginning in the late 1970s and continuing into the 1980s, the physical distribution management expanded to logistics management. This move led to purchasing's increased involvement in logistical decisions, including transportation management. Factors influencing this move included changes in the manufacturing environment and an increased emphasis on quality.^[11] The environment had changed as manufacturers moved to outsourcing a greater percentage of the content of their products to outside suppliers. These suppliers were able to produce parts at a lower cost in part because of their non-union environment. To replace the benefits of controlling all-part production, many firms moved to a just-in-time (JIT) manufacturing environment. In essence, JIT involved suppliers delivering parts to manufacturing sites just as needed for production and producing goods based on the actual customer need, rather than forecasts, resulting in lower inventory-carrying costs. However, it also meant relationships with both suppliers and transportation carriers needed to be stronger to assure the delivery within tight windows. In many instances, companies moved to single-sourcing arrangements. At the same time, the total quality management movement was in full force, enabling firms to compete with foreign goods producers. The selection of quality carriers was an important part of the equation because manufacturers had to be assured that the goods were moved safely and the customers were provided a proper level of service. Once the firms reduced the number of carriers in the pool, they focused on developing

measures of performance and monitoring carriers more closely. These measures included percentage of on-time shipments, percentage of damaged shipments, and percentage of complete orders shipped.

During the late 1980s and 1990s, the term *supply chain management* emerged and was defined as managing the total flow of goods and information from supplying agency to final “customer,” realizing that each step in the flow of goods should add a value.^[12] Potential savings within a supply chain were the driving factor behind this new trend. A supply chain has a series of “fixed points”—warehouses, distribution centers, offices, where goods are stored—linked together by transportation. This strategy includes developing partnerships between manufacturers, suppliers, and logistics-related members of the supply chain including transportation carriers. In particular, shippers spend more time qualifying their carriers. Carrier strategies that will lead to more successful relationships within the supply chain have been cited:^[13]

1. Partnerships in which both shipper and carrier challenge each other to provide innovative solutions.
2. Working jointly with shippers to control transportation costs.
3. Providing services considered the highest priority by the shipper.
4. Committing to continuous improvement processes that are critical to positive change.
5. Implementing satellite communication systems.
6. Providing information systems, logistics consulting, and specialized services based on customer needs.

TRANSPORTATION STATISTICS

Logistics costs as a percent of gross domestic product (GDP) have continued to drop with the advent of deregulation. At a high of 17.9% in the early 1980s, costs have averaged approximately 10% since 1992 and were 10.1% in 2000.^[14] In 2000, 5.9% of GDP was accounted for by freight transportation costs, the largest component of logistics costs, or \$590 billion in 2000. The most recent figures available indicate that federal, state, and local governments spend approximately \$129 billion annually on transportation.^[15]

Table 1 breaks down where organizations spend their logistics dollars.

CARRIER SELECTION

Selecting a transportation carrier can be one of the most crucial decisions a government purchaser can

Table 1 Logistics cost breakdown, 2000

Logistics cost	Cost (\$ Billion)
Carrying Costs	
Interest	\$95
Taxes, Insurance, Obsolescence, Depreciation	204
Warehousing	78
Subtotal	\$377
Transportation Costs	
Motor Carriers	
Truck—Intercity	323
Truck—Local	158
Railroads	36
Water (International, Domestic)	26
Oil Pipelines	9
Air (International, Domestic)	27
Forwarders	6
Subtotal	\$585
Shipper-related Costs	5
Logistics Administration	39
Subtotal	\$44
Total Logistics Costs	\$1,006

(From Ref.^[14].)

make for their organization. Just as choosing the right supplier of goods can help ensure the operational success, determining how those same goods are transported to their final destination is vital.^[16] While the initial choice made by an agency will be the type, or mode, of delivery, the choice of carriers will certainly be next.

When one contemplates how important quality deliveries are to an organization, they need only to look at the seven R's of customer satisfaction:^[17]

1. The right product.
2. Delivered to the right place.
3. At the right time.
4. In the right condition and packaging.
5. In the right quantity.
6. At the right cost.
7. To the right customer.

The importance of quality, accuracy, and timeliness in the delivery of goods is apparent. Logically, this places major importance on which carrier the purchasing practitioner selects to make such deliveries. With the importance of the carrier selection determined, let us examine how these decisions are best approached.

One of the first considerations a buyer must make in carrier selection is whether or not these services will be purchased separately. The predominant method in

the public sector is to shift the burden of carrier selection to the supplier.^[1] By specifying Free on Board (FOB) Destination in the bidding documents, the public agencies have charged the selected vendor with handling all the aspects of delivery, with the agency taking ownership upon delivery.

However, there are many instances where an agency can contract directly for transportation services. Many organizations ship out goods and equipment from their own facilities and hire private carriers to transport them. For instance, a water treatment facility may need to ship specialized control equipment to a manufacturer for repair or upgrade. To ensure quality transportation services of such critical equipment, they would likely have a contract with a carrier already in place. Even for incoming freight, this approach can make a lot of sense. Rather than rely on a carrier someone else has selected, a public agency may want to hire a carrier to transport that replacement compressor from the supplier's dock. By doing so, buyers are essentially stating that they are better equipped to determine the carrier than their supplier.

If the government purchaser does elect to contract for transportation services, they will face many of the same issues encountered with other types of services. A determination of the best procurement method will need to be made. An invitation to bid may be employed or possibly a request for proposal (RFP). A bid process will be driven by the lowest cost, while the RFP process will take into account the other aspects the agency determines as valuable. These can include service quality, corporate experience, financial stability, customer satisfaction, and prior dealings among other factors. According to the research by McGinnis,^[18] *service* is often more important than *cost* when selecting carriers.

As with other procurements, the opportunity for a contractor to have a large volume of guaranteed business by the agency will normally result in very favorable offers from the carrier community. Additionally, long-term agreements with multiyear options can provide an opportunity for a carrier to become a strategic partner of the public agency. Once a contract is executed and the business relationship is established, the agency has the ability to utilize the carrier whenever a transportation need arises. Similar arrangements can also be made with the third-party providers of transportation services. When an organization contracts with a freight forwarder or a broker, they can provide whatever type of transportation mode is needed, offering the buyer even more flexibility.

MODES OF TRANSPORTATION

There are five primary modes of transportation—motor carriage, rail, air, water, and pipeline. Each

mode provides its own unique capabilities as well as disadvantages. The following paragraphs provide a short analysis of each mode.

The dominant mode of transportation is motor carriage, or trucking, where approximately \$481 billion was spent in the United States in 2000 (Table 1), and they moved approximately one-half of all tonnage. The industry consists of a large number of small carriers, resulting in a high level of competition. The reasons for the heavy use of trucking include high availability, ability to provide service from and to almost any location, relatively fast transit times, and reliable service. Motor carriage has two roughly equal segments, for-hire and private carriage. Ninety-five percent of all for-hire carriage is moved by truckload, with the remaining 5% by less-than-truckload, or movements of less than 10,000 lb.^[19] The average truckload delivery spans 350 mi, with a market dominance of up to 750 mi. Trucks move raw materials, component parts, and finished goods. Trucking companies are actually a good customer of the rail industry, loading their trailers onto rail flatcars.

There are a small number of relatively large railroads, called Class I carriers, in the United States. As seen in Table 1, the railroads account for about 6% of the revenues generated by all modes, although they move approximately 25% of all the freight tonnage.^[20] Railroads carry primarily commodities including grain, metallic ores, crushed stone, and glass and clay products. Their advantage lies in moving goods long distances in large volumes with a low-cost structure. However, access is a significant disadvantage, resulting in the need for an additional mode of transportation such as trucks to complete deliveries. Another disadvantage is the long transit time as a result of consolidation and the transfer of boxcars at classification yards within the rail system. Shippers must also pay additional costs to maintain the safety of their goods. However, rail service is more reliable than other modes because it is mostly unaffected by poor weather conditions.

Water transportation also plays a significant role, accounting for \$26 billion of total logistics dollars spent within the United States (Table 1). This mode is frequently used by the organizations located in strategic areas to move raw materials, heavy items, and low-value bulk materials. Road salt is one example of a common government commodity often transported by water. The oldest technology of transport, water transportation includes barges used on inland waterways as well as ocean-going ships. The cost of the pickup and the delivery of goods to the dockside need to be included in the total cost of shipping, as they are not included in the water freight rates. Water carriers normally base their charges on either weight or

volume, depending on which one results in the highest revenue for the carrier.^[1] The federal government has long been involved with improving the waterways within the United States. The Army Corps of Engineers is the agency that provides oversight and funding for river and harbor improvements. Projects such as enlarging the locks and deepening the channels are examples of such improvements.^[3]

In 2000, a total of \$27 billion was spent on international and domestic airfreight. This transportation mode includes all shipments made by air cargo carriers. The movement of goods through the air has certainly increased in recent years, but the relatively high cost has kept its use somewhat limited in the government. As a general rule of thumb, the freight rates for air cargo are approximately twice as high as motor freight. However, along with the higher costs do come certain advantages. Because the air travel is generally much smoother than the ground freight, it is very attractive for transporting delicate equipment.^[1] Another advantage is the shorter delivery times involved with the air transport. A specialized repair part, for instance, can be shipped from New York to California in 1 day. Commercial companies that provide such service include the Federal Express and the United Parcel Service. When typical ground transportation would take several days, the decision to use airfreight is much more attractive. With such short lead times, organizations can normally reduce their inventories and carrying costs.

Referred to as the “hidden giant of American transportation,” pipelines have been around since 1865 when Samuel Van Syckel constructed a 5-mi oil pipeline in Pennsylvania.^[3] Besides its extensive use in the oil industry, pipelines are also used to transport other materials such as natural gas and coal slurry. The total transportation costs for oil pipelines alone was \$9 billion during 2000. Within the United States, there are over 110 regulated oil pipelines. From 1906 to 1977, the oil pipeline industry was governed by the ICC. Since then, it has been regulated by the Federal Energy Regulatory Commission. The dependability of this transportation mode is extremely high, as far as obtaining the product to its destination when promised. Because pipelines are virtually unaffected by the weather, they really are unparalleled when it comes to dependability. One disadvantage to the pipeline transportation is the high costs associated with installing and maintaining the piping systems. The material, labor, and equipment costs can be extensive, including such things as speciality piping, journey level labor, pumps, and computerized flow controls. However, the pipelines continue to be a viable means of transportation in the United States, particularly in the oil industry.

CONCLUSION

Transportation has played an important role in the development of the United States and thus our economy. City locations and their growth were often influenced by access to transportation. For example, Boston initially grew because of its origin as a port city; Chicago was a hub for the railroads; and Atlanta, initially a railroad stop, now has the world's largest airport.

Today, significant changes are occurring within the industry itself. A formerly heavily regulated industry now has the freedom to set prices, to determine where they want to operate and where they want to discontinue service, and to negotiate special services for customers. Cavinato^[21] suggests watching for certain trends in the next 10 years, including:

1. Continuing turnover of truckload motor carriers.
2. More rail mergers with few capacity changes.
3. Consolidation of courier and express mail firms.
4. Airline volatility, including bankruptcies and mergers.
5. More extensive use of electronic information to provide inventory visibility and efficient billing and payment.
6. Industry consolidation that allows more one-stop shopping.

Wood and Johnson^[3] also foresee an increase of environmental protection and safety laws. Additionally, recycling's impact on transportation will be significant because packaging choices will be based on reuse. Recycled materials and products mean additional hauls for transportation carriers. Finally, threats of terrorism will have a major impact on the international trade and the transportation carriers. The new challenge will be to reduce time and costs given increased security measures taken by the U.S. government. Given these trends, transportation and logistics will continue to be a dynamic and challenging part of the public procurement process.

REFERENCES

1. Corwin, S.H. *Intermediate Public Procurement*, 2nd Ed.; National Institute of Governmental Purchasing: Herndon, VA, 2000.
2. Stock, J.R. The maturing of transportation: an expanded role for freight carriers. *J. Bus. Logist.* **1988**, 9 (2), 15–31.
3. Wood, D.F.; Johnson, J.C. *Contemporary Transportation*; Prentice Hall: Upper Saddle River, NJ, 1996; p. 4.
4. Coyle, J.J.; Bardi, E.J.; Langley, C.J., Jr. *The Management of Business Logistics*, 6th Ed.; West Publishing Company: Minneapolis/St. Paul, MN, 1996.
5. Heizer, J.; Render, B. *Principles of Operations Management*; Prentice Hall: Upper Saddle River, NJ, 2001.
6. National Institute of Public Purchasing. In *Dictionary of Purchasing Terms*; National Institute of Public Purchasing: Herndon, VA, 1996.
7. Gourdin, K.N. *Global Logistics Management: A Competitive Advantage for the New Millennium*; Blackwell Publishers: Oxford, U.K., 2000.
8. Tzu, S.; Gagliardi, G. *The Art of War: In Sun Tzu's Own Words*; Clearbridge Publishing: Seattle, WA, 1999.
9. Pagonis, W.G. *Moving Mountains: Lessons in Leadership and Logistics from the Gulf War*; Harvard Business School Publishing: Boston, MA, 1992.
10. Heskett, J.L. Logistics—essential to strategy. *Harvard Bus. Rev.* **1977**, 55 (6), 84–95.
11. Gentry, J.J. *Purchasing's Involvement in Transportation Decision Making*; Center for Advanced Purchasing Studies: Tempe, AZ, 1991.
12. Ellram, L.M.; Cooper, M.C. Characteristics of supply chain management and the implications for purchasing and logistics strategy. *Int. J. Logist. Manag.* **1993**, 4 (2).
13. Wagner, W.B.; Frankel, R. Quality carriers: critical link in supply chain relationship development. *Int. J. Logist. Res. Appl.* **2000**, 3 (3), 245–257.
14. Cooke, J.A. Steering through the storm. *Logist. Manag.* July **2001**. Retrieved January 8, 2002, from <http://www.manufacturing.net/lm/index.asp>.
15. United States Department of Transportation, Bureau of Transportation Statistics. *Government Transportation Financial Statistics: Fiscal Years 1985–1995*; 2000. Retrieved January 8, 2002, from <http://www.bts.gov/ntda>.
16. Leenders, M.R.; Flynn, A.E. *Value-Driven Purchasing*; National Association of Purchasing Management and Irwin Professional Publishing: New York, NY, 1995.
17. Kuglin, F.A. *Customer Centered Supply Chain Management*; AMACOM, American Management Association: New York, NY, 1998.
18. McGinnis, M.A. The relative importance of cost and service in freight transportation choice: before and after deregulation. *Transp. J.* **1990**, 30 (1), 12–19.
19. WEFA-DR1a. In *Air Transportation Industry Yearbook 2001/2002*; WEFA-DR1a: New York, NY, 2001. Retrieved December 27, 2001, from Business Source Premier database.
20. WEFA-DR1b. *Railroad Transportation Industry Yearbook 2001–2002*; WEFA-DR1b: New York, NY, 2001. Retrieved December 27, 2001, from Business Source Premier database.
21. Cavinato, J.L. Buying transportation today. *Purch. Today* **2001**, 12, 16–18.

Malaysia

Noore Alam Siddiquee

*Department of Political Science, International Islamic University Malaysia,
Kuala Lumpur, Malaysia*

Habib Zafarullah

*School of Social Science, University of New England, Armidale,
New South Wales, Australia*

INTRODUCTION

With an area of nearly 330,000 km² and a population of 23 million, Malaysia is located at the heart of Southeast Asia. It is essentially a plural society where over 50% of the population are Malays and the rest are Chinese, Indians, and indigenous people. Politically, Malaysia is one of the most stable countries of the region. Yet despite the presence of multiparty democracy, Malaysia is often branded as a “semi-democracy”^[1–3] and its political system as “hegemonic.” A single party—the United Malays National Organization—has been dominating the political scene since independence in 1957. Economically, Malaysia has made impressive gains over the past decades; the economy is largely export-led, and it continues to enjoy high levels of growth notwithstanding the setbacks of the recent Asian financial crisis. With a human development index (HDI) ranking of 59 in 2002, Malaysia also fares well internationally in terms of other socioeconomic indicators. The economic success of the country is attributed, among other things, to political stability, dynamic and visionary leadership, and the efficiency of the administrative system in planning and managing programs of socioeconomic development.

THE GOVERNMENTAL SYSTEM

Malaysia is a constitutional monarchy within a federal and parliamentary structure. Of the 13 federating states, Selangor, Johor, Perak, Perlis, Pahang, Kedah, Terengganu, Kelantan, and Negri Sembilan are headed by hereditary *Sultans* who elect one of them as *Yang Di Pertuan Agung* (*Agung*), the supreme but nominal head of the federation for 5 years. The remaining states—Malacca, Penang, Sabah, and Sarawak—are headed by state governors appointed by the *Agung*. The Malaysian parliament consists of two chambers: a 219 member universally elected *Dewan Rakyat* (the lower house) with exclusive control over legislation and finance and a 69-member nominated *Dewan*

Negara (upper house) performing the tasks of scrutiny and review.

The federal government, with overarching control over the entire country, is responsible for key public tasks such as finance, foreign affairs, defense, internal security, trade, commerce and industry, health, education, and social welfare, whereas the state governments have jurisdiction over local government, agriculture, land, mines, forestry, Islamic matters, and Malay customs. Although the federating states share certain functions^a with the center, the balance of power is tilted heavily toward the federal government.

The Administrative Structure

The federal government is organized into ministries, departments, and statutory bodies. Currently, there are 25 ministries, 161 departments, and 75 statutory bodies through which the business of the federal government is transacted. Each ministry, headed by a minister, plays an important role in planning, coordinating, and implementing government policies and programs and has under its control a number of departments and statutory bodies which are primarily the implementing arms of the government. The secretary general, a career civil servant, as the administrative head of a ministry, assists and advises the minister and is responsible for the proper implementation of all policies and directives pertaining to the ministry. The departments are headed by directors-general and usually have their offices at the state and district levels.^[4,5]

The prime minister and his cabinet, responsible to the *Dewan Rakyat*, exercise executive authority. At the apex, the prime minister's department serves as the nerve center of government administration and is vested with considerable powers to coordinate government policymaking and implementation. Its chief

^aTwo eastern states of Sabah and Sarawak have been granted some additional powers relating to immigration and education.

secretary, the highest-ranking civil servant and principal advisor to the prime minister, is responsible for smooth and efficient implementation of the decisions of the cabinet. The other key central agencies include the Treasury within the Ministry of Finance, Public Service Department (PSD), Economic Planning Unit (EPU), Implementation and Coordination Unit, and Malaysian Administrative Modernization and Manpower Planning Unit, all under the prime minister's department.

At the state level, the governmental machinery replicates a similar fashion. The state executive council—the highest executive authority at this level—is headed by a *Menteri Besar* in states with hereditary *sultans* and a Chief Minister in other states. Although appointed by the federal government, he is required to contest and win in the state legislative assembly elections. The state secretariat is administered by the state secretary (the head of the civil service at the state level) who also acts as the chief advisor to the *Menteri Besar*/Chief Minister. Administratively, each state is divided into districts, *mukims* (subdistricts), and *kampungs* (villages). Although there are some additional levels in Sabah and Sarawak, districts play important roles throughout the country as most federal departments and agencies have their branches there, and almost all development projects planned at federal and state levels are implemented at this level.

The civil service in Malaysia has two broad divisions: the federal civil service and the state civil service. Although the federal civil service is organized into various cadres, the elite Malaysian Administrative and Diplomatic Service [*Perkhidmatan Tadbir dan Diplomati* (PTD)] is the most important. Officers belonging to PTD enjoy more power and prestige, occupying as they do most top policymaking positions at both federal and state levels.^[6] The civil service is also divided into three broad categories: top management group, professional and managerial group, and support staff. These groups are arranged vertically into 274 service schemes under 19 classifications based on similarities of roles and functions. Generally speaking, the service structure is rigid both vertically and horizontally^[5] with little or no scope for interservice or group mobility. This, apart from huge differentials between/among various categories in terms of pay and perks, has caused considerable tension and resentment in the civil service. Although some states have their own civil services, almost all senior positions at the state level are filled by PTD personnel and other centrally appointed bureaucrats.

The PSD serves as the central personnel agency, with jurisdiction on personnel matters of federal, state, and local governments, but an independent Public Service Commission has responsibility over recruitment and selection in the civil services. There are

another four specialized commissions for judicial and legal matters, railways, police, and education services. Civil servants in Malaysia are selected on the basis of merit and quota principles through a series of interviews, written examinations, and medical tests and placed on probation for 3 years before being confirmed and considered eligible for stipulated benefits and privileges.^[5]

REFORM AND MODERNIZATION EFFORTS

Since the early 1980s, the administrative system, in general, and the civil service system, in particular, have undergone significant changes. A series of reforms aimed at enhancing efficiency and effectiveness have been introduced. These have modified the structure of the public service, rules, regulations and procedures, personnel management, planning, budgeting, and financial management systems. The worldwide recession of the 1980s and the growing budget deficits caused by falling exports and public revenues prompted the government to reduce its involvement in the economy and hence “downsize” the public sector. Following global trends in public management reforms, Malaysia has undertaken major programs of privatization and deregulation. Such drives along with subsequent institutional restructuring have led to considerable reductions in the size of the public bureaucracy; yet it remains relatively large vis-à-vis the population it serves.

Early initiatives such as the introduction of time clock systems, name tags for employees, codes of ethics, manuals on office procedures, and quality control signaled the government's commitment to bureaucratic efficiency and responsiveness and clean and trustworthy government. Other moves sought to enhance civil service capacity and its relationship with the private sector and creation of clearance centers for providing swift and hassle-free services to potential investors seeking business approvals. Public service networks provide a range of services besides acting as one-stop bill payment centers. The Client's Charter, introduced in 1993 along the lines of the Citizen's Charter in the United Kingdom, pursues enhanced bureaucratic accountability and improved service delivery systems. The charter is an assurance by agencies that their outputs and services will comply with stipulated quality standards that conform to customer expectations. More recently, internationally recognized total quality management, ISO 9000 series, and benchmarking have been adopted in the public sector, and a comprehensive award system for quality and excellence has been introduced. The multimedia super corridor was established to promote the application of information communication technology (ICT) in society and implement a variety of schemes such as e-government and e-services

to enhance also the quality of public services. Although there is a gap between promise and performance, generally, reforms are seen favorably; they have helped improve the delivery of public services, thereby contributing to customer satisfaction.^[7]

PUBLIC ADMINISTRATION: TRENDS AND LIMITATIONS

The Malaysian administrative system has performed reasonably well compared with similar systems and traditions elsewhere in the region (except Singapore) and in other developing nations. Certain trends and limitations are perceptible, and these are explained below.

Centralization

Despite the arrangements for power sharing between the states and the center, the governmental system remains highly centralized. First, with exclusive control over all-important matters, the federal government plays a much bigger role than the state authorities. In the event of a conflict between the center and the states, federal laws have precedence. With its monopoly over resources and personnel, the central government exerts considerable influence on the states. Even with some autonomy in certain matters, the states rely on centrally deputed civil servants and federal grants and lack any degree of control on central agencies operating at the state or local level. Second, bureaucratic culture is characterized by a top-down approach to decision making, and there has been little or no attempt at delegating authority at lower levels. This has been reinforced by the concentration of power in central agencies. Prime minister's department symbolizes the concentration of extensive policy and managerial authority for planning and implementation of development activities at the state and local levels. The EPU serves as the center of national planning—conceiving and formulating major economic policies and strategies, including the 5-year plans. Even programs and projects initiated by individual agencies and departments need EPU approval and funding. Under such an arrangement, lower level agencies simply implement policies.

Dominant Executive

External control on administration has been weak in Malaysia mainly because of one-party hegemonic rule^[3] and a series of constitutional amendments it has made over the years increasing the power of the executive in relation to the other branches of the government.^[8] The dominance of the executive

increased dramatically since the launch of the New Economic Policy (NEP) in the early 1970s. The NEP was a two-pronged scheme that sought to correct the problems of economic imbalance among major ethnic groups through redistribution of wealth and assets in favor of *bumiputeras*. Hence a large number of quasi-public enterprises were created under the scheme giving the public bureaucracy an important role in economic planning, public policy initiatives, and the management of large sectors of the economy. Consequently, Malaysia produced one of the largest public bureaucracies among countries of comparable size and level of economic development. Notwithstanding the drives for privatization and deregulation, the bureaucracy continues to play an active role in terms of setting developmental goals, stimulating economic growth through the promotion of private sector, and distribution of incomes and wealth among various ethnic groups. Given its pervasive role in the society, Malaysia is often described as an “administrative state”^[9] where bureaucrats enjoy “positions and powers not to be found anywhere in the democratic world.”^[6]

Elitism and the Rule of Generalists

Public bureaucracy in Malaysia is characterized by elitism, a legacy of colonial administration. Members of the English-educated Malay aristocracy created by the British have continued to fill the highest positions in the bureaucracy since independence and enjoy extensive powers, privileges, and prestige.^[6] They monopolize federal and state administrative structures. The elite PTD constituents occupy important decision-making positions in the field as well and are also privileged to provide leadership in the public service as the directors-general of various ministries and departments invariably come from its ranks. The continuation of the quota system guaranteeing 80% of positions to the *bumiputeras* endows native Malays the opportunity to dominate the ranks of the elite service. This Malay prominence in the PTD has created a feeling of disenchantment among members of other services, further heightening the longstanding tensions in the civil service.^[5]

Politicization of Bureaucracy

An important feature of the public bureaucracy in contemporary Malaysia is the intimate relationship between politicians and bureaucrats. In theory, the Malaysian bureaucracy is governed by professional values and norms, and its members subscribe to the principle of “political neutrality;” in practice, the bureaucracy has never been separated from

politics.^[2,10] Although it is true that political leaders are generally capable of controlling the bureaucrats under their jurisdictions, senior bureaucrats share the same social and economic background with political leaders and play an important role in the policy process. Consequently, civil servants not only see themselves as the “paternal ruling group,” but also often get enmeshed in party politics which, in reality, makes the Malaysian civil service a “political bureaucracy.”^[11,12] Although Malaysian public servants no longer stand for party positions, the law that prohibits them from being involved in party politics is not uniformly enforced. In recent times, the government has been prompt at taking disciplinary action against those public servants who have apparently supported the opposition, but has been wittingly indifferent to those who actively back the ruling coalition. Such a policy has encouraged many civil servants, especially those who wish to build career in politics, to become passive members of the ruling party/coalition and support its policies and programs. Whereas bureaucrats, in general, are expected to be sympathetic to government policies and programs, those who work in the state-controlled electronic media or the Ministry of Rural Development and the National Civics Bureau have little choice but to virtually act as the agents of the ruling coalition.^[10]

Lack of Local Democracy and People's Participation

Local government in Malaysia is yet to emerge as a vibrant democratic institution. Presently, local agencies, established and regulated by state governments, are no more than subordinate offices of the state government. They suffer from limited capacity and resources and depend on federal and state grants to close their fiscal deficits. Furthermore, with local government elections remaining suspended since 1969, popular influence on local government and pressures for responsiveness is rather limited. In spite of administrative modernization, no significant drive has been mounted either to democratize governance at the grassroots or to rejuvenate local government bodies. Repeated overtures of the political opposition and civil society to reinstate a representative local government system have remained unheeded.

Likewise, public participation in governance has been lacking. Whereas impressive gains in eradicating poverty and promoting rapid socioeconomic development have been made, there has been no corresponding progress in creating opportunities for public participation. Reforms have fallen short of widening the presence of politically organized groups and civil society organizations in the policy process. As a result, the policy process continues to be insulated from political

pressures and influences from outside the public sector. Instead, the use of restrictive laws to curtail political participation and the absence of transparency and scrutiny of governmental affairs by a free media have not only undermined Malaysia's democratic image,^[1] but also thwarted the capacity of her citizenry to effectively influence public affairs. Without such capacities, the public at large remains simply passive recipients of government-directed services and facilities.

Corruption

The accountability system, despite changes brought about in various aspects of public management, has remained rather weak and inadequate.^[13] The incidence of bureaucratic malfeasance and corruption thus remains high. Corruption is believed to be rampant particularly at lower levels in the police service, road-transport department, immigration office, customs and excise office, and local government.^[14] Contracts are often awarded on the basis of political and kinship ties and kickbacks, and underhand financial dealings are common business approvals. The Malaysian experience with privatization shows the scale and magnitude of such corruption and patronage distribution.^[15] The anticorruption drives mounted by the government and the institutional mechanisms available to tackle the problem are yet to make any breakthrough in containing bureaucratic turpitude, and so Malaysia's position in the Transparency International's global corruption perception index has remained static over the past few years.

CONCLUSION

Public administration and the policy process in Malaysia operate within a federated pluralist political system with a dominant party playing the key role in all stages of the policy cycle. The prime minister and his cabinet determine policy agenda, make the major decisions to frame and adopt policies, negotiate legislation in parliament, and oversee their implementation and evaluate their impact. An elite band of bureaucrats, selected by a merit-based recruitment system possessing considerable power and prestige and strategically placed in key central agencies, provides support to the political executive in managing governance at the center. The state and local bureaucracies are also active in implementing development programs in their own jurisdictions.

A politically committed political leadership has succeeded in introducing innovations, akin to the advanced industrialized countries, in public management practices. New institutions and structures have been created, and existing ones have been rejuvenated to inject efficiency

in the delivery of public services. A range of reforms centered on economic rationalism as well as citizens' needs has enhanced the effectiveness of the administrative system. Deregulation and privatization have reconstructed the public sector for sound operation, and the application of ICT has simplified and improved communication both within government and between it and the private sector and citizens. Yet despite many positive features of these reforms, the administrative environment in Malaysia remains captive in enduring centralization, bureaucratic elitism, ineffective local governance, subdued public participation, bureaucratic corruption, and ineffective public accountability. These will need to be recalibrated, even if to a certain degree, to give the Malaysian administrative system a character correlating to the maxims of democratic governance.

REFERENCES

1. Case, W. Semi-democracy in malaysia—withstanding pressures for regime change. *Pac. Aff.* **1993**, 66 (2), 183–205.
2. Crouch, H. *Government and Society in Malaysia*; Cornell University: Ithaca, 1996.
3. Hilley, J. *Malaysia: Mahathirism, Hegemony and the New Opposition*; Zed Books: London, 2001.
4. Shafruddin, B.H. *The Federal Factor in the Government and Politics of Peninsular Malaysia*; Oxford University Press: Singapore, 1987.
5. Ahmad, A.S.; Mansor, N.; Ahmad, A.K. *The Malaysian Bureaucracy: Four Decades of Development*; Pearson: Kuala Lumpur, 2003.
6. Puthucheary, M. The administrative elite. In *Government and Politics of Malaysia*; Ahmad, Z.H., Ed.; Oxford University Press: Singapore, 1987; 94–110.
7. Siddiquee, N.A. Administrative reform in malaysia: recent trends and developments. *Asian J. Polit. Sci.* **2002**, 10 (2), 105–130.
8. Hai, L.H. Public administration—the effects of executive dominance. In *Democracy in Malaysia: Discourses and Practices*; Wah, F.L.K., Teik, K.B., Eds.; Curzon: London, 2002; 165–197.
9. Ahmad, Z.H. Introduction: history, structure and process in malaysian government and politics. In *Politics of Malaysia*; Ahmad, Z.H., Ed.; Oxford University Press: Singapore, 1987; 1–10.
10. Funston, J. Malaysia: developmental state challenged. In *Government and Politics in Southeast Asia*; Funston, J., Ed.; Institute of Southeast Asian Studies: Singapore, 2001; 160–202.
11. Scott, J.C. *Political Ideology in Malaysia*; Yale University Press: New Haven, CT, 1968.
12. Common, R. Malaysia: a case of business as usual. In *Governance and Public Sector Reform in Asia*; Cheung, A.B.L., Scott, I., Eds.; Routledge Curzon: London, 2003; 163–185.
13. See Chee, S. Public accountability in malaysia: form and substance. In *Public Administration in the 1990s: Challenges and Opportunities*; Pradhan, G.B.N., Reforma, M.A., Eds.; EROPA: Manila, 1991; 105–125.
14. *New Straits Times*; New Straits Times Press (M) Bhd.: Kuala Lumpur, 15 April, 2002.
15. See Gomez, E.; Jomo, S.K. *Malaysia's Political Economy: Politics, Patronage and Profits*; Cambridge University Press: Cambridge, 1997.

Management for Quality in Government

Michael E. Milakovich

Department of Political Science, University of Miami, Coral Gables, Florida, U.S.A.

INTRODUCTION

This article traces federal, state, and local government efforts to apply quality management theories and techniques since the reinvention era of the early 1990s. The principles and theories underlying quality management are a mix of various American and Japanese philosophies and strategies that are now widely practiced in many “reinvented” U.S. governmental agencies. Quality management was first applied in U.S. private industry during the 1980s, and is variously referred to as business process reengineering (BPR), companywide quality control (CWQC), total quality control (TQC), total quality management (TQM), and statistical process control (SPC). Most of these theories were based on applied knowledge of systems, empowerment of teams of employees, statistical process control of variation, and changes in organizational behavior.^[1–5] In some cases, the lack of complementary models for the public service sector led to overoptimistic expectations, experimentation, and disappointments. Despite uneven successes, a great deal has changed from a decade ago when there were few models to follow and serious questions about how committed governments were to quality management improvement.^[6] Improving the efficiency, quality, and productivity of government programs has taken on increasing urgency within a political framework of assigning accountability for results, reducing deficits, and streamlining federal executive agencies.

The application of TQM tools and techniques by federal, state, and local agencies responds to citizen demands for better service quality, improves government’s ability and reputation for effectively managing public problems, and provides a promising model for future “customer-responsive” public management practices. Specific projects and strategies to implement TQM and productivity improvement reflect our federal system of decentralized and divided authority, and differ substantially from county to county. Numerous state and local projects have been identified and recognized through formal competitions for quality awards (for examples, see the Malcolm Baldrige National Quality Award web site at <http://www.nist.gov>). They now provide a database of detailed guidelines and “best practices” for others to follow. The federal

government has implemented TQM, with nearly all agencies participating; state initiatives date from the early 1990s, with over two-thirds now involved; and about one-half the local governments (cities over 25,000) report customer service, quality improvement, or employee empowerment in at least one function. Many governments are taking customer service and quality management seriously by adopting a *process approach* to improve common areas, such as personnel administration, record keeping, vehicle maintenance, and community development.

WHAT IS TQM?

The importance of controlling the quality of industrial products was first recognized in the 1920s, when Walter Shewhart introduced SPC charts to monitor outputs in mass production manufacturing.^[7] At the Western Electric “Bell Labs” factory in Chicago during the 1930s, SPC techniques were expanded by Shewhart and his colleagues, including a young PhD in statistics from Yale, W. Edwards Deming. The techniques of SPC provided an efficient method for controlling the quality of mass-produced goods. They were expanded, perfected, and successfully applied to the mass production of weapons and war materials during World War II. In the boom years following the war, the lessons learned were largely forgotten. In the absence of economic competition, American industry leaders mistakenly believed that their products and management systems were superior. The theoretical work of several world leaders in the field of quality improvement paved the way for service professionals to apply these principles to their organizations.

In the early 1950s, Armand V. Feigenbaum^[8] coined the term TQC, and both W. Edwards Deming^[1,2] and Joseph M. Juran,^[4] among others, were invited by General Douglas MacArthur to teach SPC techniques to the Japanese during the occupation. Deming and Juran both met and influenced Kaoru Ishikawa,^[3] who became Japan’s foremost expert in CWQC. Phillip Crosby^[9] alerted the public to the importance of quality at a time when few skeptical American managers listened. Although greater numbers of private

manufacturing firms first succeeded in applying TQC in Japan, several Americans are internationally recognized as the intellectual founders of the concept later labeled TQM in the United States.

Total quality management is a continuous, cross-functional, interdisciplinary, and horizontally integrated management strategy applicable to all types of organizations. It is not an end in itself, but rather a *carefully designed and executed strategy for improving processes, products, and services through continual improvements in quality, reliability, systems, and performance*. Successful applications require an organizationwide commitment to rethinking existing production and delivery systems, not only from an *internal* systems view, but also from the perspective of *external* customers or users. When correctly applied, TQM is also a powerful, yet simple, method of process improvement for achieving valid customer quality requirements and productivity goals without the long-term need for substantial additional resources. It streamlines internal processes and merges process measurement characteristics (voice of the process) with customer satisfaction (voice of the customer) data. In the short-term, additional resources may be required for retraining, but when organizational changes and simple statistical management techniques are combined, TQM reduces internal competition, fosters teamwork, improves decision-making processes, reduces mistakes, and lowers operational costs.^[10]

Total quality management is a participatory organizational change strategy based on the idea that the greater the involvement an employee (or teams of “empowered” employees) has in determining and implementing organizational goals, the more committed those employees will be to achieving them. It encourages better monitoring of results by providing incentives for employees to increase the success of the whole enterprise, rather than offering rewards to individuals as performance incentives. In other sectors of the American economy, reliance on individual performance measures over total system and process improvement approaches has been linked to declines in both employee productivity and service quality. These approaches were discarded and replaced with newer, theory-driven, customer-focused strategies that achieve results without the need for additional long-term resources by changing management systems, enhancing the capacity for individual agencywide cooperation, and continuously improving processes.^[11] Efforts to improve quality and productivity are directly linked to concerns with efficiency in government, yet they draw on past experiences and encompass a broader range of global and national economic issues. Above all, quality management teaches managers and workers that organizational change is an achievable goal.

TQM IN THE PRIVATE AND PUBLIC SECTORS

In the 1980s, U.S. industry faced the grim prospect of greater international competition, lower profits, and shrinking global market share. Business leaders responded by “reinventing” the internal work environments of their corporations. They rediscovered the importance of all types of customers—internal and external—and began “empowering” employees with the authority, training, and trust needed to better serve customers. In the effort to regain competitive global economic market share, even the most disciplined firms found that demand for higher-quality goods and better customer services often exceeded the capacity of management systems and/or the motivation of employees to consistently “meet or exceed” customer expectations. Although several firms underestimated the discipline, time, and training required to change management procedures and empower employees, many embraced the new “quality paradigm.” As the quality message spread, more companies responded by relearning how to simultaneously be more competitive, quality focused, and customer driven.

The 1990s also witnessed a profound restructuring of many basic productive enterprises, in both manufacturing and service industries, resulting in acquisitions, downsizing, and mergers of companies within the United States and abroad. At the same time that TQM was being widely applied, a more aggressive movement emerged, urging the application of radical change to “reengineer” corporations.^[12] Many large multinational firms quite literally reengineered their business processes from the inside out using BPR, referred to by many as “TQM with a vengeance,” a short-lived management fad that was quickly seen as a transparent rationalization for downsizing and massive corporate lay-offs. The expanded use of the Internet, World Wide Web, and e-commerce has only accelerated these trends. Leading private sector firms mandated that *their suppliers* demonstrate commitment to BPR by meeting higher-quality standards and showing enhanced responsiveness to customer requirements. Many of them literally “broke up” into smaller, more manageable, decentralized service units to establish closer working relationships with customers and suppliers. Given the sheer size and scope of the effort, mistakes were inevitable.

In some instances, theories of quality improvement were misunderstood, misapplied, or not fully implemented.^[13] Unrealistic expectations prompted employee resistance, and controversial “quality bureaucracies” were created in many public and private organizations. Predictably, a “TQM backlash” fostered cynicism and dampened enthusiasm for further non-manufacturing applications. Public sector reforms are hampered in

the public sector by: 1) perceived negative experiences with private sector methods and models; 2) the long-term time frames of many TQM programs; 3) rigid management structures; and 4) a lack of theory-driven research on how to empower, motivate, and reward public employees for systems changes and customer service quality. One observer argued that so-called “pure” quality initiatives were ill suited for public sector organizations, citing four key limitations: 1) defining the customers of government is ambiguous; 2) public administrators are service rather than product oriented; 3) public agencies are input rather than output oriented; and 4) politics works against long-term leadership and constancy of mission.^[14] Others suggested that political leadership is necessary to achieve any change, and that governments are primarily small service organizations that can (and must) be responsive to customers.^[15,16] Despite some resistance, TQM was used extensively during the 1990s to improve a wide range of federal executive agencies, public utilities, and state and local governments. In addition to extensive federal reform efforts, many states and local governments replaced traditional bureaucratic hierarchies with “networked” structures, empowered employees, and redesigned performance management systems to better serve customers.

CONGRESS, TQM, AND THE NPR

Since 1988, TQM has been the official management improvement system for all federal executive agencies. The U.S. Office of Management and Budget coordinates joint public and private governmentwide quality improvement initiatives in the executive branch. Congress supported federal executive branch reform with budget and paperwork reduction agreements, spending limits, mandated use of results-oriented management systems, and the creation of a national award for quality improvement. In addition, management reforms, such as the Chief Financial Officers (CFOs) Act of 1990 and the Government Performance and Results Act of 1993, reinforced the federal TQM reinvention strategy. The National Performance Review (NPR) encouraged federal agencies to implement TQM to increase efficiency, better serve customers, develop outcome-oriented performance measures, monitor results, and publicly post key performance measures.

The NPR (1993–2001) challenged federal agencies to cut red tape, downsize, deregulate, and provide customer service equal to “the best in business.” Inspired by the best-selling book *Reinventing Government*,^[17] the NPR was a controversial mix of theory and ideology that focused on “market-driven”

solutions to the perceived inefficiencies in the delivery of government services. Specific, and often contradictory, recommendations included putting customers first, decentralizing decisions, empowering front-line employees, reducing administrative costs, and focusing on results. After nearly a decade of progress, comparative measures of federal performance are now available and CFOs in all federal agencies must provide systematic financial reports to Congress and to the president. The ultimate success of the federal government’s aggressive drive for quality management and reinvention depends on the soundness of its theories and how well it is implemented in the field.^[18,19]

REWARDING QUALITY IMPROVEMENT EFFORTS

Recognition and award programs are also used to reward quality management efforts and acknowledge employees for exemplary customer service. Although it is sometimes difficult to respond to all customers of public organizations, quality awards can demonstrate how governments in *specific* service areas are better serving customers. Governments can use quality awards to reinforce these goals, reward success, and provide objective benchmarks to help guide the transition to a customer-focused government. Awards, benchmarking, and charter marks been used successfully by many governments to compare results, achieve policy goals, highlight achievements, and guide customer service quality improvement efforts.^[20] There are now hundreds of national, state, and local quality awards that are a largely untapped, but increasingly rich, reservoir of detailed information about how various types of government agencies can successfully change management processes to effectively respond to customers. Cities, counties, and states are benchmarking the best practices from leaders in various processes. In this way, management improvements are being made on the basis of experience in other, similar jurisdictions. At the local level, public agencies are learning from each other how to best respond to the needs of their customers. Many other public agencies, non-profit organizations, and service organizations are also involved as co-partners. Forty-five states have established state quality awards patterned after the federal government’s Baldrige Awards, and many localities have also launched successful customer-driven quality improvement drives. Research findings generated by successful applications offer administrators a clearer understanding of their customers’ needs and how they can empower employees to meet them.

FUTURE OF QUALITY MANAGEMENT IN GOVERNMENT

Public organizations have now increased their capacity to measure performance, increase productivity, improve customer service, and reward quality management efforts. Links between the executive and legislative branches have been established to achieve greater accountability, efficiency, and productivity with the adoption of new management systems. The principles of TQM are widely practiced in many reinvented agencies, and numerous quality awards provide “benchmarks” for others to follow. Models for reform were based on earlier quality improvement systems initiated by the private manufacturing sector and thus not easily converted to the public service sector. Many public agencies learned the hard way that it is easier to start a productivity improvement initiative than it is to maintain one. The reinventors were accused of going too far in changing government management procedures to “run government like a business” and many public managers were criticized for resisting attempts of improve performance out of fear that their positions might be threatened. The Bush administration (at least prior to September 11, 2001) clearly signaled a preference for the use of non-governmental means to achieve the same ends: public-private partnerships; tax cuts; privatization; and greater use of charter schools, faith-based organizations, and vouchers.

The application of quality management principles in government reflects the century-old debate about the “proper” relationship between private interests, politics, policy, and administration. Which models—public, private, non-profit, or some combination thereof—best reflect the public interest in turbulent times? The future direction of the quality movement in government is now at a crossroads: at issue is whether the incremental publicly oriented efforts of the last decade will continue or be replaced altogether with contracting out, vouchers, privatization, and partnerships with non-governmental organizations.

REFERENCES

1. Deming, W.E. *Out of the Crisis*; M.I.T. Center for Advanced Engineering Study: Cambridge, MA, 1986.

2. Deming, W.E. *The New Economics of Industry, Government and Education*; The MIT Center for Advanced Engineering Study: Cambridge, MA, 1993.
3. Ishikawa, K. *What is Total Quality Control? The Japanese Way*; Prentice-Hall: Englewood Cliffs, NJ, 1985.
4. Juran, J.; Blanton, G. *Juran's Quality Handbook*, 5th Ed.; MacGraw-Hill: New York, 1999.
5. Scherkenbach, W.W. *Deming's Road to Continual Improvement*; SPC Press: Knoxville, TN, 1991.
6. Milakovich, M. Total quality management for public sector productivity improvement. In *Public Productivity Handbook*; Holzer, M., Ed.; Marcel Dekker: New York, 1992; 577–602.
7. Shewhart, W. *Economic Control of Quality of Manufactured Product*; Van Nostrand Reinhold: New York, 1931.
8. Feigenbaum, A.V. *Total Quality Control*, 3rd Ed.; McGraw-Hill: New York, 1983.
9. Crosby, P. *Quality is Free: The Art of Making Quality Certain*; Mentor: New York, 1979.
10. Milakovich, M. Total quality management and public sector productivity improvement. *Public Prod. Manage. Rev.* **1990**, XIV (1), 19–32.
11. Milakovich, M. *Improving Service Quality: Achieving High Performance in the Public and Private Sectors*; St. Lucie Press: Delray Beach, FL, 1995.
12. Hammer, M.; Champy, J. *Reengineering the Corporation*; Harper Business: New York, 1993.
13. Cole, R. *The Death and Life of the American Quality Movement*; Jossey-Bass: San Francisco, 1993.
14. Swiss, J. Adapting total quality management (TQM) to government. *Public Adm. Rev.* **1992**, 52 (5), 356–362.
15. Rago, W. Adapting total quality management (TQM) to government: Another point of view. *Public Adm. Rev.* **1994**, 54 (1), 61–65.
16. Sensenbrenner, J. Quality comes to city hall. *Harvard Bus. Rev.* **1991**, 69 (2), 64–70.
17. Osborne, D.; Gaebler, T. *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*; Addison-Wesley: Reading, MA, 1992.
18. Beam, G. *Quality Public Management: What It Is and How It Can be Improved and Advanced*; Burnham Publishers: Chicago, IL, 2001.
19. Osborne, D.; Plastrik, P. *The Reinventor's Fieldbook*; Jossey-Bass: San Francisco, CA, 2000.
20. Keehley, P.; Medlin, S.; MacBride, S.; Longmore, L. *Benchmarking for Best Practices in the Public Sector*; Jossey-Bass: San Francisco, CA, 1997.

Managing Mutual Commitments and Expectations: Contract Administration in Public Labor Relations

D. S. Chauhan

Department of Political Science, Bowling Green State University,
Bowling Green, Ohio, U.S.A.

INTRODUCTION

The collective bargaining process is governed by a set of regulatory rules, legislative policy, body of judicial decisions or opinions, and a contract mutually agreed upon by the management and the employee union. A contract establishes a set of rules, commitments, and expectations legally binding on both the labor and management. As it emerges from the negotiating process, a contract or a collective bargaining agreement is a mutually agreed upon set of rules that have been jointly developed to guide an employer–employee relationship for a fixed period of time. Because collective bargaining represents an adversary relationship involving the representatives of two groups with divergent goals, the emerging rules are likely to represent numerous compromises by both parties in dealing with contract negotiation or administration. These rules provide the framework to govern day-to-day relationships between the two parties.

Upon the ratification of the contract by the union, the labor relation's process enters the contract administration phase, where labor and management have to abide by the provisions of the agreement under the supervision of the management and the vigilance of the union Steward. Because the contract administration is a bilateral process, it recognizes the goodwill, cooperation and understanding of labor and management to build a successful labor relation's program through shared responsibilities in day-to-day administration of the contract.^[1]

NATURE AND SCOPE OF CONTRACT PROVISIONS

The length, subject matter, and details of an “agreement” or “contract” represent the complexity and unique aspect of collective bargaining relationships evolved over a long period of time. For instance, the length of contract and the complexity of contract provisions may differ in well-established contractual relationships and the newly evolving relationship. Specificity of provisions may cover 20 to 100 pages. Also, the subject matter and content of contractual provisions may differ in various contracts. Most

contracts, however, contain a preamble highlighting the purpose and importance of cooperative leadership, mutual trust, and shared responsibilities in promoting effective contract administration. For instance, the agreement reached between Kent State University and the faculty union states, “The parties to this agreement desire to establish and maintain an atmosphere of mutual understanding and cooperation . . . The parties endorse the university's commitment to the achievement of optimal conditions of intellectual discovery, human development and responsible social changes.”^[2]

The substance of most contract provisions can be divided into three major categories:^[3]

1. *Fixed provisions*: Fixed provisions in majority of contracts pertain to wage rates, union security, and the date of enforcement and reopening of new contract negotiations.^a
2. *Contingent provisions*: Contingent provisions cover lay-off regulations, promotions, transfers/reassignments, and severance and discharge.^b
3. *Dispute resolution provisions*: Because the collective bargaining is an adversarial process, all contracts contain dispute resolution provisions focusing on mediation, fact-finding, and arbitration procedures to resolve collective conflicts involving contract negotiations, interpretations, and the fulfillment of mutual commitments and expectations.^c

The other important provisions include grievance procedures, management rights clauses, and discharge procedures.

^aFixed provisions normally include contract duration, the bargaining unit, union recognition and rights, management rights, wages, hours of work, benefit plans, union security provisions, antidiscrimination clauses, residency requirements, union political activities, official time, and reopening clause.

^bContingent provisions include discharges, layoffs/RIF procedures, recall of laid-off workers, promotions, work scheduling, work assignments, transfers, and discipline.

^cConflict resolution provisions include no-strike provision, multistep grievance procedures, ULPs, contract interpretation clauses, and the use of mediation, fact-finding, and arbitration procedures.

CONTRACT ADMINISTRATION PROCESS

When the contract is agreed upon by the management and ratified by the union, it must be enforced by the management through consultative leadership with the employee union and the Steward. A contract is no better than its administration. Because managers in most public organizations have the responsibility for supervising personnel, organizing and directing the operations of the agency, and controlling the utilization of human and material resources, they undertake the primary responsibility for contract administration. Contract administration is not separate, but it is an integral part of the labor relations process providing an opportunity for cooperation and resolution of conflicts and differences through effective contract interpretation.

Because the labor agreement covers a wide range of activities in the collective bargaining process, there is a paramount need for some degree of conformity and consistency in the administration of the contract. A good contract administration also requires the development of guidelines and policies that help management to pursue a systematic approach in dealing with personnel problems, defining procedures with mutual agreement, and identifying the role of both parties. Resolution of grievances as well as the satisfaction of the union membership must be timely. Communication support among managers and between the management and the union is also vital.^[4]

MAJOR STRUCTURAL PROBLEMS

As discussed earlier, a contract covers a wide variety of activities, expectations, and a set of procedures to deal with emerging problems experienced during the process of execution. Although the primary responsibility for the execution or the administration of the agreement/contract lies with the management, the union's Steward oversees that all the provisions of the contract are being applied uniformly and consistently to guarantee the equal application of the agreement. Although the totality of the contract's execution is essential, the following are the components or processes that characterize the major elements in the contract administration process:

1. *Role of supervisors:* Although the management is responsible for ensuring effective execution of the contract, supervisors are considered to be the key in providing linkage between the management and the rank-and-file workers. These supervisors work in close cooperation with the union Steward, because they possess an intimate knowledge of work activities and problems encountered by the management and workers. Studies of the Steward function and relationships with

supervisors reveal that the relationship between the two, to a large extent, has been a positive factor in improving supervisory skills and facilitating equitable treatment of employees in personnel policies and practices.^[5]

2. *The grievance procedure:* During the course of negotiations or contract administration, differences or disagreements on a variety of issues may arise between the management (supervisors) and union Steward. Such differences or disagreements may lead to threats of strike and must be resolved through effective interpretation of related contractual provisions in a timely manner. One important aspect of contract administration is the grievance procedure that is aimed at resolving the labor-management differences or disagreements. All contracts contain a formal grievance procedure authorized by the legislative statute or agreed upon by the parties as part of the contract. The enabling legislation or a negotiated contract can also specify the nature of issues that the employees may be able to grieve. However, the civil service regulations can constrain employees to grieve on such issues as discipline and discharge, just cause, management rights, past practice, overtime, reinstatement/reimbursement from suspension, leaves of absence, promotion, union security, and disciplinary policies and procedures.^[6] If the union believes that the employer is not complying with the agreement, it often resorts to enforce its position by the same show of force that won those terms in the first place.^[7]

The grievance procedures/appeal systems in the public sector are mandated by law or are agreed upon through contract provisions. They typically involve three to five steps^d with the informal discussion or commencement of the grievance process finally requiring a written grievance leading to arbitration.^[8]

3. *Unfair Labor Practices (ULPs):* At any point in the organizing process or during the process of collective bargaining, the administration or the union or its bargaining agent may engage in some activity which may be considered unfair or may have negative effects on the outcome of either organizing efforts, collective bargaining

^dA typical formal grievance procedure may include the following steps: Employee initiates the grievance: 1) Employee discusses grievances or problem orally with supervisor or union Steward; 2) Grievance is put in writing and submitted to supervisor or other line manager; 3) Grievance is appealed to upper management or labor officer; and 4) Union decides on whether to appeal unsolved grievance to arbitration for a binding decision. Some grievance procedures may contain up to six or seven steps.

or contract administration may constitute an unfair labor practice. The employee union collectively or jointly with the individual may also grieve against any management practices or procedures that are considered by the union as unfair labor practice. A grievance can be filed by either party involving interference, domination, discrimination, discharge, refusal to bargain, or repeated failure to process grievance, and cause the union or employer to violate provisions or prohibitions imposed by the law or a negotiated contract.^e

4. *Strike threats and dispute resolution:* The major disagreements associated with interpretation, negotiations, and collective bargaining may sometimes lead to work stoppages or strike threats during the process of contract administration. Such job actions may invoke the use of such formal procedures as mediation, fact-finding, and arbitration involving neutral third

parties. Recent developments in collective conflict resolution, through building arbitration of labor-management disputes, are considered to bring some degree of finality. However, the preferable strategy of resolving disputes should be based on cooperation and not on confrontation.

In sum, important structural problems in contract administration involve infringement upon management rights, seniority and promotion qualifications, statutory limitations on grievance procedures, civil service rules limiting subjects that can be discussed in collective bargaining, and emphasis placed on the importance of past practices. Similarly, the key elements or components in contract administration focus on grievance procedure/arbitration, role of the upper management and supervisor as a primary line, role of the Steward and union leadership, management rights, interpretation of the contract language, the personnel offices as a clearinghouse, and disciplinary procedures.

^eAn ULP may be committed by the public employer or the employee organization. They may include the following:

1. Public employers/agents or their representatives may not:
 - a. Interfere, restrain, or coerce employees.
 - b. Initiate, create, or dominate in the conduct of union business.
 - c. Discriminate in hiring, layoff, and advancement of employees.
 - d. Discharge an employee for filing a grievance.
 - e. Refuse to bargain with recognized union bargaining agent.
 - f. Fail to process grievances.
 - g. Lock out or prevent employees from performing their legal duties.
 - h. Cause or attempt to cause an employee organization to violate prohibitions.
2. Employee organizations/agents or their representatives may not:
 - a. Restrain or coerce employees in exercise of their rights.
 - b. Cause or attempt to cause an employer to violate legal or contractual obligations.
 - c. Refuse bargain collectively with a public employer.
 - d. Call, initiate, maintain, or conduct a boycott against any public employer or picket any place of business of a public employer, or on account of jurisdictional dispute.
 - e. Induce or encourage any individual employed by any person to engage in a strike or work stoppage that is prohibited.
 - f. Fail to fairly represent all public employees in the bargaining unit.
 - g. Induce or encourage any individual in connection with a labor relation's dispute simply to picket at the residence of a public official.
 - h. Engage in picketing, striking, or other concerted refusal to work without written notice to the public employer or employee relation's board.

OBJECTIVES OF POSITIVE CONTRACT ADMINISTRATION

The success of contract administration will ultimately determine if all the time and energy devoted to the formulation of the collective agreement will be justified by subsequent developments.^[9] The collective bargaining agreement must be administered in a manner to:

1. Provide a basis for speedy and judicious review and resolution of labor-management differences over interpretation of the terms of government.
2. Provide the organization with information about the location of sensitive labor-management areas and about inadequacies of a particular supervisor.
3. Provide employees assurance that they will receive fair and just treatment by adherence to due process in dealing with employee problems.
4. Provide the organization with a mechanism to adjust virtually all day-to-day differences between the labor and management.
5. Provide a mechanism for identifying management employee problems not covered by the agreement.
6. Provide management officials and union officers an opportunity to meet separately with the supervisory workforce and the shop Steward to review the contract provisions.
7. Provide for an interpretation of the contract by a labor officer or personnel director or any

other designated person agreed upon by the labor and management.^[10]

The success of a labor-management relations program largely depends on the effective functioning of the cooperative leadership process in which the management, the union, and the employee groups work together toward the effective execution of the contract. In 1991, the U.S. General Accounting Office conducted a study of the federal labor relations program and reported that “a cooperative or joint problem-solving orientation was lacking... It also noted that labor-management cooperative programs in private sector reflect the growing view that an (us versus them) approach is outdated and unworkable.”^[11] Consequently, in 1993, President Clinton issued an Executive Order 12871 calling for a change in the labor-management relations “So that managers, employees, and employees’ elected union representatives serve as partners... It established a National Partnership Council to advise the President on labor-management relations, to support and promote labor-management partnership, and to recommend statutory changes needed to support a cooperative and reform oriented approach to federal labor relations.”^[12]

Achievement of goals of the Executive Order 12871 are hard to measure for a variety of reasons. However, the traditionally adversarial culture of federal labor relations was replaced during the Clinton Administration by a cooperative culture that encouraged mutual respect, partnership, and goal-oriented cooperation among federal agency heads.^[13] The joint labor-management committees (LMCs) at the state level have also emphasized to provide a mechanism for cooperative efforts to resolve a variety of problems associated with workplace safety, quality control, and communication. The major cities, such as Phoenix and San Diego, as well as the states of Ohio and Massachusetts have negotiated agreements establishing LMCs.^[14]

CONCLUSION

The impact of collective bargaining on personnel management systems in public management has produced a need for more collaborative or cooperative efforts in public affairs aimed at effective management and provision of goods and services to the people. The changing organizational culture and work environment are influenced by family and diversity issues, communication task management technologies, changing employee attitudes, behaviors, and values, and the need for participative management. The issues that may significantly influence the nature of employer-employee relationship, the bargaining powers, and the contract

administration in the twenty-first century may include new realities about domestic partners, day care, elder care, flexible scheduling, drug testing, job sharing, flexible staffing, flexible benefits, flexible leaves, and flexible location. The changing workplace would require more positive attitudes to facilitate effective communication, and exchange of information in good faith to facilitate labor-management cooperation.

Increased unionization and collective bargaining in the public sector during the past three decades have led toward the centralization of public management structures and control systems.^[15] Public managers are forced to take a proactive role in providing cooperative leadership in public labor relations producing significant decline in the legislative influence over personnel matters in jurisdictions where collective bargaining and the practice of contract administration are well developed.^[16]

Disagreements in contract’s administration would be largely determined by the size and complexity of the organizations, contractual arrangements, working relationships between the management and union leadership and members, managerial attitudes and willingness to work together, changing organizational culture, and informational technology and task management. Living with the contractual agreements would require greater harmony or precedence of contractual obligations over the traditional “rules of the workplace” characterized by civil service rules and the merit system.

REFERENCES

1. Nigro, L.G.; Nigro, F.A. *The New Public Personnel Administration*; F.E. Peacock Publishers: Ithaca, IL, 2000; 233–237.
2. Kent State University. *Collective Bargaining Agreement*; Kent State University and AAUP, Kent State Chapter: Kent, OH, September 16, 1989.
3. Craft, J.A. Notes on administration of collective bargaining agreements. *Personnel Administration and Public Personnel Review* July–August, 1972, 1, 30–33.
4. Levine, M.J.; Hagburg, E.C. *Public Sector Labor Relations*; West Publishing Company: St. Paul, MN, 1979; 85–88.
5. Sulzner, G.T. *Impact of Labor-Management Relations Upon Selected Federal Personnel Policies and Practices*; US Office of Personnel Management: Washington, DC, 1986.
6. Coleman, C.J. *Managing Labor Relations in the Public Sector*; Jossey-Bass Publishers: San Francisco, 1990; 142–147.
7. Kearney, R. *Labor Relations in the Public Sector*; Marcel-Dekker: New York, 1984; 290 pp.
8. Repas, B. *Contract Administration*; Bureau of National Affairs: Washington, DC, 1984; 69–76.

9. Paterson, L.T.; Murphy, R.T. *The Public Administrator's Grievance Arbitration Handbook*; Longman: New York, 1983; 64–67.
10. Levine, M.J.; Hagburg, E.C. *Public Sector Labor Relations*; West Publishing Company: St. Paul, MN, 1979; 84–88.
11. U.S. General Accounting Office. *Federal Labor Relations: A System in Need of Reform*; U.S. Government Printing Office: Washington, DC, 1991.
12. Nigro, L.G.; Nigro, F.A. *The New Public Personnel Administration*; F.E. Peacock Publishers: Ithaca, IL, 2000; 233–237.
13. Reeves, T.Z. Labor-Management partnerships in the public sector. In *Public Personnel Management: Current Concerns, Future Challenges*; Ban, C., Reecucci, N.M., Eds.; Longman: New York, 1997; 173–186.
14. Dawson, I.O. Trends and developments in public sector unions. In *Public Personnel Administration: Problems and Prospects*, 2nd Ed.; Hays, S., Kearny, R.C., Eds.; Prentice-Hall: Englewood Cliffs, NJ, 1990; 153–157.
15. Kerney, R.C. Monetary Impacts of Collective Bargaining. In *Handbook of Public Sector Labor Relations*; Rabin, J., Vocino, T., Hildreth, W.B., Miller, G., Eds.; Marcel Dekker: New York, 1983; 368–402.
16. Nigro, L.G.; Nigro, F.A. *The New Public Personnel Administration*; F.E. Peacock Publishers: Ithaca, IL, 2000; 230–237.

Market Failure, Theory of

Brian E. Dollery

*School of Economics, University of New England, Armidale,
New South Wales, Australia*

Joe L. Wallis

*Department of Economics, University of Otago, Dunedin,
New Zealand*

INTRODUCTION

The appropriate role of government in contemporary advanced industrial democracies is a complex and controversial issue that remains unsettled. A vast research effort has been devoted to resolving this issue. Social scientists, including anthropologists, economists, policy analysts, political scientists, public administration specialists, and sociologists, have devised numerous approaches to the study of the nature and role of government, with varying degrees of success. One of the more successful approaches to the analysis of the state has been developed by welfare economists in the form of the theory of market failure. In essence, the market failure paradigm examines the operation of the economy and prescribes government intervention when markets “fail” on the grounds of either economic efficiency or equity.

Although the origins of the theory of market failure may be traced back several hundred years in the history of economic thought, the contemporary version of this theory owes much to William Baumol’s^[1] *Welfare Economics and the Theory of the State*; Paul Samuelson’s^[2] article entitled “The Pure Theory of Public Expenditure” in the *Review of Economics and Statistics*; and Francis Bator’s^[3] article entitled “The Anatomy of Market Failure.” Notwithstanding the success of the market failure paradigm in providing a rigorous conceptual framework to guide policy makers in the design of public policies, from the beginning it contained several serious limitations. Mitchell and Simmons^[4] have described these problems as follows:

Because no human institution is perfect, it is easy to find imperfections or “failures.” It is even easier if, like the welfare economists, you only concentrate on the economy and do not recognize the effects of the political system on the shape, direction, and rules of the economy. It is also easy to call for government intervention if you have no corresponding theory of how government functions. (p. xvii)

ADAM SMITH’S “INVISIBLE HAND”

The notion that “the pursuit of rational self-interest by individuals engaged in the exchange of property rights through market institutions results in socially benevolent outcomes” represents one of the most important insights of economics, and can be traced back to Adam Smith’s famous doctrine of the “invisible hand” in the *Wealth of Nations* in 1776. In essence, this argument holds that maximizing behavior by individual economic agents in market relationships generates a socially rational use of scarce resources under certain defined conditions. Moreover, the voluntary exchange of goods and services through the market mechanism is itself a positive sum game for all participants because exchange enhances mutual welfare. Because the behavior of *homo economicus* leads automatically to Pareto optimality in consumption and production, this eliminates the necessity for active policy intervention, and thus creates a strong case for the role of government to be limited to the definition and enforcement of property rights.

Since Arthur Pigou’s momentous *Economics of Welfare* in 1920,^[5] modern welfare economics has been largely concerned with the development and refinement of the conditions necessary for the effective operation of the invisible hand—that is, the underlying conditions that must be met for a perfectly competitive or decentralized system of price determination to efficiently allocate scarce resources amongst alternative ends. The discovery of these necessary and sufficient conditions for economic efficiency led to the systematic identification of generic instances where markets “failed” to produce allocatively efficient results. In essence, the existence or absence of several factors can prevent the rational self-interest in exchange or market processes characteristic of the invisible hand from generating socially desirable outcomes. This phenomenon is market failure and provides the intellectual basis for extensive government intervention aimed at achieving economic efficiency in market

economies. Market failure in this sense refers to the inability of a market or system of markets to provide goods and services either at all or in an economically optimal manner. Market failure is thus defined exclusively in terms of economic efficiency in general, and allocative efficiency in particular. The market failure paradigm for microeconomic policy making assumes that governments intervene in the public interest to restore economic efficiency. The market failure paradigm is thus a normative theory of government intervention insofar as it prescribes how governments “should” behave. As we discuss, the market failure paradigm can be extended to include distributional or equity elements.

SOURCES OF MARKET FAILURE

Economists have identified numerous sources of market failure, and indeed instances of real or perceived market failure are probably limited only by the imaginations of the economists. However, textbooks typically emphasize six main forms of market failure, which we review below.

First, for Adam Smith’s invisible hand to operate properly, markets must be competitive rather than monopolistic or oligopolistic. Various reasons exist for the absence of competition in some defined market. Geographic factors such as large distances or isolated locations can mean limited competition. For example, the vast outback of Australia contains numerous small communities often serviced by a single retail supplier. Similarly, limited ownership of some natural resource can confer monopoly power on a producer. Governments may often create monopolies through the legal system. For instance, patent laws grant monopoly rights to inventors for specified time periods. Similarly, gaming operators are usually given exclusive control over casinos in particular jurisdictions. However, perhaps the most significant in the present context are barriers to entry into an industry, which arise from increasing returns to scale. This necessarily implied decreasing average costs over large volumes of output and the potential for natural monopolies to arise. Because of the technological nature of production in industries of this kind, competition simply cannot exist. This source of market failure is most commonly evident in the provision of services such as electricity, water, railway networks, and post offices.

Traditionally, government policies toward competition as a potential source of market failure have taken two general forms. First, laws have been promulgated that seek to prevent collusion and other defined types of anticompetitive conduct among producer groups. Second, public policies have sought to deal with the problems associated with natural monopoly

by either nationalizing the industry in question (historically, the preferred approach in western Europe), or by regulation, especially rate of return regulation (the policy instrument most favored in the United States).

A second source of market failure resides in the existence of positive or negative externalities involving an interdependence between consumption and/or production activities, and resulting in a divergence between private and social costs and benefits. In essence, the problem posed by externalities is that the resource allocation yielded by markets will not be efficient because market prices do not reflect the full costs and benefits involved, and accordingly will not generate socially efficient levels of consumption and production. Literally, thousands of kinds of externalities exist. Perhaps industrial pollution is the most widely cited example of a negative production externality. By contrast, industrial agglomeration along the lines of Silicon Valley represents a positive production externality where separate economic activities reinforce each other and lower production costs. A myriad of consumption externalities can be readily identified. Drinkers in a bar enjoying listening to a jukebox constitutes a positive consumption externality, whereas tired neighbors attempting to sleep alongside a rowdy party signifies a negative consumption externality!

Conventional policy responses to externalities have almost always resulted in government intervention. Two generic forms of intervention can be identified. First, direct intervention sought to supersede markets and embraced direct government production and regulation. Thus, governments often impose standards of food hygiene, water and air pollution, etc., and provide vaccination and other medical services in the event of epidemics. Second, indirect intervention attempts to work through the market mechanism by means of taxes and subsidies. For instance, because education is supposed to confer benefits on society at large, in addition to those bestowed on the recipients of education, it receives large subsidies from the fiscus.

However, in principle at least, the problems posed by externalities do not necessarily require government intervention. The so-called Coase theorem, which stemmed from Ronald Coase’s^[6] brilliant paper on “The Problem of Social Cost,” demonstrates the validity of this proposition. In essence, the Coase theorem holds that if nothing obstructs efficient bargaining between parties affected by an externality, then people will negotiate until they reach a Pareto efficient outcome. Accordingly, no government intervention is necessary. However, for the Coase theorem to work requires the invocation of several heroic assumptions, including the existence of costless bargaining with no transaction costs.

A third kind of market failure stems from the inability of private markets to produce public goods. Numerous public goods exist that cannot be provided through the competitive market process due to their particular characteristics. Pure public goods are said to be both non-rival in consumption and non-excludable in consumption. Non-rivalrous consumption occurs where one person's consumption does not reduce the good's availability for consumption by others as, for instance, in the case of national defense. Non-exclusion means that producers of the good are technologically and/or economically unable to prevent individuals from consuming the good as, for example, in the case of radio and television transmissions.

Given these characteristics, there is no incentive for private firms to provide public goods; accordingly, they must be furnished collectively through either private voluntary arrangements or through government agencies. If public goods are provided through the public sector, this does not necessarily imply public production, but may simply mean public provision through contracting. Thus, although some public goods, such as law and order, are produced directly by government through justice and police agencies, other public goods, such as highways and water reticulation systems, are provided by governments, but constructed under contract by private firms.

The efficient operation of the invisible hand implies that markets will provide all goods and services where demand is sufficient to cover the costs of supplying these goods and services. Where this does not occur, economists argue that incomplete markets have resulted in market failure. It is often argued that this form of market failure is especially evident in insurance markets and capital markets. In many countries, governments intervene to provide loans (or at least guarantee loans) to categories of home mortgage borrowers, small businesses, export industries, and farmers in the belief that private credit markets would not provide capital in the absence of such intervention. A related form of market failure due to incomplete markets focuses on the purported absence of complementary markets. Complementarity between markets is said to occur where activity in one market is dependent on the existence of other, related markets. For instance, large-scale property redevelopment in modern cities typically requires extensive coordination between local authorities and many private firms, which is often provided by government development agencies on the assumption that the necessary level of coordination would not be forthcoming without public intervention.

An additional source of market failure resides in the fact that economic agents on one or both sides of a market may possess incomplete information, or available information may be asymmetrically allocated

among market participants. Akerloff's^[7] famous analysis of the market for used cars is a good illustration of this source of market failure. In advancing the concept of adverse selection, or a situation in which those on the informed side of the market self-select in a way that harms the uninformed side of the market, Akerloff argued that in used car markets sellers typically possess better information than buyers. Because sellers are more adept at picking "duds" or "lemons" than buyers, buyers will realize the possibility that they may inadvertently purchase a "dud" is high and offer only "discounted" prices to reflect this perception. In turn, sellers will be unwillingly to part with good used cars and will sell only "duds" or "lemons." The final outcome will be an adverse selection process where "lemons" crowd out good used cars.

Market failure due to information failure has led to widespread government intervention in developed market economies. Much public policy making has focused on measures to protect consumers, especially product labeling and the disclosure of product content. But direct government intervention has also been evident. For example, weather forecasting is usually produced by public agencies as a means of disseminating information on weather patterns.

A final source of market failure is often argued to exist in the macroeconomics of market societies in the form of the business cycle. Periodic downswings in economic activity result in unemployment and falling incomes, whereas upswings in economic growth may generate inflationary episodes. Macroeconomic instability of this kind is usually met with government intervention in the guise of macroeconomic policies intended to flatten the business cycle. Typically, these macroeconomic policies seek increased economic activity during recessionary periods and decrease such activity during boom periods.

ETHICAL CONSIDERATIONS

Thus far, we have only considered market failure from the perspective of economic efficiency. If markets failed to yield economically efficient outcomes, then we have seen that a *prima facie* case for government intervention can be made on efficiency grounds alone. But even if markets do generate economically efficient outcomes, additional ethical arguments may still be invoked to justify a role for public policy. This, of course, requires a somewhat broader definition of the meaning of market failure. Wolf's observation that "markets may fail to produce either economically optimal (efficient) or socially desirable (equitable) outcomes" is a good example of a broader definition (Ref.^[8], pp. 19–20).

The introduction of equity or fairness considerations into economic analysis brings with it complex issues,^[9] not least the fact that a virtually inexhaustible range of plausible standards for judging equity exist. Quite apart from the difficulties raised by the existence of the myriad of possible equity criteria, further complications arise when different methods of approaching ethical questions are considered. For example, a useful distinction can be made between approaches to equity, which focus on distributive processes, opportunities, and outcomes, respectively. Thus, an approach that emphasizes distributive outcomes might call for equality between all individuals. But this would ignore the procedural aspects of distributive outcomes. For instance, it is entirely conceivable that unequal distributions could be fair if they are the result of fair processes of exchange. By the same token, an approach that embraces procedural fairness to the exclusion of outcomes would allow massive income and wealth differences to occur. Moreover, both the outcome and procedural approaches neglect initial endowments enjoyed by individuals in society. But considerations of equality of opportunity require complex interpersonal judgments on the inherited abilities of people and their socioeconomic backgrounds. It is clear that the adoption of any of these three generic approaches to fairness involves trade-offs between processes, opportunities, and outcomes, not to mention more general trade-offs between equity and efficiency.

Notwithstanding these complications, three common ethical arguments are often used to support government intervention. First, widespread support exists for the contention that the distributive results of efficient markets may not meet socially accepted standards of equity, or accord with a desire to reduce extremes of wealth and poverty. Moreover, practical politics tends to emphasize distributive issues. Wolf^[8] put the matter thus:

[M]ost public policy decisions are usually even more concerned with distributional issues (namely, who gets the benefits and who pays the costs) than with efficiency issues (namely, how large are the benefits and costs). (original emphasis) (p. 30)

Equity-based distributional arguments along these lines have been used to justify massive redistribution programs characteristic of the modern welfare state. Specific programs range from targeting beneficiaries, such as the Aid to Families with Dependent Children program in the United States, to universal coverage, such as the Medicare system in Australia.

A second line of argument is based on the notion that people do not always behave in their own best interests. For instance, unless compelled by law, some citizens might not send their children to primary school or wear seatbelts. Others might ingest dangerous narcotics or watch violent pornographic movies. Arguments against these kinds of behavior are based on the concept of merit goods, not on distinctions between private and public goods. Merit goods are defined “as goods the provision of which society (as distinct from the preferences of the individual consumer) wishes to encourage or, in the case of demerit goods, deter” (Ref.^[10] [p. 78]). Thus, government intervention in the form of subsidies to the performing arts or prohibitions against marijuana smoking can be justified.

A further common (if highly contentious) ethical argument for public policy intrusion into economically efficient market outcomes is based on the idea of equal economic opportunity. It is sometimes claimed that markets resort to ethnic, gender, or racial stereotypes as a filtering device in labor markets, and that these biases are reflected in employment patterns. Exponents of these arguments call for government intervention in labor markets in the form of equal opportunity programs and affirmative action laws to improve the job prospects for minority groups.

REFERENCES

1. Baumol, W.J. *Welfare Economics and the Theory of the State*; Harvard University Press: Cambridge, MA, 1952.
2. Samuelson, P.A. The pure theory of public expenditure. *Rev. Econ. Stat.* **1954**, *36*, 378–389. November.
3. Bator, F. The anatomy of market failure. *Q. J. Econ.* **1958**, *72* (2), 311–400.
4. Mitchell, W.C.; Simmons, R.T. *Beyond Politics*; Westview Press: San Francisco, 1994.
5. Pigou, A.C. *The Economics of Welfare*; Macmillan: London, 1920.
6. Coase, R.H. The problem of social cost. *J. Law Econ.* **1960**, *3* (1), 1–44.
7. Akerloff, G.A. The market for ‘lemons’: Qualitative uncertainty and the market mechanism. *Q. J. Econ.* **1970**, *72* (2), 388–400.
8. Wolf, C. *Markets or Governments: Choosing Between Imperfect Alternatives*; MIT Press: Cambridge, 1989.
9. Kolm, S.C. *Modern Theories of Justice*; MII Press: Cambridge, 1977.
10. Musgrave, R.A.; Musgrave, P.B. *Public Finance in Theory and Practice*; McGraw-Hill: Singapore, 1984.

Mathematical Modeling

Roger J. Beck

*Department of Agribusiness Economics, Southern Illinois University Carbondale,
Carbondale, Illinois, U.S.A.*

INTRODUCTION

To some people, it is inconceivable that mathematical models could capture economic growth. Recruiting new firms to the area or retaining existing firms does not require mathematics or even economics. It is a matter of persuasion and negotiation. However, if the goal is to become familiar with the process of business location decisions and whether a firm continues production in an area, it can be useful to develop a model that regularizes the decision-making process. Then the model can be tested with data to determine whether location and retention is a random process. If it is not, and the model has variables that can be manipulated, the community has some control over its economic destiny.

UNIT OF ANALYSIS

Modeling community economic analysis requires that the community be the unit of analysis. What constitutes the community? The community has spatial dimensions constructed from economic considerations. Where do people shop? Where is their primary place of employment? How do they interact and communicate with each other? These are questions that, if data are gathered and analyzed, can identify communities in a spatial sense. The term “region” is often used interchangeable with “community,” with a difference being that regions have historically been designated on the basis of primary production activities. The Coal and Steel Region of Europe or the Piedmont, Corn Belt, or Delta regions of the United States are examples. A more current use of community is “Silicon Valley.” That is, community implies interactions around some unifying social or economic activity.

But what really constitutes that community? There are discrete units that comprise the community, and they interact with and among each other. They are households that supply resources of production such as land, labor, capital, and entrepreneurship, and are rewarded with rent, wages and salaries, interest, and profits, respectively, for having the insights, knowledge base, and willingness to assume risk in anticipation of

returns that exceed costs and compensate the supplier of the resource. Businesses combine these resources to produce a product or service that can be sold at a profit to sustain the business. Mathematical modeling requires developing equations that describe and capture the essential elements of this process. The model developed ought to be comprehensive, yet flexible enough to be capable of addressing questions posed by the regional planners, decision makers, and even economic development researchers.

METHODS OF ANALYSIS

The essential elements of that model should include demand characteristics of the product or service being produced; supply considerations including availability of resources used to produce the product or service; location considerations of markets, resources, and associated costs; institutional considerations; and technological changes that affect all the above.

Export base models have a certain appeal because of their inherent simplicity and their apparent adaptability to econometric models. An export base model states that community growth is a function of community export performance. The value of the theory is that it emphasizes the openness of a community economy, and the role played by markets for the goods and services produced by the community that emanates from outside the community. Yet, the model ignores investment, immigration of persons, and flows of capital as well as labor. Thus, it is really not an acceptable mathematical model of community growth and development.

Economists often rely on neoclassical models to explain community growth. This would be a world in which perfect competition, complete certainty, and marginal adjustments in prices, outputs, and locations rule. This type of world probably only exists in classrooms and textbooks. The existence of space, that locations of consumption and production affect community economics, is incompatible with these restrictive assumptions. One of the reasons it has been used so extensively is that it subsumes a theory of factor mobility and economic growth. Flows of product,

labor, and money occur if the rate of return or reward differs, until these rates approach some constant allowing for differences in transport costs.

There is a set of mathematical models that vary as far from neoclassical models as can be imagined in form, substance, and predicted outcomes. Neoclassical models predict convergence of income, while cumulative causation models predict per capita income divergence. Myrdal^[1] argued that the play of economic forces tends to increase, rather than decrease, inequalities between regions. In recent times, the divergence between rural and urban regions has indeed widened rather than narrowed. Myrdal stresses locational advantages as well as resource endowments that set off a trigger, which results in ever-widening per capita incomes rather than the predicted convergence.

Econometric models have been used to predict economic growth of a community; these models are primarily forecasting models. These models normally use time-series data rather than cross-sectional data required by community growth. They are best suited when used in conjunction with neoclassical models.

Specifically, the model is expected to explain factors that affect the supply of the resources at a point in space where the production of the product or service takes place. This implies that labor force participation models be combined with commuting models to predict the quantity of labor with specific skills at given wage and fringe levels available at the point of production. Such a labor force function could be specified that hypothesizes that the labor force at a point in space is a function of socioeconomic characteristics, and a variable reflecting geographic access to employment opportunities at a point in space. Other resources such as telecommunication services, land use, and input supplies could be similarly modeled to reflect their availability.

It is known that each economic entity not only produces goods and services for final demand—that is, household, capital goods for businesses, exports from the community, and for government—but also produce goods and services for firms in the community that produce other final goods. Walras stressed the interdependence between the production sectors of an economy with his general equilibrium model in the 1870s.^[2]

The first empirical application of the input–output (I–O) model in the Western world dates from 1936, when Leontief published an I–O system of the U.S. economy.^[2] Leontief simplified Walras' generalized model so that equations associated with it could be estimated empirically. He used two simplifying assumptions. First, the large number of commodities in the Walras' model was aggregated into relatively few outputs, one for each industrial sector. Second, supply equations for labor and demand equations were

abandoned. The remaining production equations were expressed in their simplest linear form.

The model is simple, however, and lends itself to empirical analysis. Its key variables are the outputs of sectors into which the economy is divided. Each sector's output consists of the sum of sales to all other sectors and final demand. The amount of each product, which each sector consumes, depends only the level of output in the consuming sector. Equilibrium in the economy is attained when the output each sector total purchases from that sector, these purchases being determined by the output of all other sectors.

The key to Leontief's analytical system is the construction of the I–O or transactions table. The transactions table shows the flow of commodities from each producing sector to all other consuming sectors, both intermediate and final. Data provided in the transactions table can be rearranged to derive a table of technical coefficients, and a table of direct and indirect requirements from which I–O multipliers are derived.

KEY ASSUMPTIONS

The first assumption is that the production functions (an industry's list of expenditures) are assumed to have constant returns to scale. This means that if additional output is required, all inputs increase proportionately. A second assumption is that there are no supply constraints. This implies an industry has unlimited access to raw and manufactured inputs, and its output is limited only by the demand for its products. A fixed commodity input structure implies that price changes do not cause a firm to buy substitute goods. A fixed commodity input structure assumes that changes in the economy will affect the industry's output, but not the mix of commodities and services it requires to make its products. The fourth assumption is that there is homogeneous sector output. This means that the proportions of all the commodities produced by that industry remain the same, regardless of total output. An industry technology assumption assumes that an industry uses the same technology to produce all its products.

IMPACT ANALYSIS

Economic impact analysis involves applying a final demand change to a predictive economic I–O model and then analyzing the resulting changes in the community economy. In practice, economic impact analysis might mean many different things. The effect of a new manufacturing plant or discontinued operations is another use. Policy analysis might require the economic effects of generating electricity at the mouth of a coal mine. What are the economic effects of removing marginal

land from agricultural production, creating buffer zones along streams and rivers? These questions are what an operational I–O model is expected to address.

REFERENCES

1. Myrdal, G. *Economic Theory and Underdeveloped Regions*; Methuen: London, 1963.
2. Miller, R.E.; Blair, P.D. *Input–Output Analysis Foundations and Extensions*; Prentice-Hall: Englewood Cliffs, NJ, 1985.

BIBLIOGRAPHY

1. Hamilton, J.R.; Whittlesey, N.K., et al. Economic impacts, value added, and benefits in regional project analysis. *Am. J. Agric. Econ.* **1991**, 73 (2), 334–344.
2. Maki, W.R.; Lichty, R. *Urban Regional Economics: Concepts, Tools, Applications*; Iowa State University Press: Ames, IA.
3. Shaffer, R. *Community Economics*; Iowa State University Press: Ames, IA, 1989.
4. *Micro IMPLAN Software Manual*; University of Minnesota: St. Paul, 1989. University of Minnesota.

Media and Bureaucracy in the United States

Mordecai Lee

*Department of Governmental Affairs, University of Wisconsin-Milwaukee,
Milwaukee, Wisconsin, U.S.A.*

INTRODUCTION

This article reviews media coverage of public administration in the United States. First, it discusses the history of such coverage and ties it to an element of American political culture, which is consistently negative toward government. Second, the contemporary state of news coverage of American bureaucracy in the early 21st century is examined. It identifies elements that contribute to and exacerbate such negative coverage, including an antibureaucracy metanarrative that tends to be embedded in journalism, the difficulty of making public administration interesting in a visual medium, the emergence of attack journalism, and the reduction in original reporting vs. the rise in punditing.

Third, possible responses by civil servants are presented. Public administrators have little influence to attenuate the antibureaucracy tilt of news coverage. However, government managers have some modest options regarding softening the hard edges of this coverage. They can professionalize relations with reporters, take advantage of opportunities to present positive stories, and utilize new digital technologies to communicate directly with the public instead of relying on the traditional intermediary role of the news media.

HISTORY

The negative view of government in American journalism is as old as America itself. The American revolutionary war was justified as a reaction to tyranny, depicted as the abusive use of government power by Great Britain against the colonists. The 1776 Declaration of Independence comprised little more than a litany of complaints about the British government. A primary vehicle for the revolution was circulars and pamphlets denouncing uncaring and distant government by Committees of Correspondence, early versions of American journalism.

With the founding of the United States, the First Amendment to the federal Constitution guaranteed freedom of the press in order to facilitate journalism's role as an instrument of democracy. The duty of an independent press was to report what government did and did not do, so that an informed public could exercise its will, whether through formal mechanisms

(elections) or informal ones (public opinion). Therefore, the ethos of American journalism was that its *raison d'être* was to cover government critically. Freed of government regulation, American newspapers were presumed to take an adversarial attitude toward government, helping guarantee democracy and the other constitutional freedoms possessed by the citizenry.

The antigovernment strain of U.S. journalistic and civic culture has been a permanent phenomenon since then. While the emergence of a merit-based and professionalized civil service in the late 19th century was a new element in the public sector, public administration was still part of government, the presumed heavy in the given narrative of American society. While muckraking was viewed as a new form of journalism at the beginning of the 20th century, it shared a world view that was not novel: government was one of the bad guys, even though Progressive reformers often turned to government to solve problems. The expose of problems in Chicago's meat packing plants was as much a condemnation of corrupt or ineffective government inspections as it was of the mendacity of big business. One staple of journalism of that era was exposing the malfeasance or corruption of urban political machines, how they misused government agencies to line their own pockets, whether with rigged bidding for purchase of goods and services or the corruption of law enforcement to protect organized crime. Reporters often called for reform, sometimes with government as the problem, sometimes as the solution, sometimes both. It was in that context that public administration emerged in the United States. Separating politics from administration through civil service reform was viewed then as a solution (although, retrospect, a chimera).

The so-called "golden age" of media coverage of the bureaucracy began with Franklin Roosevelt's New Deal, when an alphabet soup of administrative agencies sprouted up to rid the country of the Great Depression and the power of the economic royalists. Suddenly, public administration was not only the solution, but also the possessor of nearly inexplicable expertise. That made reporters more than ever dependent on the public relations staffs of the new agencies. Journalists needed help to understand and interpret bureaucratic actions into terms meaningful and understandable to their lay readers.^[1] The intense coverage of the bureaucracy—and deference to its innate fairness, wisdom, and

patriotism—largely continued through World War II and the first decades of the Cold War. Kurtz described media coverage of that era as “page after page of gray type and boring stories about bureaucracy.”^[2] However, those mid-century decades of extensive and largely informative coverage of public administration need to be seen now as an aberration, not the normal state of affairs.

The return to journalistic normalcy, of negative coverage of the bureaucracy (in contradistinction to how reporters covered politicians in general and presidents in particular) was triggered by the coalescing of several trends. In part, reporters fell out of love with public administration, after such high-profile events as the failure of Defense Secretary Robert McNamara’s bureaucratic approach to win the Viet Nam war, the domestic bureaucracy’s inability to win the War on Poverty and the seeming failure of the economic policy bureaucracy to tame inflation in the 1970s. News reporting of government, especially after Watergate, became increasingly negative. Not just the traditional skeptical and adversarial stance, but of barely disguised cynicism, disbelief, and outright hostility. All of government, whether elected or civil service, bore the brunt of that. But, most importantly, then came television.

By the 1970s, television had fully eclipsed print as the dominant news media. This had several key consequences for coverage of the bureaucracy, mostly dictated by the medium’s technological parameters. First, TV was a visual medium and public administration was hard to cover when pictures were needed. Most developments from government departments were long-running policy issues, mind numbing in detail, and took too long to explain. Sure, efforts at securities regulation were an important story, but the subject was inherently detail oriented and lacking in dramatic pictures. Second, even more than print journalism, television was best at telling short stories, which required a beginning, middle, and end, all wrapped up in the same report. Again, disadvantage bureaucracy. TV was especially good at telling dramatic and emotion-laden stories about a person, a suffering victim, who the viewers could empathize with. This kind of storytelling was conducive to covering the bureaucracy, but only when it was the bad guy, insensitively victimizing a poor innocent citizen.

By the end of the 20th century, admitted a senior reporter for the *New York Times*, “Newspapers cover governmental *process* less than they used to. Television hardly covers it at all.”^[3] And public administration is all about *process*. The trend accelerated during the first decade of the 21st century. Kettl observed that “‘Waste-fraud-and-abuse’ has become a single word in the media.”^[4] Also, there was more and more news, but less original reporting and, of that reporting, even

less of it was about departments and agencies. And what little there was almost always was negative.^[5] “Television newsmagazines regularly alternated “give me a break!” features about government gone amuck with plaintive stories of citizens in need” who were being ignored by bureaucrats.^[6]

The impact of these trends in media coverage of the bureaucracy meant that citizens increasingly considered government as “something that may be watched, and even viewed with disgust, but nonetheless regarded as a series of events with which the individual citizen is not directly involved.”^[7] This further separated the citizenry from experiencing or exercising its ownership role in democratic government, especially public administration.

AS SEEN ON YOUR TV SCREEN: CONTEMPORARY MEDIA COVERAGE OF AMERICAN BUREAUCRACY

The technical requirements for TV storytelling greatly exacerbated the pre-existing secular trends in American journalism: a general distrust of government and pervasively negative spin to coverage. Public administration was a made-for-TV heavy. As TV news sought to be increasingly entertaining and outrage promoting, a new mutation of reporting, called attack journalism, had the daily effect of depicting the bureaucracy as deviant and arousing public indignation about these transgressors of goodness.^[8] Public administrators faced a lose-lose world of a “court of public opinion in which media is the prosecutor, judge, and decision maker.”^[9] (Somewhat similar trends have begun occurring in media coverage of the non-profit sector.^[10])

Zoonen sought to identify a typology of contemporary news coverage. He suggested there were four metanarratives that reporters consciously or unconsciously used to craft their stories: quest, bureaucracy, conspiracy, and soap. Bureaucracy stories were ones dominated by motifs of incompetence, waste and “as a force beyond the individual’s control.” Zoonen also suggested that the four metanarrative categories were not mutually exclusive, with the story line of news coverage often invoking several simultaneously. Hence, bureaucracy stories were also a good fit for the running themes of quest (bureaucracy as the evil obstacle to a successful quest to do right), conspiracy (deliberate and secretive efforts to harm society), and soap (dramatic tale of an individual up against an institution that inherently suppresses individual choice).^[11] The antibureaucracy metanarrative can be embedded in many kinds of stories, even those not specifically about the bureaucracy, rather tossed in for good measure.

For news coverage that is explicitly about government agencies, Vermeer identified three categories for

such coverage: stories about failure and conflict; about high-profile agencies (such as EPA and social security) rather than obscure ones; and stories about the particular civic interest in a local media market.^[12] An example of the latter would be the interest in southwestern states in water decisions by the U.S. Bureau of Reclamation, which is part of the Interior Department. As with metanarratives, these kinds of stories were not mutually exclusive. Continuing with the same example, a story about a new Bureau of Reclamation water policy might not only fit in the third category of local interest, but also the first category, regarding conflict. Water distribution decisions can be very controversial. Different categories of users would criticize Bureau policies for proposing policies that go against their particular parochial interests.

Then television begat cable TV, which begat 24-hour cable news networks, which begat punditry as an industry. The right-of-center tilt of talk TV (and talk radio) meant commentary was inherently antigovernment, even when the right-of-center party controlled all elected institutions in Washington. In a parallel, but decreasingly separate, reality the internet begat real-time news sites (often of newspapers seeking to adapt and survive), which begat the blogosphere, which made every American a pundit. Blogging is not, strictly speaking, journalism, but in the first decade of the 21st century it began influencing and driving the media's agenda for coverage.

BUREAUCRATIC RESPONSES TO MEDIA COVERAGE

As the Darth Vader of early 21st century America, it would seem that U.S. bureaucrats have few options regarding their media image. To paraphrase the 1990s bestseller, administrators are from Jupiter and reporters are from Mercury. In Imperial Rome, Jupiter was the god of the state, while Mercury was the bearer of messages, often eloquent, but cunning and unpredictable. Can Jovian managers and mercurial reporters coexist? The adaptations required of public administration to survive necessitate some traditional as well as innovative responses. Most of all, government managers need to be philosophical about the perennial journalistic hostility to their work. This is little different from the persistence of a baseball pitcher at striking out each batter. As batters, public administrators need to be satisfied with a .300 average, even .250, since that is about as good as it will ever get.^[13]

Media Relations and Public Information

Love 'em or hate 'em, because of the essential difference between public administration and of management in

the business and nonprofit sectors, the constitutional, legal, and professional duty of the civil servant is to cooperate with the media. The government manager is part of the public sector and, therefore, must be a participant in the democratic system. Given the First Amendment and the federal Freedom of Information Act (or its state counterparts, sometimes called open records or sunshine laws), public agencies must answer reporters' inquiries, whether the answers are in the best interest of the agency or not.

However, this necessity to cooperate with the media can sometimes be converted into a virtue. Beyond passively answering unsolicited requests for information from the media, public administration can view media relations and public information as a tool of governance, which may be used to advance the mission of the agency.^[14] Its various purposes could include informing citizens of eligibility for new services; informing the public of new laws that may affect them personally; seeking to promote socially desired behavior (only you can prevent forest fires); and promoting voluntary cooperation with laws and regulations rather than relying on expensive regulatory and enforcement approaches.^[15] Also, there are circumstances when the media seeks and disseminates agency information. For example, in times of crisis and disaster, relevant departments are given virtually unlimited access to media channels to provide instructions, advice, and updates to affected populations.^[16]

In-House Staffing and Planning

Good (or, at least, occasionally decent) public relations does not just happen. It requires the agency having professional staff in its public information bureau, training all employees in the basics of handling unsolicited media inquiries, having top managers who are attuned to public relations, involving senior public information officials in decision making, and engaging in planning for foreseeable scenarios as well as unforeseeable crises.

Bypassing the Media

Other bureaucratic counterstrategies can aim at changing the context of media coverage or bypassing the media altogether. For example, through policy entrepreneurship, the administrator can gradually build support for the "do something!" dynamic of American society. By working behind the scenes to help define a problem with various attentive audiences, the government manager can create a frame of reference that invariably puts the agency in a good light, as responding to a social need.^[17] Similarly, the traditional duty of a public administrator to report to the

citizenry on agency activities can be transformed in this digital era to e-reporting, using the internet to reach citizens directly with information about agency performance.^[18]

CONCLUSION

A natural human tendency is to complain about a situation that one is dissatisfied with. Like most Americans, reporters complain through their journalism about bureaucracy and then bureaucrats complain about hostile media coverage. The difference between the two is that reporters have the power to frame events for news consumers while public administrators do not. This asymmetry of power is at the heart of the problem of media coverage of American bureaucracy. For the foreseeable future, Mercury always trumps Jupiter. The negativity of American political culture to government is permanent and is the root cause that leads to the harshness of media coverage of bureaucracy in the United States. Developments in the news industry in the first decade of the 21st century made press coverage even more negative. There are more profit-oriented media outlets, which leads to hyper-competition for audiences, less original reporting, more punditing, more spin, and an obsession with scandal and outrage.

However, with the emergence of digital and internet-based technologies for e-government and e-democracy in the 21st century, public sector agencies have, for the first time in American history, the potential of bypassing the media and reaching the citizenry directly. Also, occasional policy openings or crisis situations provide opportunities for agencies to present a countervailing picture to the citizenry through a cooperative news media. Still, public administrators have little choice but to be philosophical about this national tendency to bureaucracy bashing. Most have experienced such negative media coverage to prove it. And, with the Americanization of global media patterns, this is becoming an international tendency as well.^[19]

ARTICLES OF FURTHER INTEREST

Freedom of Information Act—Federal, p. 805.
Nonprofit Organizations, External Relations in, p. 1351.
Popular Reporting, p. 1516.
Public Reporting, p. 1652.

REFERENCES

1. McCamy, J.L. *Government Publicity: Its Practice in Federal Administration*; University of Chicago Press: Chicago, 1939.
2. Kurtz, H. *Media Circus: The Trouble with America's Newspapers*; Times Books: New York, 1994; 357 pp.
3. Clymer, A. *Edward M. Kennedy: A Biography*; William Morrow: New York, 1999; 607 pp, emphasis added.
4. Kettl, D.F. Relentless reinvention. *Government Executive* **2000**, 32 (1), 26.
5. Lichter, S.R.; McGinnis, P. Government in and out of the news. *Public Manager* **2003**, 32 (3), 47–51.
6. Kettl, D.F. *The Transformation of Governance: Public Administration for Twenty-First Century America*; Johns Hopkins University Press: Baltimore, 2002; 63 pp.
7. Roberts, A. *Blacked Out: Government Secrecy in the Information Age*; Cambridge University Press: New York, 2006; 236 pp.
8. Patterson, T.; Seib, P. Informing the public. In *The Press*; Overholser, G., Jamieson, K.H., Eds.; Oxford University Press: New York, 2005; 196 pp.
9. Kalantari, B. Media and the bureaucracy in the United States. In *Handbook of Comparative and Development Public Administration*, 2nd Ed.; Farazmand, A., Ed.; Marcel Dekker: New York, 2001; 881 pp.
10. Stephenson, M., Jr.; Chaves, E. The nature conservancy, the press, and accountability. *Nonprofit Voluntary Sector Quart.* **2006**, 35 (3), 345–366.
11. Zoonen, L.V. *Entertaining the Citizen: When Politics and Popular Culture Converge*; Rowman & Littlefield: Lanham, MD, 2005; chapter 7.
12. Vermeer, J. Bureaucracy and the media. In *Media Power, Media Politics*; Rozell, M.J., Ed.; Rowman & Littlefield: Lanham, MD, 2003; 102 pp.
13. Lee, M. Administrators are from Jupiter, reporters are from Mercury. *PA Times* **2001**, 24 (6), 7.
14. Weiss, J.A. Public information. In *The Tools of Government: A Guide to the New Governance*; Salamon, L.S., Ed.; Oxford University Press: New York, 2002; 217–254.
15. Lee, M. Public information in government organizations: a review and curriculum outline of external relations in public administration. *Public Admin. Manage. Interactive J.* **2000**, 5 (4), 183–214; http://www.pamij.com/5_4/5_4_4_pubinfo.pdf (accessed June 2007).
16. Lee, M. Media relations and external communications during a disaster. In *Disaster Management Handbook*; Pinkowski, J., Ed.; Taylor & Francis: Boca Raton, FL, 2008; chapter 19.
17. Lee, M. Reporters and bureaucrats: public relations counter-strategies by public administrators in an era of media disinterest in government. *Public Relations Rev.* **1999**, 25 (4), 456–457.
18. Lee, M. E-reporting: using managing-for-results data to strengthen democratic accountability. In *Managing for Results 2005*; Kamensky, J.M., Morales, A., Eds.; Rowman & Littlefield: Lanham, MD, 2005; chapter 4. Also available online as a monograph: http://www.businessofgovernment.org/pdfs/Lee_Report.pdf (accessed February 2007).
19. Lee, M. Globalization and media coverage of public administration. In *Handbook of Globalization, Governance, and Public Administration*; Farazmand, A., Pinkowski, J., Eds.; Taylor & Francis: Boca Raton, FL, 2007; chapter 8.

Medicaid Reforms in the States, Administration of

Thomas E. Yatsco

U.S. Department of Health and Human Services, Washington, District of Columbia, U.S.A.

Irene J. Barnett

Department of Political Science, Kent State University, Kent, Ohio, U.S.A.

INTRODUCTION

A program, with implementation and funding that is shared between the federal government and the states, Medicaid is the source of health insurance for over 40 million Americans who are elderly, disabled, or poor. In fact, one in seven Americans receive health insurance through Medicaid. Medicaid is administered at the federal level by the Centers for Medicare and Medicaid Services (CMS), a unit of the U.S. Department of Health and Human Services (DHHS).

Medicaid experienced dramatic change during the 1990s, especially in regard to how the federal government and states coordinate their efforts to provide health care insurance, and the next decade looks to be equally dramatic. This article reviews major trends in spending and eligibility as well as federal issues that affect states' administration of Medicaid and their policy decisions. Specifically, examined are Medicaid waivers that states use to implement health care innovation; changes in the relationship between Medicaid and managed care; developments in providing health care to needy children under the State Children's Health Insurance Program (SCHIP); issues related to Medicaid long-term care; and efforts to improve federal management and oversight of Medicaid.

Medicaid spending grew at rates in excess of 25% per year. Spending moderated in the late 1990s due to federal legal and regulatory efforts aimed at cracking down on states' use of "creative" financing mechanisms like excessive disproportionate share hospital payments, intergovernmental transfers, and provider taxes and donations. Furthermore, expenditures remained constant as enrollment remained steady due to a strong economy along with state welfare reform and fewer expansions in eligibility. Services provided to Medicaid recipients were provided at costs that remained constant due to moderate health care price inflation along with the expanded use of managed care and the increased use of home and community-based alternatives to costly institutional long-term care.^[3]

Observers predict that Medicaid expenditure growth will start escalating in the new century and will once again become a major drain on state budgets. The Congressional Budget Office predicts Medicaid spending to increase by 8 to 9% over the next several years. Factors that will likely cause a reversal of recent declines include rising prescription drug costs, increasing enrollment of the disabled who require costly long-term care, pressure from the health care industry to increase payments levels, and the growing likelihood of an economic slowdown.^[3]

TRENDS IN MEDICAID SPENDING

Spending on Medicaid's 40 million beneficiaries is projected to total over \$220 billion during federal fiscal year 2001.^[1] Medicaid expenditures are shared between the states and the federal government. Federal matching funds are provided to the states based on their per capita income. In 1998, the federal government's share of Medicaid expenditures averaged 57% across all the states.^[2]

Medicaid spending grew at an average rate of 11% between 1980 and 2000. However, Medicaid expenditure growth has been relatively modest since the mid-1990s, and spending grew by only 5% in 1998. This marked a significant change from the early 1990s, when

TRENDS IN MEDICAID ENROLLMENT

Medicaid provides care to over 40 million individuals. The majority of Medicaid beneficiaries are children, who account for 20 million beneficiaries. In fact, one in four children in the United States are covered under Medicaid. The rest of Medicaid's recipients are comprised of more than eight million non-elderly adults with children in the home, seven million disabled persons, and four million elderly adults.^[2]

Medicaid enrollment grew rapidly in the early to mid-1990s, increasing from 29 million in 1990 to nearly 42 million in 1995. However, from 1995 to 1998, enrollment actually decreased. This reversal in trend largely can be attributed to a strong economy and the effects

of welfare reform. In addition, the large eligibility expansions of the 1980s are now fully implemented by states, and fewer people are being enrolled. However, there are indications that Medicaid enrollment is increasing, due to factors such as better outreach efforts on the part of states to extend health care for uninsured children and a slowing economy.^[3]

FEDERAL WAIVERS FOR STATES

The authority of DHHS to waive Medicaid statutory and regulatory provisions is a tool that can be used to significantly transform state Medicaid programs. These waivers give states additional flexibility and enable them to test innovative service delivery options to expand health care coverage to uninsured populations. All programmatic changes must be carried out within the limits of budget neutrality, which means that the overall cost to the federal government must not exceed the amount that would have been spent in the absence of the waiver.^[4] DHHS has three types of waivers that it can grant to states to exempt them from federal statutory and regulatory requirements (all refer to sections of the Social Security Act): a section 1915(b) freedom of choice waiver; a section 1915(c) home and community-based services waiver; and a section 1115 research and demonstration waiver. A section 1115 waiver allows wide-ranging, comprehensive health care reforms, while sections 1915(b) and 1915(c) waivers are narrower in scope and only waive specific provisions of Medicaid law.^[5]

Section 1915(b) Freedom of Choice Waivers

This type of waiver allows states to waive Medicaid's requirements of statewideness, comparability of services, and beneficiary freedom of choice with respect to providers. In return, states must ensure that their waiver programs are budget neutral, meaning they must not exceed the cost of traditional fee-for-service programs. Specifically, this 2-year waiver allows states to require Medicaid beneficiaries to enroll in managed care plans; establish primary care case management programs; selectively contract with hospitals, nursing facilities, or other providers; and create tailored programs that do not operate statewide.^[6]

Section 1915(c) Home- and Community-Based Service Waivers

With this type of waiver, the federal government allows states to implement home- and community-based alternatives to institutional care, as long as these alternatives do not cost more than institutional care.

Specifically, 1915(c) waiver authority permits states to provide services, such as home health aides or nurse's aides, for the elderly, the disabled, or people with chronic illnesses, in a community setting rather than in nursing homes or hospitals. Home- and community-based services programs essentially enable beneficiaries to receive care in a cost-effective and less restrictive manner.^[6]

Section 1115 Research and Demonstration Waivers

This type of waiver gives DHHS broad authority to waive virtually all aspects of Medicaid law without congressional approval. In theory, section 1115 waivers are used to test unique approaches to program design and administration, but many states have sought to implement nearly identical demonstrations. Accordingly, some section 1115 waivers are a means to achieve general program changes outside of the legislative process, and tend to reflect the current administration's policy priorities. All section 1115 waiver programs are subject to strict budget neutrality standards.^[4]

From the perspective of states, waivers enable them to reduce federal requirements that apply to Medicaid program participation and implement delivery system changes to expand health insurance coverage to populations who otherwise would be ineligible for Medicaid coverage, or both. Some states also target benefits to special populations. The extent of states' policy innovation is varied. For example, some states have essentially restructured their entire Medicaid programs using 1115 waivers, with Oregon's Oregon Health Plan and Tennessee's TennCare being two of the more ambitious examples. Other states' waiver programs are less ambitious in scope or basically duplicate what other states have done.

Although the federal government possessed this authority for a long time, waivers, particularly section 1115 waivers, did not proliferate until the Clinton Administration. The Clinton Administration actively promoted states' use of federal waivers, and states made extensive use of these waivers during the 1990s.^[4] Currently, 28 states have 1915(b) waivers, every state has a 1915(c) waiver, and 18 states have 1115 waivers. The Bush Administration is even more active, approving hundreds of new waivers as well as renewing existing waivers. The Bush Administration's waivers are clearly targeted at giving states greater flexibility and even autonomy over their Medicaid programs. As the Bush Administration continues to redefine the parameters of state and federal jurisdictions over Medicaid policy, it is likely that the number and nature of waivers approved by DHHS will increase.

MEDICAID MANAGED CARE

The federal and state governments look to managed care as a means to reduce costs and improve quality of care by eliminating unnecessary services and relying more heavily on primary care and the coordination of care. The number of Medicaid managed care plans increased more than sixfold during the 1990s—from 2.7 million in 1991 to 17.8 million in 1999.^[7]

States use a variety of managed care arrangements, which generally fall into two models: risk-based and primary care case management. Under the risk-based model, a managed care organization (MCO) is paid a fixed monthly fee per beneficiary and assumes either partial or full risk for the delivery of health care services. Some MCOs deliver more limited services, such as only ambulatory care, while others provide a full range of health care services. About four-fifths of Medicaid managed care programs operate under risk-based contracts. The second model of managed care, fee-for-service primary care case management, is not risk-based. Under the primary care case management model, a provider is responsible for acting as a gatekeeper, who approves and monitors services provided to Medicaid beneficiaries. Gatekeepers are paid a monthly fee by states and do not assume financial risk for the provision of services. About one-fifth of Medicaid managed care enrollees receive care under the primary care case management model.

Managed care is not a new idea. In fact, states have been able to enroll their Medicaid beneficiaries in managed care programs for several years. Furthermore, states can mandate that Medicaid beneficiaries enroll in managed care programs through a section 1915(b) or 1115 waiver. The Balanced Budget Act (BBA) of 1997 gave states new authority to mandate enrollment in Medicaid managed care without a waiver by submitting to the federal government an amendment to the state's Medicaid plan. Also, the BBA eliminated the so-called "75-25 rule," thereby allowing Medicaid-only managed care plans. So far, few states have taken advantage of the new BBA authority by amending their state plans, largely because BBA-related final regulations are still not in place. As a result, most mandatory Medicaid managed care continues to be delivered through section 1915(b) and 1115 waivers.

The BBA also mandates beneficiary protections for Medicaid managed care enrollees in exchange for giving states greater flexibility to use managed care. However, the regulations detailing requirements regarding beneficiary protection, quality, and capacity are not yet implemented. These requirements will include direct access to emergency rooms under managed care arrangements, second opinions as needed, direct access to women's health services, and an adequate network of providers. The Clinton Administration issued a

version of these regulations in January 2001; however, the Bush Administration replaced these regulations with a new set of regulations that were released in August, 2001, for review. The new proposed regulations are considered more favorable to states, because they are less prescriptive and allow states more flexibility to formulate their own specific requirements.^[8] The final regulations are supposed to be in place by the spring of 2002.^[9]

There are several issues that remain unresolved concerning Medicaid managed care. For example, federal and state policy makers will have to grapple with how to attract an adequate number and mix of providers and how to ensure access and quality of care for beneficiaries enrolled in these managed care plans. Questions also exist about broader use of managed care for special populations, such as the elderly and disabled. Currently, most Medicaid managed care programs provide care for children and families who only account for about 25% of spending. Finally, using managed care to contain costs may not be the panacea it was envisioned to be, as the most expensive and chronic forms of care are, for the most part, not delivered through managed care arrangements.^[10]

MEDICAID AND CHILDREN'S HEALTH INSURANCE

In recent years, the federal government has taken steps to increase the number of insured children by encouraging states to expand the public health insurance safety net. Most significantly, the BBA of 1997 offered states \$40 billion over fiscal years 1998 to 2008 through the State Children's Health Insurance Program (SCHIP) to provide health coverage to children who did not qualify for Medicaid. The BBA allows states to choose one of three options for accessing the SCHIP funds to cover additional children: 1) create a stand-alone program that is separate from Medicaid; 2) expand Medicaid; or 3) do a combination of both. As of 2001, every state had a program in operation, providing coverage to more than three million children.^[6] Fifteen states have stand-alone SCHIP programs, 18 states have combination programs, and 23 states expanded coverage to previously uninsured children exclusively through Medicaid. The stand-alone option gives states greater flexibility and greater control over expenditures, as a state may limit annual contributions, create a waiting list, or stop enrollment once the funds it budgeted for SCHIP are exhausted.^[11]

SCHIP is a relatively new program and has grown recently, thereby fulfilling its mission of reducing the number of uninsured children. Initially, enrollment in the program and state spending were slow, but states improved outreach efforts, devoted more resources to

the program, and enrollment grew from nearly 2 million children in 1999 to over 3 million children in 2001.^[12] In addition, DHHS issued final regulations that spell out requirements states must follow in implementing SCHIP in June, 2001. These regulations include how to determine if a child is eligible, how to enroll children, how children must be screened for Medicaid, how limits are set for out-of-pocket expenses, and how states can design a SCHIP benefits package. For the most part, the new regulations give states greater flexibility to design and implement SCHIP.^[6]

LONG-TERM CARE

Medicaid continues to be the nation's long-term care safety net, paying for more than one-third of all long-term care. This amounted to \$65 billion in 1998, or 42% of total Medicaid spending. Over 70% of Medicaid long-term care is delivered in an institutional setting to elderly and disabled beneficiaries—in skilled nursing facilities and intermediate care facilities, or intermediate care facilities for the mentally retarded.^[13] Most states have a spend-down requirement that precludes adults from receiving Medicaid-financed long-term care until their personal assets are depleted to a certain state-specific level. Considering that on average nursing home care costs \$55,000 a year, most Americans deplete their own savings and assets rather quickly.^[14]

Although most long term care spending is still devoted to institutional care, states are increasing their commitment to home and community-based service alternatives. In fact, spending on home and community-based services as a portion of total Medicaid long-term care increased from 10 to 30% during the 1990s. These services, which are supposed to prevent a person from being institutionalized, include case management, homemaker, home health aide, personal care, habilitation, and respite care. However, states must seek a section 1915(c) waiver to provide home and community-based services or obtain a section 1115 waiver for more comprehensive programs involving long-term care. In addition, the availability of home and community-based services continues to vary by state.^[13]

The landscape for Medicaid long-term care is posed for change in the upcoming years, especially regarding developing alternatives to institutionalized care like home and community-based care services. Most significantly, the U.S. Supreme Court's *Olmstead v. L.C.* (1999) decision challenged federal, state, and local governments to administer their services and programs in the most integrated setting appropriate to the needs of qualified individuals with disabilities of any age. States have the option to develop a greater array of appropriate care services through section 1915(c)

home and community-based service waivers. To further expedite implementation of the *Olmstead* decision, in June, 2001, the Bush Administration directed federal agencies, including CMS, to assist states in expanding community-based services for people with disabilities. The agencies also were directed to create an interagency council to coordinate federal efforts for the disabled.^[15] These efforts are likely to have far-reaching consequences for Medicaid and may hasten the trend toward home and community-based care services as opposed to institutional care.

MEDICAID MANAGEMENT AND FEDERAL OVERSIGHT

CMS has experienced several limitations that compromise its ability to administer Medicaid effectively and to interact constructively with states. These problems are understandable to some extent, considering CMS' vast responsibilities; it administers Medicaid, Medicare, and SCHIP along with overseeing the implementation of many Health Insurance Portability and Accountability Act (HIPAA) provisions. However, these limitations have resulted in questionable oversight and major gaps between CMS' capabilities and stakeholders' expectations, leaving many state officials frustrated. Also, CMS' programs have been described as vulnerable to waste, fraud, abuse, and mismanagement.^[16]

Recently, the Bush Administration launched a "waive of reforms" designed to improve CMS' management, enhance accountability, and improve relations with states, providers, and contractors. In 2001, DHHS announced several initiatives including a name change for the agency from the Health Care Financing Administration to the Centers for Medicare and Medicaid Services (CMS); a restructuring of the agency to three centers representing each of the major program areas; assigning specific agency liaisons for each state; and establishing boards to interact with major groups of stakeholders, such as the nursing home industry. Also, CMS plans to implement regulatory reform, such as only publishing regulations once a month on the same day in order to provide predictability for states, providers, and contractors.^[17] These actions appear to be symbolic of the department leadership's serious commitment to deal with the problems that states and other observers have identified for years. It remains to be seen whether or not these reforms will actually improve CMS' management and oversight capabilities, improving federal-state relations.

CONCLUSION

Medicaid experienced tremendous change over the past decade. The proliferation of state waivers and

state health policy innovation, the expanded use of managed care, and the introduction of SCHIP are just a few of the many notable changes that occurred since 1990. Even more changes will take place in upcoming years as the Bush Administration implements its own agenda for Medicaid and federal-state relations. So far, one major theme is evident: greater autonomy for states. This change is due at least in part to the efforts of Secretary of Health and Human Services Tommy Thompson, the former Governor of Wisconsin, who brings a uniquely state-based perspective to his job and committed DHHS to enhancing state flexibility to implement federal requirements. Other trends to monitor in upcoming years include the extent to which Medicaid spending and enrollment grow; changes in the nature of federal waiver granted to states; developments in the occasionally uneasy relationship between Medicaid and managed care; how the SCHIP program matures and states use money to reduce the number of uninsured children and families; changes in the way long-term care is provided to the elderly and the disabled; and how successful the federal government is in improving its management and oversight of Medicaid.

REFERENCES

1. U.S. General Accounting Office. *Medicaid: State Efforts to Control Improper Payments Vary*; U.S. General Accounting Office: Washington, DC, 2001. In GAO 01-662.
2. Kaiser Commission on Medicaid and the Uninsured. *The Medicaid Program at a Glance*; Kaiser Family Foundation: Washington, DC, January 2001.
3. Bruen, B.; Holahan, J. *Medicaid Spending Growth Remained Modest in 1998, But Likely Headed Upward*; Kaiser Family Foundation: Washington, DC, February 2001.
4. Lambrew, J. *Section 1115 Waivers in Medicaid and the State Children's Health Insurance Program: An Overview*; Kaiser Commission on Medicaid and the Uninsured: Washington, DC, July 2001.
5. Callahan, J. *Testimony Before the Subcommittee on Government Management, Information*; Committee on Government Reform, U.S. House of Representatives, September 30, 1999.
6. U.S. Department of Health and Human Services. *HHS Fact Sheet—Medicaid and SCHIP Waivers: Promoting State Flexibility and Innovation*; U.S. Department of Health and Human Services: Washington, DC, 2001; Code of Federal Register. June 25, 2001; 66(122).
7. Kaiser Commission on Medicaid and the Uninsured. *Medicaid and Managed Care*; Kaiser Family Foundation: Washington, DC, February 2001.
8. Health care daily report. Medicaid: HHS issues state-friendly proposal amending Clinton-Era managed care rule. *Health Care Dly. Rep.* **2001**, 6 (159).
9. Code of Federal Register. August 20, 2001, 66(161).
10. U.S. General Accounting Office. *Medicaid: State Efforts to Control Improper Payments Vary*; U.S. General Accounting Office: Washington, DC, June 2001; GAO-10-662.
11. U.S. General Accounting Office. *Medicaid and SCHIP: Comparisons of Outreach, Enrollment Practices, and Benefits*; U.S. General Accounting Office: Washington, DC, April 2000; HEHS-00-86.
12. Kennedy, G.; Ullman, F.; Weil, A. *Three Years into SCHIP: What States Are Spending and Are Not Spending*; Urban Institute: Washington, DC, September 2000.
13. Smith, G.; O'Keefe, J.; Carpenter, L.; Doty, P.; Kennedy, G.; Burwell, B.; Mollica, R.; Williams, L. *Understanding Medicaid Home and Community Services: A Primer*; George Washington University Center for Health Policy Research: Washington, DC, October 2000.
14. U.S. General Accounting Office. *Long-Term Care: Baby Boom Generation Increases Challenge of Financing Needed Service*; U.S. General Accounting Office: Washington, DC, 2001; GAO-01-563T.
15. Bush, G.W. *Executive Order 1321: Community-Based Alternatives for Individuals with Disabilities*; White House Press Office: Washington, DC, June 19, 2001.
16. U.S. General Accounting Office. *HCFA Management: Agency Faces Multiple Challenges in Managing Its Transition to the 21st Century*; U.S. General Accounting Office: Washington, DC, February 1999.
17. Thompson, T. *Testimony Before the Committee on Finance*; U.S. Senate, June 19, 2001.

Merit System

J. Edward Kellough

*School of Public and International Affairs, University of Georgia,
Athens, Georgia, U.S.A.*

INTRODUCTION

The term “merit system” refers to a set of rules and procedures commonly established to fill and manage positions in government that comprise the civil service. A merit system is the primary means by which the merit principle in public employment is implemented. The merit principle, in turn, refers to the idea that public employees should be selected and promoted on the basis of merit defined in terms of a demonstrated ability to perform the jobs at issue. Merit systems are contrasted most clearly with patronage systems in which government employment is achieved largely on the basis of political connections as opposed to technical qualifications. Merit systems are designed to limit the intrusion of politics into the staffing and management of the civil service.

THE CONCEPT OF MERIT SYSTEMS

Merit systems are based on three interconnected ideas that form a foundation for the structure of civil service and public sector human resources management. The first of these ideas is the belief that positions should be filled on the basis of the results of open competitive examinations designed to measure applicants’ abilities to perform job-related tasks and duties. In other words, applicants are to be screened on the basis of examinations that measure their qualifications for employment in specified positions. The examinations may be written or oral. They might also consist of an evaluation or assessment of an applicant’s education, training, and experience. Sometimes, performance exams are also used that actually require an applicant to perform selected aspects of a job, such as operating a specific piece of equipment. Mental and physical aptitude tests may also be used in the examination process.

A key assumption in the merit examination process is that the tests used are valid in the sense that they actually do measure the ability to perform on the job. Three approaches to determining examination validity are available. The first, known as criterion-based validity, is based on an assessment of the empirical relationship between observed examination scores and subsequent job performance. One variation of this

approach requires that an examination be given to a group of job applicants, that they then be hired on a random basis or that they all be hired, and finally that their examination scores be correlated with subsequent levels of performance. The examination would be considered valid if those who scored high on the exam also scored high on measures of performance on the job. Although this approach may seem like a definitive assessment of validity, it requires that employees be selected on some basis other than their scores on the initial examination, and it assumes that poor performers will be brought into the organization in order that there be sufficient variation to correlate examination scores to performance measures. An alternative variation would be to give the examination to current employees and then correlate their scores to their performance outcomes. Although this avoids some of the problems associated with the earlier system, both variations assume that performance assessments are accurate, and that there will be sufficient variation in performance outcomes for a correlation between those outcomes and examination scores to be computed. Because of these difficulties, criterion-based validity is seldom used in practice. A second approach to measuring validity is based on what is known as construct validity. Here examination scores are correlated with known psychological constructs that have been determined to be associated with superior job performance. Such constructs might include honesty, ingenuity, or integrity. This approach suffers many of the same difficulties as criterion-based validity, so it too is not frequently used. The approach that is most useful is based on idea that an examination should measure knowledge of actual job content or the ability to perform certain aspects of a job. This tactic, known as content validity, is accepted by the courts and is the basis for most validity assessments. A test designed to measure an applicant’s ability to use a word processing program, for example, would have content validity provided that the job at issue primarily required the incumbent to perform word processing.

The second idea that forms the basis for merit systems is that public employees will have relative security of tenure. More specifically, this means that members of the civil service will not be subjected to dismissal or other adverse action on the basis of political

considerations. The objective is to ensure that public employees are protected from political manipulation. As a practical matter, this protection for public employees typically takes the form of civil service rules that prohibit adverse action against employees without just cause. In the United States, for example, this limitation means that public employees will have a reasonable expectation of continued employment provided they are performing their jobs satisfactorily. That expectation, in turn, is sufficient to establish a property interest in the job for each employee covered by the rule. Once that property interest is established, provisions of the U.S. Constitution that prohibit government from denying employees their property without due process of law become applicable. The due process requirement means, at a minimum, that dismissal proceedings against public employees may not proceed without prior notice and a hearing in which employees have a chance to respond to allegations brought against them.

The third major concept underlying merit systems is the notion that public employees should behave in a politically neutral fashion. This idea is seen as a necessary complement to the protections that shield employees from direct political control. The concept suggests that public employees must serve their political superiors equally well, regardless of party affiliation. In other words, civil servants will perform their duties in a politically neutral fashion, regardless of the political affiliation of their superiors.

CONCLUSION

Because merit systems are grounded in complex sets of rules and procedures, some level of centralized oversight of these systems is necessary to ensure their integrity. Generally, this oversight is provided by a central personnel agency, such as the Office of Personnel

Management for the federal government. The function of such an agency is to formulate rules regarding merit system procedures and to supervise their implementation. The rules and procedures govern personnel actions, ranging from selection, to training, job evaluation, and compensation. They are intended to ensure fidelity with merit principles, but the restrictions they place on management are seen by some as robbing management of flexibility that could make civil service systems more efficient.

BIBLIOGRAPHY

1. Ban, C.; Riccucci, N.M. *Public Personnel Management: Current Concerns, Future Challenges*, 3rd Ed.; Longman: New York, 2002.
2. Berman, E.M.; Bowman, J.S.; West, J.P.; Van Wart, M. *Human Resource Management in Public Service: Paradoxes, Processes, and Problems*; Sage Publications: Thousand Oaks, CA, 2001.
3. Daley, D.M. *Strategic Human Resources Management: People and Performance Management in the Public Sector*; Prentice Hall: Upper Saddle River, NJ, 2002.
4. Hays, S.W.; Kearney, R.C. *Public Personnel Administration: Problems and Prospects*, 4th Ed.; Prentice Hall: Englewood Cliffs, NJ, 2002.
5. Ingraham, P.W. *The Foundation of Merit: Public Service in the American Democracy*; The Johns Hopkins Press: Baltimore, MD, 1995.
6. Rosenbloom, D.H. What Every public personnel manager should know about the constitution. In *Public Personnel Administration: Problems and Prospects*, 3rd Ed.; Hays, S.W., Kearney, R.C., Eds.; Prentice Hall: Englewood Cliffs, NJ, 1995; 20–36.
7. Shafritz, J.M.; Rosenbloom, D.H.; Riccucci, N.M.; Naff, K.C.; Hyde, A.C. *Personnel Management in Government: Politics and Process*, 5th Ed.; Marcel Dekker: New York, 2001.
8. Van Riper, P.P. *History of the United States Civil Service*; Row, Peterson and Company: Evanston, IL, 1958.

Metropolitan Governance

Ulf Zimmermann

*Department of Political Science and International Affairs, Kennesaw State University,
Kennesaw, Georgia, U.S.A.*

INTRODUCTION

Metropolitan governance refers to the governing of metropolitan regions. This can be accomplished by a variety of mechanisms, ranging from a comprehensive metropolitan “government” to a variety of forms of cooperation among the numerous jurisdictions in a metropolitan area, which will be termed “governance” here. Over the period of modern metropolitan development, there has, accordingly, been a debate among scholars and practitioners (in government and business) between the respective virtues of a single unitary metropolitan government on the one hand and a multiplicity of independent, autonomous jurisdictions on the other. The unitary form is said to promise efficiencies of scale and greater equity throughout the metropolitan area. The polycentric form, as it has come to be called, is said to promise “public choice,” which affords area residents as well as government and business interests a variety of choices of how to govern themselves and regulate local land use. Both approaches will be outlined briefly below.

DEFINITIONAL PROBLEMS

Policymakers and scholars generally use the White House Office of Management and Budget’s (OMB) definition of metropolitan areas, usually referred to as “metropolitan statistical areas.” This definition specifies a county or counties with a central city or contiguous cities of 50,000 or more, plus all adjacent counties that are metropolitan in character and are economically and socially integrated with the central county as measured, respectively, by density and degree of commuting between the counties. Varying portions of MSAs have, over the years, taken on various more or less formal forms to promote interlocal cooperation on regional issues and/or to meet federal policy requirements, as illustrated by Councils of Governments (COGs) and Metropolitan Planning Organizations (MPOs). (It is worth noting that such metropolitan agencies tend to significantly trail the definitions of their MSAs; metropolitan growth steadily outdistances organizational development.^[1] The Atlanta MSA, for example, comprised 20 counties until

the latest census expanded it to 28, but the Atlanta Regional Commission, the area’s COG and MPO, represents only a 10-county area that was the MSA in 1970.)

THE MODERN METROPOLIS

While there have been metropolises since ancient times, these remained confined to little more than the scope of “walking cities” until the emergence of modern transportation technologies, beginning with horse trolleys. And though the gentry has had its villas beyond the outskirts of the city since at least Roman times, it was only over the last two centuries that technological developments made suburban dwellings available to progressively broader segments of society. Railroads made possible the first remoter suburbs for the wealthy, electric trolleys the denser closer-in ones for the “common man,” and, finally, the unconfined automobile, with its steadily increasing affordability, allowed cities and suburbs to grow in all directions. Automobiles enabled the huge residential expansion of suburbs and, importantly, these were followed by trucks, which overwhelmingly enabled the commercial and industrial “sprawl,” which gave us the complex metro areas of today.

GOVERNING IN THE METROPOLIS

The concern with governing these metropolitan areas began, not surprisingly, with their very emergence in the late 19th/early 20th centuries, as cities began to expand because of demographic pressures, an expansion, which was facilitated, indeed, encouraged by the abovementioned developments in transportation technology. Those who could afford more transportation could afford larger properties and housing farther out from the center of the city, a pattern that still characterizes metropolitan areas today. While at first a good deal of this metropolitan expansion was captured by expanding the existing city via annexation, suburbanization soon took instead the form of developers and homebuyers creating their own independent jurisdictions. This creation of multiple autonomous suburban “small republics,”^[2,3] came about, in part, because of Americans’ antiurban bias, the lower prices of land

beyond the city limits, and the highways and home mortgages created and promoted by the federal government (mobilized, in turn, by automobile, banking, construction, and a host of related lobbying groups).^[4] In addition, many states liberalized incorporation policies and made annexation more difficult for central cities. This created the polycentric metropolitan areas championed by public choice scholars.^[5]

There is, therefore, a certain aptness to the increasingly strong federal incentives for regional approaches insofar as the major problems that need to be addressed regionally can be said to have been produced by the federal government's highway and housing policies (interstates and mortgages), as noted. States, too, have taken up the cause, promoting various "smart growth" policies.

Thus, while in 1898, New York was still able to form a metropolitan government by consolidating Manhattan with five suburban counties into "Greater New York," subsequent efforts to expand and increase regional government have been stymied by suburban interests maintaining their own power and autonomy.

But by the time suburbanization reached an early crest, in the 1950s, many of these suburban authorities recognized the necessity of joint solutions for regional issues and problems, and between that recognition and the knowledge that if they, the local authorities, did not act voluntarily, a higher level of government would impose its authority on them, local leaders began to create regional conferences at which city and county officials could voluntarily negotiate certain regional policies. The first of these, which would become known as COGs, emerged among the counties around Detroit in 1954, with others following around Washington, DC and the Puget Sound in 1957. The initial ones may have been influenced by federal housing legislation, which established a regional planning assistance program in 1954; the many COGs formed in the 1960s certainly were responding to federal highway and housing legislation that mandated regional approaches.^[6]

For example, the 1959 Housing Act called for regional jurisdictions; the 1962 Federal-Aid Highway Act mandated a regional approach to transportation planning; and the 1965 Housing and Urban Development Act specified that the new department could make grants to regional councils. The 1966 Demonstration Cities and Metropolitan Development (Model Cities) Act required certain federal grant applications to be reviewed by metropolitan review agencies, as did the Intergovernmental Cooperation Act of 1968. This process was codified and strengthened by OMB's 1969 Circular A-95, which enhanced regional cooperation somewhat.^[7]

The A-95 regional review process was abolished in the 1980s, and with it even the little regional work the COGs did was further diminished. (COGs have

pursued less "political" activities like infrastructure improvements and stayed away from social program areas such as housing, health, and welfare).^[8] But new federal legislation has called for new regional goals. The Clean Air Act Amendments of 1990 placed requirements on metropolitan transportation planning. The Intermodal Surface Transportation Efficiency Act of 1991 and its successor Transportation Equity Act for the 21st Century require metropolitan areas to have a designated MPO. By imposing air quality standards and freeing up gasoline taxes for uses other than highways, this legislation in combination has fostered a certain amount of metropolitan governance by giving MPOs more power.

This has been the case in the Los Angeles area, where the Southern California Association of Governments, a COG, has taken over the transportation role of an MPO, which works together with a largely coterminous special district covering five counties, 200 cities, as well as numerous other local governments, the South Coast Air Quality Management District.^[9] A similar district, combining pollution and transportation management as well as limited land-use controls, was created around Atlanta in 1999, extending beyond the existing MPO, which had grown out of a COG, by three additional counties that fell afoul of air quality standards.^[10]

These examples do not constitute full-scale, comprehensive metropolitan government, but they have raised the metropolitanwide salience of issues that allows regional residents and their leaders to recognize that, despite the hundreds of governments and often 3000 or more square miles of such metropolitan areas, they indeed share a region and a common set of interests.^[11] And while there are areas that have opted for some form of genuine metropolitan government to address such issues, the vast majority have relied on more limited forms of metropolitan governance. Regionwide cooperation usually requires incentives from higher up, as illustrated by the Minnesota and Portland cases below and MPOs generally; local governments may work together if they perceive that they will all share losses, as in common pool resources such as air and water, but they do not see such losses coming from sprawl and are generally not willing to adopt, for example, metrowide land-use policies (as illustrated by the failure of Atlanta's metropolitan district to do this).

METROPOLITAN GOVERNMENT

The idea of a single unitary metropolitan government, which would provide all (urban) services to everyone in the metropolitan area, is one that emerged out of the progressive movement of the early 20th century, which promoted business values that claimed that metropolitan

centralization would bring benefits of economy and efficiency, thanks to its larger scale.^[12] And though research has not supported this claim, this literature's other claim, that the polycentric approach of many "small republics" exacerbates the segregation of rich and poor and of minorities, has some evidence supporting it and thus also supports the claim of metropolitan government advocates for its greater equity.^[13]

Those areas in which actual new, larger-scale governments have been formed can be viewed as having taken place in one, two, and three tiers.^[14] A one-tier system is a consolidation, which peaked with the integration of Greater New York's five boroughs in 1898 (following the previous consolidation of Manhattan and New York Borough). But most of the more recent examples have been just partially complete consolidations of a central city with its county, which tends to cover only a small proportion of the actual metropolitan area.^[15]

In some cases, these consolidations are more akin to a two-tiered system such as that in Miami-Dade (1957), known as "Metro," in which the county provides some services, while the remaining 27 cities provide others (leaving unincorporated county residents somewhat short on services). The three-tier system, which extends beyond a single county and is thus the closest to any real form of metropolitan government, is exemplified only by Minnesota's Twin Cities and Portland, Oregon.

Here neither existing cities nor counties are eliminated or merged, but a whole new third level is added, like a legislatively empowered COG or MPO, in the form of an areawide coordinating agency. The Portland Metropolitan Service District ("Metro"), for example, covers 3 counties and 24 cities, originally with responsibilities for waste disposal, zoo administration, and other services. In 1992, voters gave it a home-rule charter and expanded its authority to regional planning via land-use control as well as an affordable housing plan.^[16] This "Metro" is the only directly elected regional government in the U.S.A. The Twin Cities Metropolitan Council ("Met"), though its members are appointed by the governor, has similar powers and can levy property taxes and issue bonds, and it oversees sewers and waste management, protection of open spaces, and the development of sports facilities, and it also serves as the region's housing authority. The metro transit and airport commissions were placed under its authority. It likewise has land-use powers and, perhaps, most unusually, a tax-sharing plan among cities.^[17] The near metropolitan governments of the Twin Cities and Portland areas did not emerge out of local initiative and cooperation (any more than the Greater New York consolidation did), but were created by their respective state legislatures.

While this latter approach aspires to, and, to some extent, achieves, more regionwide equity (and perhaps efficiency), the one-tier consolidation approach tends

to give more power to suburban interests.^[18] Of course, this comes at the expense of city residents, as minority representatives have been quick to point out, which thus raises issues of representation in metropolitan areas.^[19] There is the "one-government, one-vote" side, and the "representation by population" side, with the former favoring suburbs and the latter cities. (A US District Court has ruled that the "one person, one vote" does not apply to COGs or other such voluntary organizations as they are not governments as such).^[20] This is another factor discouraging the creation of more actual metropolitan governments, as these would then come under the Supreme Court's "one person, one vote" dictum.

THE POLYCENTRIC METROPOLIS

This "tier" approach is the best shorthand for summing up the governmental side, but it obviously applies to very limited areas (counties) and cases (the Portland and Twin Cities metros). What prevails broadly across the rest of the country's metropolitan areas is what has increasingly been called "regional governance," with a focus on enhanced cooperation and collaboration among the countless "small republics." There is in American political culture a clear preference for such self-government and local autonomy, and it gained its academic legs with the "public choice" theory that emerged in the 1950s.^[21]

These "small republics" have been able to sustain themselves autonomously by a number of service arrangements like interlocal contracting. This is federalism writ small across metropolitan America: Citizens of a small city keep their autonomy by entering a contract (it is worth remembering that the concept of federalism comes from the Latin "foedus" for "compact"^[22]) with a larger city or the county for services they do not want to produce or provide themselves. The classic example of such contracting is the "Lakewood Plan." To avoid being annexed by Los Angeles, a number of suburban residents incorporated cities in the mid-50s, Lakewood the first of these, and bought services such as water and sewerage, among many others, from the county. As this illustrates, it bears repeating that in metropolitan areas real "public" choice only goes to some publics—those who can afford to choose in the suburbs—often practicing "exclusionary" zoning, and not those in the city.^[23]

One of the earliest studies of how such separate jurisdictions operate autonomously and at the same time deal cooperatively with metropolitan issues analyzed intergovernmental relations in the Quad cities in Illinois and Iowa.^[24] Here public officials and business leaders whose operations transcended local jurisdictions promoted cooperative governance in a limited

set of areas: they all share an airport in one city, a museum in another, a municipal art gallery in a third, and a hospital in a fourth. On the other hand, there is also the older “Tri-City Symphony” and a metro area minor league baseball team. The metro area business interests have promoted metropolitan government, but, of course, local officials (and residents) have shown no interest. Area residents prefer to retain their city identity, but they seem happy to identify with their metro area. Insofar as this may amount to metropolitan governance, it is limited chiefly to economic issues.

CONCLUSIONS

While metropolitan government advocates stress social equity, attempts at regional governance, as in consolidations, are more often driven by economic motives and spearheaded by area business leaders, chambers of commerce, non-profit organizations, and other such agents. As noted, business interests have thus become advocates for regional governance—they have recognized that the economic climate depends on a environmental factors such as clean air and water and “amenities” like easy commuting.^[25] Business interests have advocated metropolitan cooperation, but they have also promoted the creation of metropolitan area special purpose districts to address what they see as the primary issues (regional transportation and water districts in Georgia, for example).^[26]

Given the economic commons metropolitan areas represent,^[27] it would appear that, while more suburban governments will continue to be created, these economic pressures and other regional interests will promote increasingly cooperative governance. Many scholars see these pressures and interests leading to a “new regionalism,” in networks of business, non-profits, and governments, with positive efforts to address the environment, poverty and inequity, governmental inefficiencies, and regional competitiveness.^[28]

REFERENCES

- Stephens, G.R.; Wikstrom, N. *Metropolitan Government and Governance: Theoretical Perspectives, Empirical Analysis, and the Future*; Oxford University Press: New York, 2000; 16.
- Jackson, K.T. *Crabgrass Frontier: The Suburbanization of the United States*; Oxford University Press, 1985; 20–44, 87–115, 157–171.
- While we know “small republics” from the ancient Greeks on, the American concept is usually credited to Thomas Jefferson, cf. Wood, R. *Suburbia: Its People and Their Politics*, Houghton Mifflin: Boston, 1958; 25.
- Jackson, 190–218.
- Miller, D.Y. *The Regional Governing of Metropolitan America*; Westview: Boulder, CO, 2002; 94ff.
- Ross, B.H.; Levine, M.A. *Urban Politics: Power in Metropolitan America*, 6th Ed.; Peacock, F.E.: Itaska, IL, 2001; 400–401.
- Wikstrom, N. *Councils of Governments: A Study in Political Incrementalism*; Nelson-Hall: Chicago, 1977; 35–41.
- Wikstrom, 41.
- Wikstrom, 106.
- Ross; Levine, 372–374.
- Morcol, G.; Zimmermann, U.; Stich, B. The Georgia regional transportation authority: a smart growth machine? *Polit. Policy* **2003**, *31* (3), 488–508.
- Miller, 91.
- Judd, D.R.; Swanstrom, T. *City Politics: Private Power and Public Policy*, 4th Ed.; Pearson Longman: New York, 2004; 71–74.
- Feiock, R.C., Ed.; *Metropolitan Governance: Conflict, Competition, and Cooperation*; Georgetown University Press: Washington, DC, 2004; 5.
- Ross; Levine. 385.
- Stephens; Wikstrom, 24.
- Ross; Levine, 345.
- Ross; Levine, 346–350.
- Ross; Levine, 340.
- Harrigan, J.J.; Vogel, R.K. *Political Change in the Metropolis*, 7th Ed.; Pearson Longman: New York, 2003; 273ff.
- Wikstrom, 113.
- Tiebout, C.M. A pure theory of local expenditure. *J. Polit. Econ.* **1956**, *44*, 416–424; Ostrom, V.; Tiebout, C.M.; Warren, R. The organization of government in metropolitan areas: a theoretical inquiry. *Am. Polit. Sci. Rev.* **1961**, *55*, 831–842; (and numerous subsequent publications).
- Miller, 14.
- Wikstrom, 11; Saltzstein, A.L. *Governing America's Urban Areas*; Thomson-Wadsworth: Belmont, CA, 2003; 298–299.
- Friesema, H.P. *Metropolitan Political Structure: Inter-governmental Relations and Political Integration in the Quad-Cities*; University of Iowa Press: Iowa City, 1971.
- Kantor, Rosabeth Moss. Business coalitions as a force for regionalism. In *Reflections of Regionalism*; Katz, B., Ed.; Brookings: Washington, DC, 1999; 154–184.
- Morcol et al.
- Barnes, W.R.; Ledebur, L.C. *The New Regional Economies: The U.S. Common Market and the Global Economy*; Sage: Thousand Oaks, CA, 1998; 83–100.
- Savitch, H.V.; Vogel, R.L. A symposium: new regionalism and its policy agenda. *State Local Govern. Rev.* **2000**, *32* (3), 155–240; 363.

Metropolitan Governmental Fragmentation

Stephen J. Ziegler

*Public and Environmental Affairs, Indiana University/Purdue University,
Fort Wayne, Indiana, U.S.A.*

David C. Nice

Department of Political Science, Washington State University, Pullman, Washington, U.S.A.

INTRODUCTION

Metropolitan governmental fragmentation refers to the tendency for metropolitan areas in the United States to have large numbers of local governments, many of which overlap with one another. The typical metropolitan area has approximately 100 local governments in it, and larger metropolitan areas tend to have even more governments, most of which have only one or at most a few responsibilities. School districts are the best-known type of special district governments, but many other types exist. In addition, there are likely to be about two dozens municipal governments and, except in smaller metropolitan areas, no general-purpose local government that covers the entire metropolitan area.

DIMENSIONS OF METROPOLITAN FRAGMENTATION

Discussions of fragmentation in metropolitan areas usually emphasize the fragmentation of local governments, which are fragmented in two major ways: geographically and functionally.^[1-7] Geographic fragmentation involves having a multiplicity of the same type of local government, each of which covers only a portion of the metropolitan area. For example, a single metropolitan area may have 20 or more city governments or 12 or more school districts. Functional fragmentation involves the same geographical area served by several different local governments, some of which have different responsibilities. The downtown area may lie within a single city government, one or more school districts, a transportation district, a park authority, a sanitation district, and so forth. All of those governments are legally independent of one another. Because many metropolitan residents live in one part of the metropolitan area but work in another part and may shop and seek recreation in still other parts of the area, many people will be affected by the decisions of several dozen different local governments on a regular basis.

Metropolitan governmental fragmentation in the modern era has another important facet as well.

Many governmental programs and services are affected by the decisions of state governments and the national government. State governments are increasingly involved in virtually all programs that were once essentially local responsibilities, and the national government is involved in a host of programs, from health care to law enforcement. In addition, private organizations are involved in the delivery of many local services, often with some type of official relationship to one or more local governments.

Metropolitan areas are fragmented in a number of non-governmental ways as well, and the non-governmental fragmentation may be related to the governmental fragmentation. One important type of non-governmental fragmentation involves the economy. Many metropolitan areas display considerable economic segregation, with wealthy families tending to cluster together in some parts of the area and poor families clustered together in other areas. Instead of a single downtown business center, business activity is likely to be dispersed across many parts of the metropolitan area. Some of those economic centers may emphasize manufacturing, while others may emphasize retail sales, and still others, services such as health care. In addition, many businesses in a given metropolitan area are part of regional, national, or even international firms, which may cause many business leaders to have primary loyalties outside the metropolitan area.

Metropolitan areas are also likely to be fragmented socially in a number of respects. Different racial groups tend to live in different areas of a given metropolis, and some ethnic groups may have their own distinct enclaves. Gated communities restrict the access of “outsiders” to some parts of the metropolitan area, and inadequate public transportation can have the same effect on poorer citizens who may not have access to cars. Social life becomes segmented, as people live in one area, shop in other areas, and seek recreation in still other areas. As a result, many interactions take place among strangers, a development that is likely to weaken social controls on behavior and lead to more demands for governmental action.

These different types of fragmentation may reinforce one another. Governmental fragmentation appears

to encourage social and economic segregation, and economic segregation appears to discourage adoption of proposals to reduce governmental fragmentation. Whether and how governmental fragmentation should be reduced or managed are matters of considerable controversy.

PERSPECTIVES ON GOVERNMENTAL FRAGMENTATION

There are at least four perspectives on governmental fragmentation in metropolitan areas: classical administrative, public choice, class conflict, and the intergovernmental perspective.^[2-11] The classical administrative perspective argues that too many local governments exist in the United States. Because of their size, most small governmental entities are unable to reap the benefits of economies of scale and, thus, waste resources due to their inability to acquire sophisticated equipment and hire experienced, well-trained personnel. Although cooperation among local governments would likely reduce the amount of wasted resources, municipalities rarely cooperate with each other unless such cooperation is mutually beneficial and the policy area is noncontroversial. Analysts taking a classic administrative perspective argue that the overlapping nature and virtual invisibility of local jurisdictional boundaries make it difficult for citizens to determine which governmental official is responsible for what.^[5,10,11] In addition, fragmentation encourages neglect of spillover effects and often produces some governments with many needs and few resources and other governments with few needs and many resources. Coping with problems encompassing the entire metropolitan area is hampered by disagreements among the many local governments.^[5]

In contrast to the classic administrative perspective, the public choice perspective views the myriad of local governments as beneficial. Public choice analysts view local governments like competitors in a free market, competing for citizens instead of purchasers. For example, in an effort to attract students, one school district may emphasize one approach to education that is in stark contrast to the approach of an adjacent school district. This competition among districts is seen as healthy and, at times, Darwinian—school districts that fail to satisfy their customers (taxpayers) will experience a loss of students and the accompanying tax revenue. However, supporters of the public choice perspective also realize that although small local governments often waste resources or lack a tax base to support a multitude of city services, large jurisdictions can also be inefficient and experience greater costs per unit. Moreover, not all types of services have the same cost-size ratio; some services can be provided cheaply

in a large jurisdiction or be provided via a service contract to a smaller jurisdiction. In fact, some argue that smaller jurisdictions may be inherently more efficient because of the increased ability to access the proper officials, provided, of course, that the identity of that official can be determined in light of the overlapping jurisdictional boundaries. In the end, public choice theorists argue that the benefits of small local governments far outweigh the problems.^[2,5,8,9,12,13]

Although separate, the class conflict perspective is related in some ways to the classical administrative and the public choice perspectives. The public choice perspective argues that local governments compete for citizens in a free marketplace and, as a result, attract the “best” taxpayer (customer). In contrast, the class conflict view of interlocal relations recognizes that not all citizens are permitted to shop for the best community. For example, some communities are able to effectively exclude people of modest means by establishing particular zoning regulations or building codes that only the wealthy could afford to comply with. Consequently, communities composed of large, expensive, single-family homes enjoy a large tax base with a low tax burden, enabling them to enjoy whatever city services they desire. Moreover, these communities continue to thrive because of their ability to attract people and businesses with ample resources and few needs. Even if residents of the inner city had at least the legal eligibility to move into the more affluent suburbs, the historical effect of previously enforceable race-restrictive covenants or realtors who steer non-Whites away from particular neighborhoods, as well as the absence of inexpensive housing, has the effect of perpetuating segregation. Therefore, families who are unable to acquire housing in the more affluent communities are forced to remain in environments where the need for services remains high and their ability to pay for those services, low. In addition, economic opportunities grow scarcer. Consequently, businesses and the more affluent are able to leave, yet the poor must remain—their problems compounded by the lack of governmental consolidation (fragmentation). In addition, when business interests cannot find a suitably cooperative local government, they often are able to create a new one. In the end, it is only the affluent whom governments seek to attract, not the poor.^[2-5,11]

Finally, the intergovernmental perspective offers a fourth view of interlocal relations, which is essentially concerned with how metropolitan areas became so fragmented and what implications fragmentation has for interlocal relations. First, fragmentation stemmed from the confluence of several factors over time. Although changes in lifestyle and the desire of the more affluent to leave the central city and locate in more spacious and personal communities were ample reasons, fragmentation largely stemmed from the state

policies regarding consolidation, annexation, and incorporation. Before the beginning of the 1900s, annexation by local governments was relatively easy; as its citizens moved out of the city, annexation followed them. However, by the 1920s and 1930s, changes in state laws contributed significantly to a reduction in annexation, much to the relief of rural residents who had long reaped the benefits of a malapportioned state legislature that effectively diluted the vote of city residents. As the cities were unable to annex land as before and keep up with the population, the cities fragmented, new municipalities became incorporated, and the fragmented metropolis began to form. In fact, research indicates that one of the primary factors that contributes to urban fragmentation is the existence of tension between states and their cities. In short, non-metropolitan environments fear the influence of large cities and, as that fear increases, so does the tension within the states and likelihood of fragmentation. This same intergovernmental perspective supports the likelihood that special districts will result when state governments increase controls on a city's financial power. In light of the foregoing, the implications of fragmentation on intergovernmental relations are multiple. The nature of fragmentation separates people and places. Instead of one consolidated, unified polity, fragmentation leads to marginalized groups, segregated by race and class. Consequently, the more fragmented, the less cohesion among groups, and the less these groups would be willing to cooperate and resolve the many interjurisdictional problems that arise. Moreover, consistent with the class conflict perspective, some problems are inherently confined within certain jurisdictions. Thus, problems confined to one particular segment or jurisdiction within a metropolitan area are less likely to be resolved by the community as a whole, and, therefore, resolution must be accomplished by higher levels of government, leading to a reduction in autonomy between the metropolis and state and federal governments.^[2,5]

REDUCING OR COPING WITH GOVERNMENTAL FRAGMENTATION

For those who believe that large numbers of local governments in a metropolitan area are harmful, there are a number of possible remedies. They include informal cooperation, service contracts, councils of government (COG), special districts, annexation and consolidation, county-based solutions, federated metropolitan government, and non-local solutions.^[2-7,10] Although various strategies have been proposed and implemented, the success of these efforts varies. For instance, attempts have been made among fragmented jurisdictions to participate in informal cooperation. Certainly

the least intrusive method, jurisdictions have cooperated to resolve mutual problems in such areas as traffic management or exchanges of information among police agencies. However, if informal agreements are ever reached, most are in noncontroversial areas and have a short shelf life, lasting only so long as particular personnel remain in key positions.^[5]

A second approach to solving the problems associated with fragmentation is service contracts. Service contracts are often attractive, enabling local governments or private individuals to purchase from another government, a private firm, or a non-profit agency such services as water, fire protection, or even trash collection. Los Angeles County is one of the nation's leaders in providing contractual services under agreements that are legally binding. Similar devices are widely used by county and city governments all over the United States. However, service contracts require mutual agreement among the parties and will only work if the community or individual is able to pay for the services. Lack of agreement is often the by-product of fragmentation, particularly when dealing with controversial issues and policies. Local officials have sometimes encountered problems, due to inadequately written contracts and inadequate contract enforcement. Adding large numbers of private contractors can slow the implementation of major policy changes and can greatly complicate the task of determining who is responsible for results.^[2,5,14]

A third strategy to overcome fragmentation-related problems is COG. COGs are essentially a local version of the United Nations. Representatives from each member community meet to iron out problems, resolve conflicts, and coordinate their activities. These relationships were fostered years earlier, when federal grant monies were made available to local COGs and necessitated cooperation. However, like the United Nations, agreement among members has proven difficult, particularly when they address controversial issues and policies.^[2,5,10]

Special Districts

Special districts represent another approach when dealing with several local governments. These districts, e.g., school and water districts, are specific governmental entities that have the ability to raise taxes, assess user fees, and sell bonds and use the revenue to finance services they provide. Many of these districts encompass several smaller local governments like municipalities and, as a result, provide some sort of coordinated effort when dealing with a particular issue and reaching binding decisions. However, although this approach is able to handle one type of fragmentation over a particular geographical location, by the

nature of its individual function and purpose, it creates a different type of fragmentation: functional fragmentation. Thus, citizens could have different governmental units responsible for water, education, and even a sports arena, and, consequently, this leads to more government, more agents, and more effort expended to ascertain their identity and hold them accountable.^[5,10]

Annexation and Consolidation

These two related approaches are of the more drastic variety. A city can expand its geographical boundaries by annexing unincorporated areas or by consolidating with other municipalities to form one consolidated government. As a result of unification, these new metropolitan governments can take advantage of economies of scale, coordinate efforts, and offer increased visibility. Because efforts to consolidate government shift decision making from several local governments to one centralized government, resistance to this approach is common. Oftentimes, attempts to consolidate or annex are limited by feasibility. Local elites fear the loss of their power and influence, residents fear dilution of their votes, and affluent areas fear an increase in taxes to remedy problems in the central cities, which may have initially contributed to the fragmentation of government in the first place.^[5,10]

County-Based Solutions

Two county-based solutions aimed at solving the problems of fragmentation concern the transfer of some responsibilities from multiple city governments to the county and a merger between the cities and counties. For example, a county may include 10 different police agencies. Under a transfer approach, the county would be responsible for law enforcement service. Economies of scale could be realized due to standardization of procedures, personnel, and equipment. The second county-based approach is even more drastic: Cities within the county consolidate as a single political unit. This has occurred in Jacksonville, Indianapolis, Philadelphia, and Louisville, as well as some other communities. As with the annexation and consolidation approach noted earlier, county-based approaches are often difficult to enact (especially when a multicity merger is proposed). Moreover, many metropolitan areas fall within two counties or more (two governmental units are certainly preferable to 20 in the minds of many reformers). Finally, the structure of county government serves as a limitation for city-county mergers. Many county governments lack sophisticated personnel systems and executive leadership and are accustomed to working with local governments in a supplemental instead of primary

capacity. This seems to be less true of some of the more populous counties, however.^[2,5,10]

Federated Metropolitan Government

As the name implies, this fragmentation reform approach is essentially a miniature federal system. One level of government handles local responsibilities, while the other level handles area-wide responsibilities. Currently, the government in place at Toronto, Canada, illustrates this approach and enables attention to local concern, while, at the same time, permitting economies of scale and consolidation of efforts across the greater metropolitan area. The primary difficulty in this approach is its acceptance. Federated metropolitan governments are rare in the United States, and when they exist, the inherent conflict about what level of government should handle what matter is commonplace, as in the Miami-Dade County Metropolitan plan. In the end, the approaches that have the greatest likelihood of overcoming fragmentation are the least likely to be adopted. This is, in large part, the result of the original conflict that created fragmentation, a hyperpluralistic approach to the creation and maintenance of local government.^[5]

Non-Local Solutions

Finally, if effective solutions to the problems associated with fragmentation are not feasible at the local level, the solution may lie at the state or national level. For instance, grants in aid to poorer jurisdictions may help resolve the inequality of needs and resources evident in the class conflict perspective. Alternatively, a higher level of government could provide the service currently provided by a multiplicity of jurisdictions, such as a state agency vested with the authority and responsibility for education (instead of individual school districts, where disparities exist on account of their location and tax revenue sources). However, even if a higher-level government agency took a leadership role in, say, education, this approach would only serve to ameliorate the current situation, because the existing districts would remain. Moreover, higher levels of government may not always be effective in targeting assistance to where it is most needed and may not be sufficiently sensitive to local needs.^[5]

FUTURE PROSPECTS

Most of the available evidence indicates that governmental fragmentation in most metropolitan areas is likely to persist for the foreseeable future. The more drastic proposals to reduce governmental fragmentation are almost always defeated in the United States,

particularly if local residents must approve the proposal. Moreover, special district governments (other than school districts) continue to proliferate, and they bring serious problems of accountability in light of their low visibility to the general public. However, special district governments, with responsibility for several area-wide services, may offer some hope for coping with metropolitan fragmentation in a politically feasible manner, especially with some encouragement from state governments.^[7] Whether that encouragement can be found on a dependable basis remains to be seen.

New developments in information technology and the global economy have significant implications for metropolitan fragmentation, although their full implications are not altogether clear at this point.^[15] New information technologies may help improve some aspects of central city life, such as coping with traffic congestion, but the same technologies may also encourage more movement to the metropolitan periphery and even beyond it. The continuing development of the global economy may further erode the attachments of businesses to particular locations and, as a result, increase the economic vulnerability of individual localities and decrease the attachment of business leaders to the communities where they reside. Fostering a sense of community and developing future generations of local political leaders may grow increasingly difficult in many communities.

CONCLUSION

It has often been said that all politics are local, although getting a clear sense of the meaning of "local" is often difficult. This is certainly the case in many large metropolitan areas where hundreds of local governments and special districts abound. Although these governmental entities were created in the interest of being more responsive to its citizenry, it has nevertheless resulted in fragmentation, overlapping jurisdictions, and reduced accountability. However, that is not to say that governmental fragmentation has been all bad. Redundancy is a common way to reduce error, particularly in light of the fact that democracy has been regarded as a form of government for

people who are not exactly sure that they are right and who may not trust public officials very much.

REFERENCES

1. Bardo, J.; Hartman, J. *Urban Sociology: A Systematic Introduction*; P.E. Peacock Publishers, Inc.: Itasca, IL, 1982.
2. Bollens, J.; Schmandt, H. *The Metropolis: Its People, Politics, and Economic Life*; Harper and Row: New York, 1982.
3. Burns, N. *The Formation of American Local Governments*; Oxford University Press: New York, 1994.
4. Downs, A. *Opening up the Suburbs*; Yale University Press: New Haven, CT, 1973.
5. Nice, D.; Fredericksen, P. *The Politics of Intergovernmental Relations*, 2nd Ed.; Nelson Hall: Chicago, IL, 1995.
6. Ross, B.; Levine, M. *Urban Politics: Power in Metropolitan America*, 6th Ed.; P.E. Peacock Publishers, Inc.: Itasca, IL, 2001.
7. Stephens, G.R.; Wikstrom, N. *Metropolitan Government and Governance*; Oxford University Press: New York, 2000.
8. Bish, R.; Ostrom, V. *Understanding Urban Government*; American Enterprise Institute: Washington, DC, 1973.
9. Ostrom, E. Metropolitan reform: propositions derived from two traditions. In *Readings in Urban Politics*, Hahn, H., Levine, C., Eds.; Longman: New York, 1984; 329–347.
10. Herson, L.J.R.; Bolland, J.M. *The Urban Web: Politics, Policy, and Theory*; Nelson Hall: Chicago, 1997.
11. Harrigan, J.J.; Vogel, R.K. *Political Change in the Metropolis*, 6th Ed.; Addison-Wesley Education Publishers, Inc.: New York, 2000.
12. Dye, T. *American Federalism*; Lexington Books: Lexington, MA, 1990.
13. Ostrom, V. *The Meaning of American Federalism*; Institute for Contemporary Studies: San Francisco, CA, 1994.
14. Agranoff, R.; McGuire, M. *Collaborative Public Management*; Georgetown University Press: Washington, DC, 2003.
15. Ruchelman, L. *Cities in the Third Wave*; Burnham: Chicago, 2000.

Microenterprise Finance and Economic Development

Douglas R. Snow

School of Public Management, Suffolk University, Boston, Massachusetts, U.S.A.

INTRODUCTION

Microenterprise finance refers to strategies designed to create or expand small businesses that are unable to attract financing from commercial sources. Non-profit microfinance institutions (MFIs) have been established by both non-governmental organizations (NGOs) and governments to channel capital to microenterprises. Funding for MFIs comes primarily from foundations, governments, the World Bank, and MFI operations. Donors hope that lending by MFIs, commonly called microcredit, will expand local economies in disadvantaged areas, increase household income of borrowers, and create a base of entrepreneurs for future development.

Strategies to increase the supply of capital to microenterprises began over 30 years ago. Realizing that centrally planned development efforts funded by international development agencies were not making the poor better off, NGOs began developing strategies to bring small loans to poor entrepreneurs in the informal sectors of developing economies. In 1973, ACCION International may have been the first NGO to use microcredit as a development strategy when it began making small loans to businesses in Brazil.^[1] Loans from MFIs, often as small as \$50, have been used to finance a wide range of income-producing activities, including produce stands, curbside repair shops, and milk and egg production. Microenterprise finance received its first big public relations boost in the early 1980s, with the widely reported successes of Grameen Bank in Bangladesh. Grameen Bank, founded by Professor Muhammad Yunus, began lending in 1976.^[2] The Grameen model, based on small peer groups who approve and guarantee each other's loans, is now copied throughout the developing world. In addition to lending, many MFIs offer business management training, savings programs, literacy, and health care programs.

Since the early 1990s, microenterprise finance has become an important strategy for economic development in Europe and the United States. Internationally, another boost came in 1995, with the creation of the Consultative Group to Assist the Poorest (CGAP).^[3] Twenty-nine donors—development agencies of industrialized nations, multilateral development agencies, and private foundations—formed the new organization (see Table 1). In 1997, a Washington, DC, conference attracted 2900 people from 137 countries. The conference, the Microcredit Summit, set a goal

to reach 100 million of the world's poor by 2005. The Microcredit Summit Campaign emerged from the conference as an organization that collects data on MFIs, tracks progress toward reaching the campaign's goals, and promotes best practices research.^[4]

Microenterprise Finance Defined

Strategies for financing microenterprises are commonly called *microcredit*, a generic term for the diverse strategies employed to bring capital to microenterprises. Other terms in common use are *microlending* and *microfinance*. All have one thing in common, however: loans are made to entrepreneurs who are otherwise unable to access commercial banking services. The growth of microcredit as a development strategy has created a broadening of definitions. In developing countries, microcredit borrowers are generally poor women living in urban or rural areas with very small enterprises. In recent years, microcredit programs have expanded to former newly independent countries of Eastern Europe and to areas of poverty in industrialized countries. Microcredit is now a common economic development strategy in urban enterprise zones in the United States. Microcredit programs now serve diverse clienteles and enterprises throughout the world.

One of the most comprehensive definitions of a microenterprise comes from the Organization for Economic Cooperation and Development (OECD). In a 1996 report, OECD defined a microenterprise as

A sole proprietorship or family business with fewer than 10 employees (most often owner-operated); and requiring relatively small amounts of capital or credit financing (less than US\$25,000). In addition, microenterprises include part-time or self-employment activities to supplement family income.^[5]

Today, microcredit clearly involves more than just the very poor.

TYPICAL MICROCREDIT STRATEGIES

Microcredit programs are sponsored by a diverse array of institutions. Large NGOs franchise local programs that function independently using a standardized methodology. In the United States, ACCION International's

Table 1 CGAP member nations

Nations		Multilateral development agencies	
Australia	Japan	African Development Bank	International Labour Office
Belgium	Luxembourg	European Bank for Reconstruction and Development	United Nations Development Program
Canada	Netherlands	International Fund for Agricultural Development	United Nations Capital Development Fund
Denmark	Norway	European Commission	United Nations Conference on Trade and Development
Finland	Sweden	Inter-American Development Bank	World Bank
France	Switzerland		
Germany	United Kingdom		
Italy	United States	Argidius Foundation	Ford Foundation

Source: Consultative Group to Assist the Poorest. Member Donors; http://www.cgap.org/html/ac_members_donors.html (accessed March 2002).

franchises partner with local non-profit and government organizations such as the Small Business Administration and the Service Corps of Retired Executives.^[6] In developing countries, microcredit programs are operated by governments (using funds from bilateral development agencies such as USAID) and by NGOs. Sometimes the microcredit programs are extensions of existing credit unions or village associations.

The lending methodology of MFIs varies, but the most popular, and the one most familiar to casual microcredit observers, is the group-lending model popularized by Grameen Bank. The group model is designed to reduce credit risk. Because the poor have little or no experience with formal credit, they lack credit history. Even with credit history, their poverty would deny them access to commercial credit. A “typical” program involves a small group of women (six or seven) formed to vet business plans and guarantee the credit of other group members. When a loan is secured, the borrower makes payments at weekly meetings. If the member defaults, all other members lose access to credit. Because of the risk of losing credit, members are careful to only approve good business plans. Peer pressure assures repayment. Repayment rates are high, approaching 100% in some programs. The loans are very small by the standards of industrialized countries, in the neighborhood of \$100. The term of the loan is usually short and measured in weeks. Interest is high. Monthly rates of 3% to 4% are not uncommon.^a

In many African countries, a tradition of savings and rotating credit predates the colonial experience. Some NGOs have been able to use rotating savings and credit associations (ROSCAs) as a tool to implement microcredit programs. In ROSCAs, women bring savings to each meeting. The women take turns

withdrawing savings from the ROSCA and repay their loans at future meetings. Microcredit programs take advantage of existing ROSCA groups to infuse new capital into an area. Many microcredit programs require their members to save. In a peer group model, savings are brought to each meeting and are sometimes deposited with a local bank. In lieu of savings, some programs require members to bring a small amount each week to insure their loans, removing some of the risk of the group model. Some programs include training and education with lending. Members participate in training in advance of receiving loans. Other social services may also be offered.^[7]

Grameen Bank “replications” are not always successful. The peer group model is vulnerable. Default by one member may create a domino effect as other members realize that their access to credit is about to end. Without future access to credit, why repay the loan? It is also possible for group members to cheat by repaying loans in one microcredit program with loans from another source, possibly family members, another MFI, or local loan sharks. The inevitable day of reckoning brings default.^b

In the United States, the peer group lending model is used much less than it is in developing countries. Microfinance institutions in the United States that do use the peer group methodology report much lower average loan sizes than those MFIs that do not, \$1802 and \$10,631, respectively.^[9]

SIZE OF THE MICROENTERPRISE INDUSTRY

Documenting the size of the microcredit industry is a daunting task. But by any count, microcredit is a large industry. The Microcredit Summit Campaign (MSC) reports that between 1995 and 2000, over 1500 MFIs reached 30.7 million clients. Most of the MFI clients

^aThis is a composite example based on the group lending methodologies of The Foundation for International Community Assistance (FINCA), <http://villagebanking.org> (accessed March 2002) and ACCION International, <http://accion.org> (accessed March 2002).

^bFor a discussion of defaults, see Ref.^[8].

reported to MSC are in Asia (Table 2). The largest MFI reported by the campaign is Grameen Bank, with over 2.3 million clients as of calendar year 2000.^[4]

In the United States, the *1999 Directory of Microenterprise Programs* reports 341 microcredit programs in 46 states and the District of Columbia. These programs have served over 250,000 clients and dispensed \$160 million in loans since start up.^[10] The U.S. Small Business Administration has added microcredit to its small business programs. Community Development Agencies have broadened their activities to include microcredit. Many banks either provide direct funding or indirect support through lines of credit or other banking services. Microcredit programs operate in urban enterprise zones as part of both state and federal urban development strategies.^[11] The Charles Stewart Mott and Ford Foundations, as well as the Aspen Institute, fund, evaluate, and publicize microcredit research and/or programs.

MICROCREDIT AS PUBLIC POLICY

Microcredit has its origins outside the control of governments and large international development institutions such as the World Bank, United National Development Program, and United States Agency for International Development. During the 1990s, microcredit grew into a popular economic development strategy in developing countries. This was largely due to the popular opinion that the economic development projects begun in the 1960s had failed. Industrialized nations and international financial institutions now accept microcredit as an important element of development policy. Success has been so complete that microcredit is also an accepted strategy for economic development in OECD countries.

Microcredit strategies are heavily debated among scholars and practitioners. The entry of large multilateral institutions has contributed to the vast microcredit literature. Schools of thought have formed around the

proper role of MFIs. Should MFIs become self-sustaining institutions only, recovering all their costs through fees and interest, or should they provide social intermediation, acting as more comprehensive social welfare institutions? The reality is that MFIs do not neatly fall into either camp. Nevertheless, the arguments have important public policy implications. The very standard upon which MFIs will be evaluated depends on expectations of what the MFI is supposed to achieve.^c

The argument for self-sustaining MFIs is grounded in the fear that cheap credit will distort local markets, leave the poor burdened with debt that they cannot pay, and ultimately lead to failure of MFIs when donor subsidies turn to other development strategies. The poor have shown that they can save money and repay loans with high interest rates. Microfinance institutions that recover all their costs through fees and interest become self-sustaining and part of the local financial infrastructure. One highly publicized success is that of BancoSol, which began as a subsidized MFI in Bolivia and later became a self-sustaining bank in 1992. Self-sustaining MFIs should, according to this school of thought, graduate to commercial banks. In developing countries, especially, this deepens the financial infrastructure and brings more economic activity into the formal economy. Building sustainable institutions will benefit the poor in the long run and avoid the disheartening effects of “flash in the pan” projects.

Advocates of social intermediation argue that microcredit will be judged on its ability to improve the condition of the poor. For this reason, many MFIs offer far more than credit. Training in writing business plans and management is common. But in developing countries, public health and education may supplement services offered by MFIs. Those who advocate social intermediation approaches to microcredit fear that true self-sustaining MFIs may have to serve a population that is better off in the first place, leading to goal displacement. Reaching the very poor and yet recovering all costs through fees and interest are incompatible.

Debate over the appropriate role for microcredit will continue. But while the debate has raged, microcredit has continued to grow and diversify. An important trend is the linkage of microcredit programs with both not-for-profit and for-profit organizations. Commercial banks offer savings to microcredit borrowers and letters of credit to MFIs. An MFI in the United States may have relationships with commercial banks, non-profit organizations that provide management training, and governmental agencies.^[13] Through linkages, MFIs become important components of urban and rural economic development strategies, not just stand-alone operations. Looking at microcredit from

Table 2 Microcredit clients reported by Microcredit Summit Campaign by region

Region	Clients reported in 2000	Percentage of total
Africa	5,180,881	16.9
Asia	23,576,938	76.8
Latin America and Caribbean	1,672,541	5.5
Middle East	54,272	0.2
North America	111,204	0.4
Europe	85,271	0.3
Total	30,683,107	100.1^a

^aDoes not total 100% due to rounding.

Source: Adapted from Ref.^[4].

^cFor a complete discussion of the microcredit debate, see Ref.^[12].

this point of view, sustainability may be better viewed from the point of view of an overall economic development strategy, not just the MFI itself.

Microcredit is generally perceived as a “teach a person to fish,” rather than a “give a person a fish,” development strategy. Proponents believe that providing the means for the poor to own businesses and increase their incomes will lead to strong and prosperous communities. However, evaluation has tended to focus on the design of microcredit programs, so-called “best practices,” as opposed to actual outcomes for program participants. Fortunately, evaluations of microcredit programs are maturing. Research funding by the World Bank, USAID, and philanthropic foundations is likely to produce better research designs. A promising evaluation approach involves measuring the overall effects of microcredit borrowing on household wealth and income. Although these evaluations are needed, they are not without their challenges. Microcredit participants are self-selected, so true comparison groups cannot be found. Separating increases in wealth and income due to microcredit from other factors is difficult.^d

CONCLUSION

Microcredit has become a well-accepted economic development strategy. Once targeted at the very poor in developing countries, microcredit has become an important tool in the United States and other industrialized countries. The basic premise of microcredit has not changed: given access to credit and training, the poor are able to develop businesses, increase their incomes, and expand the economic base of their communities. However, the expansion of microcredit as an economic development strategy has led to diversification in the field. The very poor are no longer the only target of microcredit programs. This has led to considerable debate over the proper role of microcredit in economic development.

Research on microcredit has become more sophisticated with the increased interest of large development institutions and foundations. In an area once dominated by NGOs, governments are now playing a larger role. Evaluation of microcredit programs must now take into account multiple goals advocated by a complex set of political actors.

ACKNOWLEDGMENTS

Past collaborations with Professors Terry Buss and Gary Woller contributed greatly to the author's knowledge on the subject of microcredit.

REFERENCES

1. ACCION International: Boston, Massachusetts, *ACCION's History*; <http://www.accion.org/about/main.asp> (accessed March 2002).
2. Grameen Bank: Dhaka, Bangladesh, *Statistical Updates*; <http://www.grameen.org> (accessed March 2002).
3. World Bank, Consultative Group to Assist the Poorest: Washington, DC, *History*; http://www.cgap.org/html/about_cgap.html#Overview (accessed March 2002).
4. Druschel, K.; Quigley, J.; Sanchez, C. *State of the Microcredit Summit Campaign: Report 2001*; Microcredit Summit Campaign: Washington, DC, 2001; 2, 12.
5. Organization for Economic Cooperation and Development. *Microcredit in Transitional Economies*; OECD: Paris, 1996; 8 pp. OCDE/GD (96) 40.
6. Snow, D.R.; Buss, T.F.; Dumas, C. Sustaining microcredit programs: lessons learned over two decades of practice. In *Creating Sustainable Community Programs*; Daniels, M., Ed.; Praeger: Westport, CT, 2001; 95–113.
7. Snow, D.R.; Buss, T.F. Development policy and the role of microcredit. *Policy Stud. J.* **2001**, 29 (2), 296–307.
8. Bhatt, N.; Tang, S.-Y. Delivering microfinance in developing countries: controversies and policy perspectives. In *Microcredit and Development Policy*; Snow, D.R., Woller, G., Buss, T.F., Eds.; Nova Science Publishers: Huntington, New York, Chapter 4.
9. The Aspen Institute. *Highlights from the 1999 Directory of Microenterprise Programs*; Aspen Institute: Washington, DC, 1999; <http://www.fieldus.org/directory/index.html> (accessed March 2002).
10. The Aspen Institute. *Highlights From the 1999 Directory of Microenterprise Programs*; <http://216.147.121.117/directory/highlights.html>, linked at aspeninstitute.org (accessed March 2002).
11. Snow, D.R. Strategic planning and enterprise zones. *Policy Stud. Rev.* **2000**, 17, 13–28 (Summer/Autumn).
12. Woller, G.; Dunford, C.; Woodworth, W. Where to microfinance? *J. Int. Econ. Dev.* **1999**, 1 (1). www.spaef.com (accessed March 2002).
13. Snow, D.R. Microcredit: an institutional development opportunity. *J. Int. Econ. Dev.* **1999**, 1 (1). <http://www.spaef.com> (accessed March 2002).
14. Barnes, C.; Morris, G.; Gaile, G. An assessment of microfinance service impacts in Uganda. *J. Int. Econ. Dev.* **1999**, 1 (1). <http://www.spaef.com> (accessed March 2002).
15. McNelly, B.; Lippold, K. Practitioner-led impact assessment: a test in Mali. *J. Int. Econ. Dev.* **1999**, 1 (1). <http://www.spaef.com> (accessed March 2002).

BIBLIOGRAPHY

1. Servon, L.; Bloustein, E.J. *Structuring and Sustaining the Relationships that Support U.S. Microenterprise Programs*; International Labour Organization: New York, 2000.

^dTwo examples of evaluations are summarized in Refs.^[14,15]

Middle East and Regulation

Ori Arbel-Ganz

Department of Political Science, Bar-Ilan University, Ramat-Gan, Israel

INTRODUCTION

There is almost no doubt that regulation has become the most prominent expression of those involved in democracy and global capitalism. Starting in the United Kingdom during the middle of the 18th century and continuing in the United States, regulation policy has spread during the last decades to Australia, New Zealand, Latin America, Western and Eastern Europe, Scandinavian countries, the Far East, and also to the Middle East.

Complex questions arise at the heart of this topic, such as the power of regulators, their institutional arrangements, and discretion and doctrine, while the political agenda also criticizes the state becoming a *regulatory state*.

Theoretical Framework

Within the public policy context, regulation refers to any action taken by the government to introduce order into the activities undertaken within its sphere of responsibility and authority. From a broader perspective, regulation pertains to the government laws and instructions that order the exchange of goods and services.^[1] Others have narrowed the definition by focusing on the executive branch, to a process or activity in which an executive government branch sanctions, prevents, or prohibits the performance of specific individual, institutional, or organizational actions or behavior, and do so under a long-term administrative regime.^[2] From a functional point of view, some have argued that the factor common to all definitions of regulation lies in their inclusion of three distinctive factors: 1) *determination of norms*—formal decisions that set standards based on professional skills, and that prescribe or prohibit specific behavior; 2) *control*—the outcome of application of standards or directives, meaning limitation of the number of actors and/or the arena for their permitted freedom of action; and 3) *monitoring*—the enforcement of standards and directives.^[3] Although regulatory agencies vary considerably with respect to the internal balance between these three characteristics, there is consensus for the priority of determination over control and monitoring. The main task of all regulatory agencies is to

structure specific domains, which they do by initially determining what norms are required and only afterwards by enforcing norms and creating subsidiary norms.

Others view the setting of norms as *ex ante* crisis prevention and enforcement of norms as *ex post* crisis solution.^[4] Some scholars differentiate between regulation's cautionary (red light) functions and its authorization (green light) functions. That is, the exercise of regulation narrows an individual's (or organization's) freedom; by doing so it sustains the freedom available to the public or other organizations.^[5,6] Also some point to three layers of regulation: 1) targeted rules; 2) all modes of state intervention in the economy; and 3) all mechanisms of social control by whosoever exercised.^[3,6] While regulation policy must be executed by rule, and since whatever regulator does normatively must be referred to any value, we may define it as an *authoritative arrangement of exchange between society's values*.

Regulation—normative aspects

In democratic societies, citizens waive some of their rights by delegating them to the government in order to fulfill their rights. However, the norms maintained and the rights awarded the individual are many; hence, the norms activated by government agencies to protect citizens frequently conflict with one another, especially on economic exchanges. Therefore, the essence of regulatory policy was originally involved in correction of market failures. Market failures represent those situations where the equilibrium between supply and demand does not maximize a society's production potential.^[7,8] As a result, the freedom to maintain exchange relations between the individual and the society is not sustained,^[9] as revealed in the presence of monopolies (or cartels), information asymmetric, externalities, public goods, and circumscribed property rights and ownerships.^[10] From this perspective, regulation is a series of competitive rules meant to overcome or at least minimize the negative effects of market failures or, alternatively, to prevent their appearance so that the market will continue to subsist as a leading society mechanism.

Market failures will be controlled by economic regulation that monitors prices of goods and services,

controls the amount of goods and services available, and determines entry and/or exit barriers and/or incentives to specific industries.^[11] Some scholars have made distinctions within theories of economic regulation between 1) *self-regulation*—meaning the determination of norms and rules by the firms themselves; 2) *regulation of competition*—referring to a policy of maintaining branch stability by industry regulation; and 3) *metaregulation*—the integrated enforcement by regulatory agencies of the regulatory regime's competitive laws through industry as well as general market self-regulation.^[8]

Nevertheless, regulation as a policy tool was also introduced to improve social well-being, that is, to increase or more equitably distribute social resources.^[12] Stated differently, regulation that does not promote competition for this purpose will not further individual freedom, but it will directly provide values and rights that the market did not provide, provided partially, provided inadequately, or conversely prevented the individual from acquiring those values and rights.^[3] Thus, social regulation focuses primarily on issues of environmental quality, health, and education as well as the safety of products and services. One of the main characteristics of these domains is the inability to accurately price their respective externalities, since the supply-and-demand equilibrium considers only in part, if at all, changes in the equilibrium of secondary markets resulting from the externalities originating in the primary market. Whereas, economic regulation evolved to support competition as a social mechanism, social regulation controls the negative impact of competition in domains lacking markets.^[13]

Some scholars also differentiate between *paternalistic* regulation, which focuses on domains in which individuals tend not to take responsibility for their actions,^[3] and *procedural* regulation, which focuses on administrative and production processes operating in the private sector for the purpose of the reducing consumer costs and preserving public trust in the economic systems active in the market.^[14]

Regulation—positivistic aspects

According to *pluralist theory* and its discussion of interest groups, regulatory policy reflects the outcome of reciprocities between social groups and government entities. The regulatory-making process therefore indicates that an interest group managed to best exploit its power vis-à-vis that of the governing agency at any given place or time in order to further its own interest.^[15]

The *institutional approach* views regulation as the outcome of an interinstitutional system of governing authorities. It focuses on the reciprocities between the

public agencies, laws, procedures, and rules that invite exchange. The same perspective is applied internally with respect to the organization's rules, culture, and policy outputs. Therefore, if the institutional systems found in different states themselves vary, the regulatory policies formulated when confronting the same market failures will also vary.^[16]

From the *private interest approach*, which focuses on the individual decision maker, the individual regulator employs regulation as a means to further personal interests under the umbrella of public interests or a particular pressure group.^[17] Members of this school view regulators as none other than *captured* by the firms to be regulated. It is therefore quite understandable that in anticipation of employment in a senior position in one of those firms, the regulator will translate his expectations into the easing of restrictions for the regulated firms and the exchange of his power into some form of future reward.^[18]

According to the *ideological approach*, regulation reflects philosophical conception of reciprocities maintained between society and the state.^[12,19] Policy makers essentially formulate regulatory policy prior to their election to senior office, and their personal worldview regarding the limits of the state when intervening into the market is what dictates their satisfaction with regulatory policy in general.

According to the *diffusion approach*, regulation spreads through scholars' networks, experts, bureaucrats, social networks, and via personal until its adopted by states, authorities, markets, and institutional procedures. As with all other diffusions, so did regulation appear over time as an "S" diagram: at the beginning, only a few players had adopted it, and by the second and third stages, most countries had implemented it as main practice of government and society.^[20]

Regulation—historical review

The perception of regulation as a distinct policy exhibiting unique characteristics was born in Britain in the middle of the 19th century. The federal government in the United States began to deal with regulatory profit policy in the last decade of the same century. Most democratic countries in Western Europe began to contend with regulation especially during the second-half of the 20th century.^[21] The Scandinavian democracies undertook a similar journey during the 1970s, but have especially contended with regulation policy since the 1980s.^[22] In Canada, the government instituted a regulatory policy beginning in the second-half of the 20th century that was constructed of a complex set of regulatory agencies,^[23] as did many South American countries^[24–26] and Australia.^[27] The Japanese began to mull over the American approach to regulation in the 1960s, with an operative version

in place since the 1980s.^[28] Creation of the European Union also induced the establishment of regulatory agencies.^[29]

However, the institutional development of regulatory policy was different in each country. This is due, among other things, to the divergence in regime position and economic ideology within each nation. For instance, in the United Kingdom regulatory departments were first established in government ministries and subordinated to the full and direct power of the respective minister. Later on, regulatory authorities were established next to government ministries with the subordination to the mandated minister varying from one ministry to another and from one country to another. In still other countries, such as the United States, regulatory authorities were set up external to the government and endowed with almost complete independence from the elected echelons with respect to policy formation.

Beginning in the 1980s, a trend stressing reform began to gain currency regarding the place and significance of regulatory policy in several democratic countries, even in those that were slowly adopting democracy as a doctrine. This trend was expressed, for example, by the transformation of U.K. ministerial authorities into “independent” agencies, usually parallel to the unification of several regulatory agencies under an inclusive superagency such as the Financial Services Authority (FSA), Office of Communications (OFCOM), Office of Gas and Electricity Markets (OFGEM), all in the United Kingdom. Reform aimed at “unification” has also been initiated in Norway, Denmark, and Sweden for more than a decade, systematically and steadily although slowly, and primarily in the sphere of financial markets. Such superagencies are also to be found in Austria, Bermuda, the Cayman Islands, Estonia, Germany, Gibraltar, Hungary, Iceland, Ireland, Japan, Latvia, Moldova, Malta, Nicaragua, Singapore, and even South Korea.^[30] Changes and reassessment of procedures can also be observed in Jamaica, Bolivia, Chile, Peru, Mexico, Argentina, and Venezuela.^[25]

Regulation in the Middle East countries

From a historical and primarily cultural perspective, the Middle East is a subregion of Africa and Asia. Some cartographers include countries from southwest Asia together with Egypt, while others also include North Africa and Central Asia. According to the International Air Transport Association (IATA), the standard organization of the airline industry, the Middle East encompasses Bahrain, Egypt, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Palestinian Territory, Oman, Qatar, Saudi Arabia, Sudan, Syrian Arab Republic, United Arab Emirates, and Yemen.

However, it is delineated, the Middle East has always been considered by the western world to be a predominantly Islamic Arabic community. Yet, within the Middle East there are several cultures and particularly different ethnic groups, such as Arabs, Armenians, Assyrians, Azeris, Berbers, Chaldeans, Druze, Greeks, Jews, Kurds, Maronites, Persians, and Turks.

A frequent description refers the Middle Eastern countries as *economies in transition*,^[31] moving from heavily state-dominated and closed economies to systems based on private initiative, market forces, and greater integration with the rest of the world. Most of the countries in the region still have limited and declining economic relations with western countries, large public sectors, and low levels of private investments, however, globalization, technology, and free markets do affect structural and institutional reforms. While social regulation may be implemented as part of a socialist doctrine, economic regulation can only take place parallel to the development of free markets. In Saudi Arabia, for example, regulation of the telecommunications industry was established in 2002 in order “to ensure the provision of high-quality universal telecommunications services at affordable prices.”^[32] Since the last quarter of 2004, the Communications and Information Technology Commission (CITC) has consistently opened telecommunications markets for competition and has liberalized mobile telephone use. In addition, the structure of the regulatory authority was following western countries’ reforms: it is governed by a Board of Directors who are responsible for policy formation, budgetary approval, and administrative reviews. The Minister of Communications and Information Technology who moderates between the government macropolicy and CITC regulations chairs the board. Below the chairperson, there are seven more directors; one serves as the chairperson deputy and as the governor of the commission (CEO). The same type of development happened in Bahrain, when the Telecommunication Regulatory Authority (TRA) was established in 2002 in order to protect “the interests of subscribers and users, and [promote] effective and fair competition among established and new licensed operators.”^[33] The TRA is governed by a board of five directors who serve independently from the cabinet of ministers. In the Hashemite Kingdom of Jordan, a Telecommunication Regulatory Commission (TRC) was established in 1995 and governed by a board of five full-time commissioners, one of whom served as Chairperson and CEO. The Jordanian TRC is a quasi-independent regulator that is directly accountable to the Prime Minister but subordinate only to the general policy of the government.

The inherent dependency between social and economic affairs can be understood by considering the Egyptian reform of environment regulation. As had

happened dozens of years earlier in the United States, Canada, Australia, Europe, Latin America, and also Japan. Egypt established an Environmental Affairs Agency in 1997 in order to prevent and reduce negative externalities caused by air and water pollution, urban sprawl over natural resources, ozone effects, and other detrimental conditions. Since the 20th century has proven that environmental hazards are no longer a state issue but an issue for all humankind, all governments are expected to take part in the global and international efforts to reduce the risks. Egypt, as one of the leading countries of the Arab modern world, has adopted the regulatory doctrine of the western world by establishing an environmental protection agency.

Governments in the Middle East have also adopted finance regulation, especially through monetary police and central banks. Bahrain, for instance, established the Bahrain Monetary Agency in 2002 in addition to the Sharia Supervisory Board. While the former regulates banking and finance within capital market standards, the latter regulates banking and finance under Muslim orthodox laws and tradition.^[34] Saudi Arabia also instituted finance regulation in 2003 with the establishment of the Saudi Arabia Monetary Agency. Similarly, Kuwait established an independent central bank in 2003 in order to “secure stability of the Kuwaiti currency and its free convertibility into foreign currencies; [and] to direct credit policy in such manner as to assist the social and economic progress and the growth of national income.”^[35] The Kingdom of Jordan transferred regulatory power to the central bank in 2000, and Lebanon followed in 2004.^[36]

The state of Israel is the only non-Islamic or Christian state in the Middle East. Regulatory policy has been designed and implemented in Israel since its infancy in the early 1950s. Yet, regulation was especially developed in the 1980s as part of privatization policy, and even more so after the mid-1990s. Most markets and companies are regulated by statutory regulators, which are subordinate directly to the executive government. Other regulators serve as quasi-independent and some agencies earned total independence from government. Today, the state of Israel regulates banks through the Central Bank; the securities are regulated by the Israel Securities Authority; the commercial TV and radio broadcasting are regulated by Board of Directors; electricity is regulated by the Public Utility Authority—Electricity; and higher education and academic activities are regulated by the Israeli Council of Higher Education. Parallel to sectoral regulation, the Israel Antitrust Authority is responsible for the competitive regulation of all markets. Administrative departments that are structured as part of the ministries and subordinate to the ministers mainly control social regulation.^[37]

CONCLUSION

Regulation has become the most common and well-known policy tool and characteristic of modern states during the last decade. Usually its development was, and still is, a part of markets liberalization and globalized economies. Since its very beginning over a 150 years ago and until recent years, regulation seemed to be a unique reflection of the western world and modern democracies. Yet, the last 15 years prove that developing countries as well can adopt the same doctrines and institutional measures of regulation. One fascinating example is the development of regulation policy in the Middle Eastern countries, particularly considering the fact that the majority are Muslim states by law. Within these countries that are in transition, it will be exceptionally interesting to view how economic and social regulation will be designed and implemented within their politics during the next decade. As part of globalization and international trade relations all over the world, the age of regulatory governance will not skip over the Middle Eastern modernization and will likely contribute to its growth.

REFERENCES

1. Moran, M.; Wood, B. *State, Regulation and the Medical Profession*; Open University Press: Philadelphia, 1993.
2. Reagan, M.D. *Regulation: The Politics of Policy*; Little Brown and Company: Boston, 1987.
3. Baldwin, R.; Cave, M. *Understanding Regulation*; Oxford University Press: Oxford, 1999.
4. Weimer, D.L.; Vining, A.R. *Policy Analysis: Concepts and Practice*, 4th Ed.; Prentice-Hall: New Jersey, 2004.
5. Harlow, C.; Rawlings, R. *Law and Administration*, 2nd Ed.; Cambridge University Press: Cambridge, 1997.
6. See also in Ogus, A. *Regulation: Legal Form and Economic Theory*; Oxford University Press: Oxford, 1994.
7. Doern, G.B.; Wilkis, S. Introduction. In *Changing Regulatory Institutions in Britain and North America*; Doern, G.B., Wilkis, S., Eds.; Toronto University Press: Toronto and London, 1998; 3–25.
8. See also in Jordana, J.; Levi-Faur, D. The politics of regulation in the age of governance. In *The Politics of Regulation: Institutions and Regulatory Reforms for the Age of Governance*; Jordana, J., Levi-Faur, D., Eds.; Edward Elgar: Cheltenham, UK and Northampton, MA, 2004; 1–28.
9. Friedman, M. *Capitalism and Freedom*, 40th ANV Ed.; University of Chicago Press: Chicago, 2002.
10. Swann, D. The regulatory scene: an overview. In *The Age of Regulatory Reform*; Button, K., Swann, D., Eds.; Clarendon Press: Oxford, 1989; 1–19.
11. Breyer, S. *Regulation and Its Reform*, 6th Ed.; Harvard University Press: Cambridge, 1994.
12. Harris, R.A.; Milkis, S.M. *The Politics of Regulatory Change: a Tale of Two Agencies*, 2nd Ed.; Oxford University Press: New York and Oxford, 1996.

13. Williams, B.A.; Mathney, A.R. *Democracy, Dialogue, and Environmental Disputes: The Contested Languages of Social Regulation*; Yale University Press: New Haven and London, 1998; See also in Whitley, R.; Kristensen, P.H. *Governance at Work: The Social Regulation of Economic Relations*; Oxford University Press: Oxford, 1997.
14. Hahn, R.W. *Reviving Regulatory Reform: A Global Perspective*; American Enterprise Institute: Washington, DC, 2000.
15. Bernstein, H.M. *Regulating Business by Independent Commission*; Greenwood Publishing Group: New York, 1977. See also Francis, J. *The Politics of Regulation: A Comparative Perspective*; Blackwell Publishers: Oxford, 1993; Mitnick, B. *The Political Economy of Regulation*; Columbia University Press: New York, 1980; Wilks, S. Corporate strategy and state support in the European Motor Industry. In *Capitalism, Culture and Regulation*; Hancher, L., Moran, M. (Eds.); Oxford University Press: Oxford, 1989.
16. Black, J. New institutionalism and naturalism in socio-legal analysis: institutional approaches to regulatory decision-making. *Law Policy* **1997**, 53, 52–56; See also in Levy, B.; Spiller, P.T. *Regulations, Institutions and Commitment*; Cambridge University Press: Cambridge, 1996.
17. Heinz, J.P.; Laumann, E.O.; Nelson, R.L.; Salisbury, R.H. *The Hollow Core: Private Interest in National Policy Making*; Harvard University Press: MA, 1997; See also in Mucciaroni, G. *Reversals of Fortune: Public Policy and Private Interest*; The Brookings Institution: Washington, DC, 1994.
18. Eckert, R.D. The life cycle of regulatory commissioners. *J. Law Econ.* **1981**, 24, 113–120; See also in Olson, C.V.; Trapani, J.M. Who has benefited from regulation of the Airline Industry? *J. Law Econ.* **1981**, 24, 75–93; Paltzman, S. Towards a more general theory of regulation. *J. Law Econ.* **1976**, 19, 211–228.
19. Hall, P.A. Policy paradigms, social learning and the state: the case of economic policy-making in Britain. *Comparative Politics* **1993**, 25, 275–296.
20. Jordana, J.; Levi-Faur, D. Toward Latin American regulatory state?: The diffusion of autonomous regulatory agencies across countries and sectors. In *International Public Policy and Management: Policy Learning Beyond Regional, Cultural, and Political Boundaries*; Levi-Faur, D., Vigoda-Gadot, E., Eds.; Marcel Dekker: New York, 2004; 155–187.
21. Coriat, B. France: the end of Fordism... and no successor inside. In *Regulation Theory—the State of the Art*; Boyer, R., Saillard, Y., Eds.; Routledge: London and New York, 2002; 247–253; Muller, M.M. ‘Reconstructing the new regulatory state in Germany: Telecommunications, broadcasting and banking. *German Politics* **2001**, 10 (3), 37–64.
22. Jacobzone, S.; Frison-Roche, M.A. *Regulatory Reform in Norway: Modernising Regulators and Supervisory Agencies*; OECD: Paris, 2003; Levi-Faur, D. Governing the Dutch Telecommunications Reform: regulation, deregulation and re-regulation within the context of European Policy Regime. *J. Eur. Public Policy* **1999**, 6 (1), 102–122; Taylor, M.; Fleming, A. *Integrated Financial Supervision: Lessons of Northern European Experience*; The World Bank: Washington, DC, 1999.
23. Sghults, R.; Doern, G.B. No longer “Governments in Miniature.” Canadian sectoral regulatory institutions. In *Changing Regulatory Institutions in Britain and North America*; Doern, G.B., Wilks, S., Eds.; University of Toronto Press: Toronto, Buffalo and London, 1998; 108–130.
24. Aboites, J.; Miotti, L.; Quenan, C. Regulationist approaches and accumulation in Latin America. In *Regulation Theory—The State of the Art*; Boyer, R., Saillard, Y., Eds.; Routledge: London and New York, 2002; 280–288.
25. See also in Guasch, J.L.; Spiller, P. *Managing the Regulatory Process: Design, Concepts, Issues, and the Latin America and Caribbean Story*; The World Bank: Washington, DC, 1999.
26. Levi-Faur, D. The politics of liberalization: privatization and regulation-for-competition in Europe’s and Latin America’s Telecoms and Electricity industries. *Eur. J. Political Res.* **2003**, 42 (5) 705–740.
27. Brennan, G.; Pincus, J. Australia’s economic institution. In *Australia Reshaped: 200 Years of Institutional Transformation*; Brennan, G., Castles, F.G., Eds.; Cambridge University Press: Cambridge, 2002; 53–85; See also in Castles, F.G. Australia’s institutions and Australia’s welfare. In *Australia Reshaped: 200 Years of Institutional Transformation*; Brennan, G., Castles, F.G., Eds.; Cambridge University Press: Cambridge, 2002; 25–52.
28. Amyx, J.A. Informality and institutional inertia: the case of Japanese financial regulation. *Japanese J. Political Sci.* **2001**, 2 (1), 47–66; See also in Maswood, J. Regulatory reforms in Japan: issues and prospects. In *Japan—Change and Continuity*; Maswood, J., Graham, J., Miyajima, H., Eds.; Routledge: London and NY, 2002; 30–53.
29. Del-Valle, I.; Astorkiza, I.; Astorkiza, K. Is the current regulation of the VIII Division European Anchovy optimal? *Environ. Resource Econ.* **2001**, 19 (1), 53–72; Feinberg, R.M.; Harper, R.K. Regime effects of EU market integration policies on the UK financial sector. *Rev. Ind. Org.* **1999**, 15 (4), 357–365.
30. Martinez, J.D.L.; Rose, T.A. *International Survey of Integrated Financial Sector Supervision*; The World Bank: Washington, DC, 2003.
31. Page, J. From boom to bust and back? The crisis of growth in the Middle East and North Africa. In *Prospects for Middle Eastern and North African Economies*; Shafik, N., Ed.; MacMillan: London, 1998; 62–79; Also: World Bank. *Claiming the Future: Choosing Prosperity in the Middle East and North Africa*; The World Bank: Washington, DC, 1995.
32. See Council of Ministers Decision No. 74, March 5th, 2002; CITC website: www.citc.gov.sa/CITC/EN.
33. Legislative decree No. 48 of 2002 (TRA website: www.tra.org.bh/en).
34. Islamic Banking regulation... 2002.
35. CBK Law 32/1968, ch. 2, section 2, article 15 (2–3) (in CBK website: www.cbk.gov.kw/www/law).
36. Law no. 575 (see on the central Bank website: www.bdl.gov.lb/circ/en/law575e.html).
37. Arbel-Ganz, O. Regulation policy in Israel: regulatory agencies and policy networks. In *The State of Israel: New Thoughts*; Richman, U., Nachmais, D., Eds.; Mifalot, The IDC Publishers: Herzlia, 2006; 149–179 [Hebrew].

Middle East: Corruption

Doron Navot

Political Science Department, Hebrew University, Jerusalem, Israel

INTRODUCTION

Corruption is a complicated and multifaceted phenomenon and it is a contested concept. The meaning of corruption is theory laden and the various uses of the term are influenced by political interests and developments. By its very nature, corruption is concealed, and part of what makes a behavior corrupt is the intention of the public officer to promote a private gain. As a result, it is hard to measure corruption directly or indirectly. When dealing with corruption in the Middle East, those challenges make the mission of exploring the phenomenon scientifically, almost impractical. Raw data on the subject hardly exists, and empirical researches of corruption in specific countries in the region are not to be found. Indeed, there are indications for a serious predicament of corruption, but very few substantive materials that support any unequivocal claim.

The effects of corruption, or so-called corruption, should also be considered when dealing with corruption in the Middle East. There are indications that in the recent years public opinion within the Middle East countries is bothered by the lack of democratic features and the existence of corruption. In addition, transnational organizations like the World Bank and OECD work to curb corruption in the region.

This article continues as follows. The first part presents the concept of corruption. The second part provides empirical findings and estimations of the range of corruption in the Middle East. It also briefly deals with methodological problems and presents possible theoretical explanations for the existence of political corruption in the area. The potential role that corruption and anticorruption activities will play in the future of the current regimes will be described in the last part.

THE CONCEPT OF CORRUPTION

Scholars agree that a clear-cut definition of corruption is not to be found. While it is acceptable to define the phenomenon as the misuse of public office for private gain, the precise meaning of the essentials that compose the definition, and the applicability to real cases and events, are far from being straightforward.^[1] In order to use the concept for public policy, we should

bear in mind that while corruption should be regarded as misbehavior or abusive activity, there are different perspectives of what is proper and what is not.^[2]

There are at least two types of corruption in politics, or two types of corrupting politics: the inappropriate struggle for power and exercising power in order to get illegitimate personal gains. Another distinction is between corruption of an appointed civil servant and corruption of an elected one. Similarly we can distinguish between political corruption per se, which is related to political processes, and bureaucratic corruption, which is related to the public.^a Finally, we should consider the principal divergent between political corruption—which is often hidden—and political scandals or “corruption affairs”—which by definition occur in public. Behind accusations of corruption, corruption does not necessarily exist. On the other hand, the fact that we do not hear about corruption does not necessarily mean that it is not present.

CORRUPTION IN THE MIDDLE EAST

What do we know about corruption in the Middle East? We have three different indicators, which point toward a problem of corruption: 1) Corruption Perceptions Index (CPI) that were made for Transparency International (TI) and published by TI; 2) “impressionistic evidence;” and 3) structural analysis. Each one of these measures does not observe the phenomenon directly, and separately each one is not a credible measure of corruption. Yet, the fact that they

^aArnold Heidenheimer suggested distinguishing between three types of definitions: public service-centered definitions; economic-centered definitions; and public opinion definitions. See: Heidenheimer, A.J., Ed.; *Political Corruption—Readings in Comparative Analysis*; Holt, Rinehart and Winston, Inc., 1970. After 2 years, James Scott offered a classical typology of definitions. Scott distinguished between three types: legal-centered definitions; public interest-centered definitions; and public opinion-centered definitions. Scott, J. *Comparative Public Corruption*; Prentice-Hall Inc.: New-Jersey and Englewood Cliffs, 1972. However, in the 1990s, scholars like Michael Johnston and Mark Philp showed that all the definitions are based on the same vague criterion, which are the sound condition of politics. For elaboration, see: Johnston, M. The search for definition: the vitality of politics and the issue of corruption. *Int. Social Sci. J.* **1996**, *149*, 321–335. Philp, M. Defining political corruption. *Political Studies* **1997**, *45* (3), 436–462.

all point to the same direction, provides us with reasons to believe that corruption is pervasive.

The opinion surveys are the most commonly used tools to assess corruption.^[3]

The data in Table 1 shows that Israel is perceived as the least corrupt country, with Iraq in the worst position. This is not surprising, if we take into account the fact that Israel is the only state with free and fair periodic elections and an independent judiciary system, and that it is identified as the single democracy in the region. Iraq has been in a chaotic situation for several years, and to perceive this country as corrupt under these circumstances is natural reaction. It is also clear that the range of the ranking is wide, and that the perceptions of corruption among the states varied. Jordan is traditionally in the middle low ranking, with Egypt just behind. Countries closer to western influence are perceived as least corrupt.

But the index has several limitations: 1) there is a difference between perceived corruption and experience of corruption^[4]; 2) the CPI is interested in the business environment, and less with issues like political appointments, nepotism, and the relative strength of the state vis-à-vis the business community^[3]; 3) not all the countries are included in the CPI, and there are differences in the number of the countries participating in the survey over the years; and 4) when it comes to surveys in the region, the samples on which the surveys are based upon are relatively quite small.

There are plenty of impressionist evidences that the regimes in the area are corrupt, some are a bit hasty. A good example is the identification of Arafat's regime with corruption. The founding father of the Palestinian authority kept millions of dollars that he received for the Palestinian people in his private accounts, and frequently used the money to buy political support (bribing opponents and keeping allies "loyal"). However, Arafat did not use the money for (non-political) personal gain.^[5] His behavior fits well to

what was termed by Thompson as mediated corruption^[6] or institutional corruption,^[7] but not with conventional corruption. In other words, even if Arafat's behavior is a sort of corruption, it should be distinguished from briberies, embezzlement, theft, bid-rigging, kickbacks, fraud in contract performance, and product substitutions. Another feature that creates the impression that regimes in the area are corrupt is the fact that most of the elections taking place in countries like Iraq or Egypt end up with the victory of the current ruler, who usually wins 99.8% of the votes. Such unreasonable results are due to undemocratic elections and their mere existence presents those regimes as corrupt. In conclusion, impressionistic evidence supports the data provided by the CPI. However, this source of data is damaged by biased judgments and suffers from inaccuracy.

Probably the best type of data about corruption in the Middle East is the existence of factors that contribute to corruption, or associated with it.^[8] Arab countries are characterized by a centralized executive, inadequate institutional checks and balances, and lack of institutional mechanisms of political participation. There are no basic mechanisms for accountability such as free and fair elections. Public figures are not answerable to the people, there is little public transparency, and a lack of available information on what government does. In addition, most of the countries in the region have weak civil societies.^[9,10] Another indirect indication for corruption is poor governmental performance of most of the countries in the region.^[11] Poor governance might be a cause for corruption or the result of it (or both). One way or another, it is another indication for the pervasiveness of corruption.

The effects of corruption on Middle Eastern countries are very negative: 1) Corrupt politicians and corrupt states that allegedly exist in the Middle East cause injustice and suffer among the people^[12]; 2) political corruption is destructive to trust and institutional

Table 1 Corruption Perception Index—countries in the Middle East: 1996–2005

Year	Egypt	Iran	Iraq	Israel	Jordan	Kuwait	Lebanon	Palestine	Saudi Arabia	Syria
1996 (54)	41	—	—	14	30	—	—	—	—	—
1997 (52)	—	—	—	15	—	—	—	—	—	—
1998 (85)	66	—	—	19	38	—	—	—	—	—
1999 (99)	63	—	—	20	41	—	—	—	—	—
2000 (89)	63	—	—	22	39	—	—	—	—	—
2001 (91)	54	—	—	16	37	—	—	—	—	—
2002 (102)	62	—	—	18	40	—	—	—	—	—
2003 (133)	70	78	113	21	43	35	78	78	46	66
2004 (145)	77	87	129	26	37	44	97	108	71	71
2005 (158)	70	88	137	28	37	45	83	107	70	70

trust,^[13] which are both necessary conditions for any political or collective action; and 3) it impoverishes economies, undermining the rule of law and facilitates other threats such as terrorism.^[14] In conclusion, while corruption might have positive effect in industrialized countries, in the short run, when it comes to the pervasiveness it is allegedly in the Middle East, corruption is disastrous.^[15]

The simplest and straightforward explanation for corruption in most of the countries in the area is the fact that their regimes are not liberal or citizen oriented; they are authoritarian with exclusive tendencies and there is a shortage of any democratic experiences or traditions. If we recall Acton's famous dictum "power tends to corrupt and absolute power corrupts absolutely," or consider simple and logical economic models that analyze the causality between discretionary and monopoly power without accountability and levels of corruption,^[16] we realize that corruption in the area is only to be expected. In other words, there are many reasons for corruption to exist—and only few not to. In conclusion, there are several indications that corruption is pervasive in the region, and none that tells the opposite. Israel is the exception, probably because of its different political system. The primary reason for the predicament of corruption can be termed as nondemocratic structures.

CONCLUSION

Since the fall of the Soviet Union in 1989, globalization and democratization movements have increased all over the world. The concern for corruption in the Middle East has grown accordingly. The phenomenon is regarded as a serious problem and a major impediment to the development of countries like Jordan, Egypt, and Lebanon. The failure to address the challenges of curbing corruption puts governments under the dangers of reduced levels of foreign direct investment and disadvantages in international/global trade. The term "corruption" is used in very intensively lately and almost every government in the region has to deal with accusations of corruption, as well as demands for liberalization and reforms.

Because of the large amounts of corruption in the various levels of the political systems and public services of the countries in the region, it is safe to surmise that more and more countries will have to deal with demands for curbing corruption, from within the state and/or from outside of it. Probably, economic interests will shape the structure of those political reforms and play an important role in their design.

Globalization has the potential for improving the state of corruption in the region. It provides new economic opportunities, and in order to realize them it is

required to curb corruption. In addition, there are actors like the World Bank, who recently played an important role in the fight against corruption in the Middle East. However, in the meantime, it is evident that economical incentives and international pressures are inadequate for changes to occur. Paradoxically, what is needed the most for curbing corruption in government is a healthy civil society, vital political system, and a good government—three elements that are absent in most of the countries in the region.

Possibly, the future of corruption and anticorruption in the area lies in the hands of the international community and global actors. Another possibility is the rise of integrity ethos from within one or more of the countries, but past experience does not leave us too optimistic in this point.

The political responses to pressures for structural and institutional reforms might lead to democratization in the region and/or in specific countries, like Egypt. But the accusations of corruption, if indeed they will be stronger and more powerful in certain countries as can be expected, might end up differently—with the rise of new types of non-democratic regimes. Among the many important and interesting questions we can ask about corruption in the Middle East, is whether the efforts for democratization, as an institutional effort to curb corruption (and other political evils like poverty and terror), will lead to democracy or not, and whether democracy, if there will be democracies in the region, will decrease corruption and lead to better regimes and politics. Nevertheless, because of the existence of large amounts of corruption in politics all over the area, the growing awareness to the subject, and because of the politics of corruption all over the world, one can speculate that in the near future the issue of corruption in the Middle East will function as catalyst for constitutional and political changes in the whole region.

ACKNOWLEDGMENTS

The author is grateful to Nick Duncan for useful advices and sharing time and knowledge, and to Fredrik Galtung, the co-director of TIRI.

REFERENCES

1. Philp, M. Corruption definition and measurement. In *Measuring Corruption*; Sampford, C., Shacklock, A., Connors, C., Galtung, F., Eds.; Ashgate: Hampshire, England and Burlington, 2006; 45–56.
2. Euben, P. Corruption. In *Political Innovation and Conceptual Change*; Ball, T., Farr, J., Hanson, R.L.,

- Eds.; Cambridge University Press: Cambridge, 1989; 220–247.
3. Galtung, F. Measuring the immeasurable: boundaries and functions (Macro) corruption indices. In *Measuring Corruption*; Sampford, C., Shacklock, A., Connors, C., Galtung, F., Eds.; Ashgate: Hampshire, England and Burlington, U.S.A., 2006; 103–120.
 4. Seligson, M.A. The measurement and impact of corruption victimization: survey evidence from Latin America. *World Dev.* **2006**, *34* (2), 381–404.
 5. Jamal, A. Institutionalizing the political system and the politics of control. In *Media Politics and Democracy in Palestine*; Sussex Academic Press: Brighton Portland, 2005; 22–26.
 6. Thompson, D.F. Mediated corruption: the case of the Keating Five. *Am. Polit. Sci. Rev.* **1993**, *87* (2), 369–381.
 7. Thompson, D.F. *Ethics in Congress*; The Brookings Institution: Washington, DC, 1995.
 8. Langseth, P. Measuring corruption. In *Measuring Corruption*; Sampford, C., Shacklock, A., Connors, C., Galtung, F., Eds.; Ashgate: Hampshire, England and Burlington, 2006; 15 pp.
 9. Essoussi, A. *Public Sector Transparency and Accountability in Selected Arab Countries: Policies and Practices*; United Nations Development Programme Regional Bureau for Arab States: New York, 2004.
 10. World Bank. *Better Governance for Development in the Middle East and North Africa: Enhancing Inclusiveness and Accountability*, MENA Development Report (no. 27146), World Bank: Washington DC, 2003.
 11. World Bank. *Economic Development and Prospects: Oil Blooms and Revenue Management*. Middle East and North Africa Region (paper no. 32994). World Bank: Washington DC, 2005.
 12. Shklar, J. *The Faces of Injustice*; Yale University Press: New-Haven and London, 1991.
 13. Chang, E.C.C.; Chu, Y. Corruption and trust: exceptionalism in Asian democracies? *J. Politics* **2006**, *68* (2), 259–271.
 14. Webb, P. The United Nations convention against corruption. *J. Int. Econ. Law* **2005**, *8* (1), 191–229.
 15. For the theoretical argument, see: Friedrich, C. *The Pathology of Politics*; Harper and Row Publishers, 1972; Elster, J. *The Cement of Society*; Cambridge: Cambridge University Press, 1989.
 16. Klitgaard, R. *Controlling Corruption*; University of California Press: Los Angeles, 1988.

Milgram Experiments

Holona L. Ochs
Andrew B. Whitford

Department of Political Science, University of Kansas, Lawrence, Kansas, U.S.A.

INTRODUCTION

Stanley Milgram's classic experiments, conducted between 1960 and 1963, sought to demonstrate the power of social influence by examining how subjects related to authority. The initial studies were later followed by several variants intended to test the factors mediating the conformity identified in those first experiments. The results of this research agenda are quite robust, having been replicated in numerous cultural contexts and appear to hold up well over time. In short, subjects of many social groups, and even some of arbitrary definition (i.e., men and women, and even Americans, Germans, Europeans, and Latinos), exhibit high levels of obedience to authority, and the overall levels of conformity remain relatively stable over time. Milgram shows that authoritarianism and conformity are not confined to particular social groups or "types" of people or even to types of situations. In fact, six out of every ten people will kill you if told to do so by someone in authority under a variety of circumstances.^[1]

Moreover, Milgram's experiments are important because they have strong implications for ethical behavior on the part of experimenters who study authority relations in controlled settings. Many recent innovations in the protection of human subjects in experimental and other research settings are broadly due to the innovations introduced—and the resulting findings—by Milgram. For example, the American Psychological Association's *Ethical Principles in the Conduct of Research with Human Participants* (1982) make direct reference to these experiments and their findings on social authority in controlled settings. Most importantly, Milgram's experiments have great importance for the study of public policy generally, and public administration specifically. The current use of human subjects committees is particularly relevant, as well as the more pressing matters of social regulation and control in hierarchical settings.

HISTORICAL CONTEXT OF THE MILGRAM EXPERIMENTS

In the early 1960s a considerable amount of research was being conducted attempting to explain the

atrocities of World War II. Generally, researchers sought an explanation for the human cruelty exhibited during this period, and there was a tremendous desire to attribute the violence to a readily identifiable characteristic that would allow people to identify that thing that must be something fundamentally different about someone who would participate in such depravity—or at least that the degeneracy of the circumstances produced such behavior in some individuals. Of course, efforts to assign responsibility for the atrocities exacted by the Nazi regime had resulted in many participants being charged during the Nuremberg Trials. The primary defense for the accused was that they were merely following orders. The vast majority of the research examined the authoritarian traits of the German people,^[2–5] that is—with the exception of Milgram's studies.

The central importance of Milgram's studies^[1,6–8] is that they attempted to pit moral values against the demands of authority in an experimental test of human behavior under social pressures to conform.^[1] Milgram placed his experiments in the discursive field that linked his research with momentous examples of human barbarism, such as the treatment of the Jews by the Nazis in Germany and the My Lai massacre in Vietnam. He reasoned that most people would exact harsh consequences on others under certain conditions, particularly when given orders by someone in authority. Most importantly, he shows that people have an incredible tendency to conform and that in fact, most people would participate in violence under similar circumstances.

The Milgram experiments profoundly influenced the discourse on human nature, group behavior, ethics, public policy, and public administration. Specifically, the findings challenge some widely held beliefs about the capacity for human cruelty. The dispositional view that there must be something in the German character that makes them particularly cruel did not hold up when tested. Instead, the Milgram experiments undermine notions that human behavior is genetically determined or culturally determined. The results illustrate that human behavior is constrained by individual values to the degree that social influence and the situational characteristics are imposed.

METHODOLOGY

The initial experiment conducted by Milgram on obedience to authority began in a lecture theater where a group made up of psychiatrists, university students, and middle-class adults of various occupations and ages gathered for a lecture on obedience to authority.^[8] Here we describe the situation Milgram asked his audience to imagine at the end of the lecture:^[9]

In response to a newspaper add offering \$4.50 for one hour's work, you participate in a Yale University psychology experiment investigating memory and learning. You are introduced to a stern looking experimenter in a white coat and a mild-mannered individual introduced as your co-participant. The experimenter explains that the study examines the role of punishment in learning, and one of you will be the teacher and one will be the learner in this experiment. You are randomly assigned to the role of the teacher, and the three of you then proceed to an adjacent room, where the other participant, the "learner," is strapped into a chair. The experimenter explains that this is to prevent excessive movement during the experiment, but it is also obvious to you that the learner is unable to remove himself from the chair if he so chose. Then, conductive gel is applied to an electrode is subsequently attached to the learners arm, and the experimenter explains that this is to prevent burning and blisters. Both you and the learner are shown the electric shock generator that will deliver electric shocks that will serve as punishment for incorrect responses. The generator has 30 switches labeled with a voltage ranging from 15 to 450 volts. Each switch also has a rating, ranging from "slight shock" to "danger: severe shock," and the final two switches are labeled "XXX." The other participant assigned to the learner role asks if "the shocks will hurt" to which the experimenter replies, "although the shocks will be painful, they cause no permanent tissue damage."

Your role is to teach the learner a simple association task, but you also are told by the experimenter that you are required to punish the learner for incorrect responses. You are told that for every incorrect response you must increase the voltage by one more switch (15 volts). You are given a 15-volt shock to check that the generator is functioning adequately and to demonstrate the tingling sensation of 15 volts. The learner makes frequent errors during the experiment, each resulting in a higher voltage shock than the previous one. At 75 volts, 90 volts, and 105 volts, you hear the learner "grunt" through the wall. At 120 volts, the learner says the shocks are getting painful. At 150 volts, he screams, "get me out of here! I refuse to go on!" His protests continue as the voltage increases.

At any point that you question whether you should continue, the experimenter tells you to keep going, saying, "you can't stop now", "he is getting paid to do this experiment" or "the experiment depends on

your continuing compliance." He may even say "you have no choice." At 300 volts, the learner pounds on the wall and demands to be let out. After 330 volts, there is no longer any response from the learner, at which point the experimenter tells you that the learner's failure to respond should be interpreted as an incorrect response and to continue increasing the shock level. The experiment concludes when the highest shock level is reached or the learning task is completed.

Following the lecture in which the audience was asked to image the scenario described, each member of the audience was asked to anonymously record how he or she would have responded under the circumstances described. Members of the audience reported that they would have disobeyed the experimenter early in the experiment because they "didn't want to hurt anyone," and no one in the audience reported that he or she would continue beyond 300 V, the level indicating "danger: severe shock."^[9]

What we now know is that when Milgram conducted the experiment described to the audience in the preliminary study, the "learner" was actually a confederate.^[8] He found that more than 60% of the participants inflicted the maximum level of electric shock, and none of the participants disobeyed before 300 V.^[1] Despite the prediction by the psychiatrists in the preliminary study that only "pathological sadists" would administer electric shocks to 450 V, the experiment revealed that ordinary people will inflict pain on another person under the right set of circumstances, particularly when told to do so by a legitimate authority.^[8] The study also exemplifies the tendency to overemphasize dispositional factors and underestimate situational factors when making attributions about actions, referred to as a *fundamental attribution error*.^[10] For example, people tend to assume that if someone has not done much today, it is because he is lazy; rather than assuming that he is tired or lacks the right resources.

Milgram conducted several variants of the experiment to test the factors that might mediate conformity, including surveillance by the experimenter and proximity of the learner to the teacher. Several characteristics of the study participants, experimenter characteristics, attributes of the confederate learner, and features of the experimental setting were manipulated in subsequent studies. In one variation, the learner asks for the shock, and not a single participant exceeds the constraints of the test manager.^[1] It is evident that the presence and the demands of legitimate authority define the boundaries of behavior for the participants to a considerable extent.

It should be noted that a number of complaints about the Milgram experiments—beyond ethical considerations—are possible and have been voiced. For example, a central component in all of the Milgram experiments is the incremental shock procedure.

By using a graduated series of shocks it may be possible that subjects faced an innocuous beginning that elicited compliance before any of the frightening implications of the procedure were clear to the subject. That means that subjects were not presented with any sort of “qualitative breakpoint” at which they might shift from obedience to disobedience. This procedure is held by some as a fundamental component that should have been varied within the experimental context.^[11] Additionally, the studies took place during the epoch of behavioralism in the field of psychology. Consequently, the criticisms of a strictly behavioral approach are applicable to evaluating the Milgram experiments. In particular, Milgram’s focus on the appearance of authority and the ambiguity of notions of authority and obedience make drawing conclusions about human nature a substantial leap.^[12] However, subjects’ protections put in place because of the original protocol now make it difficult to replicate or extend it in any way that would meaningfully test these complaints.

IMPLICATIONS FOR PUBLIC POLICY AND PUBLIC ADMINISTRATION

Efficient and legitimate institutions can be used for constructive or destructive purposes.^[13]

During Adolf Eichmann’s testimony at the Nuremberg Trials, he claimed that he had nothing personal against the Jews or anyone else and stated that he found the concentration camps repugnant, but at the same time he described the meticulous organization of the deportation and extermination of several hundred thousand people.^[2] He contended that he had an obligation to perform the administrative duties as assigned by the Third Reich because had he refused to be obedient and responsive to the political authorities, then every soldier from any army would have the right to disobey any order considered personally objectionable. In fact, the Holocaust would not have been possible without the obedience and conformity of the public service, implementing the policies of the Nazi regime efficiently through routine practices.^[13] In essence, latent tendencies toward dehumanization exist in the routinization of any process.^[13]

However, some degree of conformity is necessary for the adequate functioning of any society. It is difficult to conceive of how decisions could be made and enforced—and actions implemented—without the recognition of some sort of authority. There is a reciprocal nature to the use and acknowledgment of authority. Public trust is essential to acquiring the compliance necessary to accomplishing civic purposes, and public servants are obligated to champion the public interest with empathy, respect, and consideration for future generations to obtain that trust.^[14]

Additionally, hierarchy is an implicit component of any social regulatory regime because people vary in their readiness to obey and to extend trust. Useful hierarchical relationships require competence of the authority system.^[1] For example, spectators in a concert hall do not respond to the instructions of the conductor. Authority is considered legitimate only by those who are incorporated into a system supporting that authority’s legitimacy. This point is crucial when it comes to understanding policy and administration because those who are not integrated into the system have no investment in maintaining it. In fact, the alienation of people from society’s institutions, processes, benefits, etc., undermines the legitimacy of the system, threatening the security and stability of the entire social group.

Conformity can also serve society in important ways.^[1,15] For example, hierarchically organized groups are tremendously advantageous in addressing dangers or threats to security. Second, organizations provide stability and serve to facilitate cooperation among group members. Third, contesting hierarchical organization can lead to greater levels of violence. However, resolving the questions that arise when individual values and the demands of authority conflict remains among the central philosophical problems. Determining the nature of man and the role of values in society is integral to developing adequate policies.

Studies of public policy and public administration can draw several insights from the Milgram experiments. For example, Milgram outlines several preconditions for obedience.^[1] Some basic attributes—such as the authority structure of parental authority within the family structure—directly condition conformity. The larger institutional framework within which we all operate also solicits obedience. Even while institutional settings teach us that one’s opinion is possible and important, for smooth functioning obedience is necessary. Rewards and sanctions of various types in a range of settings are structured to elicit conformity. This is essential to easing relations and facilitating transactions between individuals and groups.

In addition, an assortment of policy changes can be attributed to the Milgram experiments. The study itself raised numerous ethical considerations. First, participants volunteered for a learning experiment, a choice that is itself not a test of obedience. Although deception was essential to testing the obedience to authority, participants were not afforded critical information that may have affected their decision to participate, and they were deceived as to the nature of the experiment (i.e., the shocks were not real, and the learner was a confederate). Second, participants were filmed without their consent, with important implications for current practices about subjects’ consent. Third, the strength of the social pressure made it very unlikely

that participants could actually discontinue their participation if they so chose. Fourth, many of the participants exhibited severe and persistent psychological reactions to participating in the experiment. It is important to point out that the experiment itself exemplifies the principles under investigation. A study inquiring on the cruelty of humanity was continued through numerous trials despite the knowledge that there are detrimental effects on participants.

Eventually, the experiments were discontinued, and standards were developed for research that involved human subjects. The resulting policy requires research proposals to be reviewed by a *human subjects committee*. Only research meeting the standards for the treatment of human subjects is approved by the committee and capable of being carried out. The authority of human subjects committees and the standards of research practice hold considerable legitimate authority, and there are very high degrees of conformity to those standards in the United States. The federal policy for the protection of human subjects is implemented by the Department of Health and Human Services and is integrated into the procedural process of every public and private organization conducting research involving human subjects in the United States. Further, information about the federal human subjects research guidelines can be found at the United States Office for Human Research Protections (<http://ohrp.osophs.dhhs.gov/index.htm>).

CONCLUSION

Because the results of Milgram's research agenda are quite robust, having been replicated in numerous cultural contexts and over time (see Ref.^[16]), we now have a fundamentally different understanding about how authoritarianism and conformity are not confined to particular social groups or "types" of people or even to types of situations. We have also discussed the strong implications for ethical behavior on the part of experimenters who study authority relations in controlled settings, and recent innovations in the protection of human subjects in experimental and other research settings are broadly due to the innovations introduced—and the resulting findings—by Milgram. Readers seeking more information about the Milgram experiments should consult the works cited here, and general studies such as Ref.^[17] but a most instructive

introduction is located at the Milgram Reenactment (<http://www.milgramreenactment.org/pages/index.xml>), an online resource about the experiments and their lasting contribution to our understanding of authority, in all its forms.

REFERENCES

1. Milgram, S. Liberating effects of group pressure. *J. Pers. Soc. Psychol.* **1965a**, *1*, 127–134.
2. Adorno, T.W. *The Authoritarian Personality*; Harper: New York, 1950.
3. Christie, R., Jahoda, M., Eds.; *Studies in the Scope and Method of 'The Authoritarian Personality'*; Free Press: Glencoe, IL, 1954.
4. Scharrner, B. *Father Land; A Study of Authoritarianism in the German Family*; Columbia University Press: New York, 1948.
5. Shaver, J.P.; Hofmann, H.P.; Richards, H.E. The authoritarianism of American and German teacher education students. *J. Soc. Psychol.* **1971**, *84* (2), 303–304.
6. Milgram, S. Behavioral study of obedience. *J. Abnorm. Soc. Psychol.* **1963**, *67*, 371–378.
7. Milgram, S. Some conditions of obedience and disobedience to authority. *Human Relat.* **1965b**, *18*, 57–75.
8. Milgram, S. *Obedience to Authority: An Experimental View*; Harper-Collins: New York, 1974.
9. Adapted and Abridged from the Stanley Milgram Papers. Manuscripts and Archives, Yale University Library.
10. Kelley, H.H. Attribution in social psychology. *Neb. Symp. Motiv.* **1967**, *15*, 192–238.
11. Gilbert, S.J. Another look at the Milgram obedience studies: the role of the graduated series of shocks. *Personal. Soc. Psychol. Bull.* **1981**, *7* (4), 690–695.
12. Helm, C.; Morelli, M. Stanley Milgram and the obedience experiment: authority, legitimacy, and human action. *Polit. Theory* **1979**, *7* (3), 321–345.
13. Adams, G.B.; Balfour, D.L. *Unmasking Administrative Evil*; Sage Publications: London, 1998; xxix pp.
14. Lewis, C.W. *The Ethics Challenge in Public Service: A Problem-Solving Guide*; Jossey-Bass: San Francisco, 1991.
15. Zimbardo, P.G.; Lieppe, M.R. *The Psychology of Attitude Change And Social Influence*; Temple University Press: Philadelphia, 1991.
16. Smith, P.B.; Bond, M.H. *Social Psychology Across Cultures: Analysis and Perspectives*; Allyn and Bacon: Boston, MA, 1994.
17. Miller, A.G. *The Obedience Experiments: A Case Study of Controversy in Social Science*; Praeger: New York, 1986.

Mitigating HIV/AIDS' Impact on Teachers and Administrators in Sub-Saharan Africa

Terry F. Buss

*Center for International Affairs, National Academy of Public Administration,
Washington, District of Columbia, U.S.A.*

Mumukshu Patel

National Academy of Public Administration, Washington, District of Columbia, U.S.A.

INTRODUCTION

As is well known, Sub-Saharan Africa has been devastated by the HIV/AIDS pandemic. Recent estimates (as of December 2004) by the World Health Organization show 29 million people infected, with an additional 2.5 million being infected annually and cumulative AIDS-related deaths at 13.7 million. Although only 10% of the world's population lives in Sub-Saharan Africa, nearly 60% of the world's HIV-positive population resides there.^[1,2]

If the direct, immediate impacts of the epidemic were not enough, the longer-term consequences could be even worse. One emerging area of concern is the impact on the education sector—not only on teachers (including university professors) and school administrators but also on civil servants in Ministries of Education. Although figures are not widely available, preliminary statistics^a are frightening.^[3,4]

Teacher attrition from AIDS is estimated to be highest in Kenya (25,000), Nigeria (22,100), South Africa (44,900), Uganda (14,900), and Zimbabwe (16,200).^[3] Some 200,000 of Sub-Saharan Africa's 650,000 teachers are projected to die from the disease.^b At present, more than half of education positions are unfilled. Globally, it is estimated to cost \$1 billion annually to compensate for the loss and absenteeism of teachers from AIDS.^[2,5] Estimates suggest that an

infected teacher loses about 6 months of professional working time before developing full-blown AIDS, while an average of 1 year lapses between the onset of clinical AIDS and death: Teachers experience about 18 months of increasing disability before leaving the school system.^[6] What does the loss of large cadres of teachers, school administrators, and civil servants in the education sector mean to Sub-Saharan African countries? To begin with, the public sector is arguably the largest employer in Sub-Saharan African countries, and the education sector makes up the lion's share. Heavy casualties imply that this and the next generation of young people will not receive the education they need to become productive members of society. Young people will be unable to find employment either in the private or public sector, and will experience difficulty in starting their own entrepreneurial endeavors. Loss of administrators and civil servants will remove skilled workers from the bureaucracy that manages education. Remaining public sector employees will be pushed to the limit, trying to cover work previously performed by others. Productivity will decline precipitously, not only from deaths but also from absenteeism among those affected or caring for relatives. Advances in economic development and growth will be reversed, perhaps for decades in the future. Development will not be sustained (see Fig. 1).

BACKGROUND

In the past, many believed that loss of teachers, school administrators, and civil servants would not be terribly devastating, essentially self-correcting, for at least two reasons: First, they believed that civil service employment was fairly low-skilled and that the vast pool of unskilled workers in many countries could be easily trained to replace those lost to the system. Second, they argued that populations devastated by AIDS necessitated a decrease in administrators and teachers, so the system would maintain its equilibrium. Those holding

^aThe Association for the Development of Education in Africa (ADEA) is a group of Ministers of Education from African countries attempting to identify the impacts of the AIDS pandemic on education and to devise strategies for addressing it (see www.adea-net.org/workgroups/en_adhocvih.html).

^bThere are no definitive studies explaining why teachers are so vulnerable to HIV infection, or whether they have higher prevalence rates than the general public. Anecdotal evidence suggests that they, like the general public in rural areas, lack information about prevention—"safe sex." Teachers' sexual exploitation of students in exchange for "good" grades may increase transmission of disease. And tribal beliefs—e.g., having sexual relations with virgins to cure AIDS—may also spread the disease.

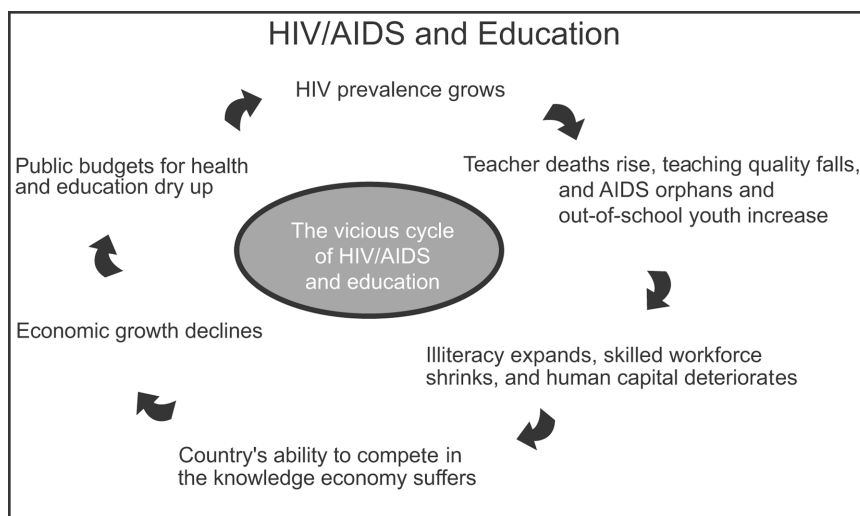


Fig. 1 HIV/AIDS impacts on education and the economy. (Adapted from World Bank, www.developmentgoals.org, 2004.)

these beliefs are clearly wrong. The pool of unskilled labor generally cannot be converted to education sector employees, and those who could be trained require more skills than can be transferred in a short time. Even if they could be trained, they would be likely absorbed into higher-paying jobs in the private sector, non-governmental organizations, or development agencies. On the second point, it is true that millions of children are dropping out of school and dying because of AIDS. In fact, by 2010, between 10% and 20% of primary school age children will be lost to the disease. It is also true that there are many more who are not infected, but have become orphaned,^c drafted into the labor force to help parents, or have become caregivers for infected relatives. These children still have educational needs, but they simply cannot access the traditional education system given their circumstances.

National governments need to devise strategies for preventing AIDS' impacts on teachers, school administrators, and civil servants where they can, and managing the consequences of the epidemic on the education sector that are already well underway. Formal strategies are now required as a prerequisite for receiving funding from the Global AIDS Fund and for loans from the World Bank. A national strategy will require:

- Stemming further attrition in the ranks of teachers, school administrators, and civil servants through prevention.
- Training a cadre of workers to replace those lost in the public sector.

- Re-engineering the workforce to cover functions affected by attrition.
- Re-engineering ministries, schools, universities, and educational bodies to make them more efficient and effective in light of a changing workforce.

Government reform and advancement has been difficult to attain and sustain in many Sub-Saharan African countries, irrespective of AIDS. Overlaying the effects of the epidemic makes educational workforce issues even more problematic for policy makers. Ironically, it might compel African governments to rethink government and education generally, perhaps with positive outcomes. Uganda, for example, realized that the disease was a problem early on, and did something about it, reducing its prevalence substantially more than any other African country.^[7,8]

Most countries initially relegated management of the epidemic to the Ministries of Health with little involvement from or concern about other ministries, including Education. With assistance funding now flowing to many countries in Sub-Saharan Africa, disease management has now been made the responsibility of National AIDS Commissions. This strategy, in hindsight, only makes sense if the epidemic is solely a health problem, requiring prevention and treatment. But the downward spiraling effects of AIDS on the education workforce itself require a broader strategy. Thus far, Sub-Saharan African policy makers and the international community are just beginning to address the problem and few countries have crafted responses to it.^[9] The Global AIDS Fund, for example, is yet to award funding to countries wishing to pursue education workforce capacity-building strategies, focusing instead on prevention and treatment. Malawi has become proactive on the issues and is exploring the possibility of providing antiretroviral (ARV) drugs to its employees to extend their work-life. The country

^cThe AIDS pandemic has created legions of orphans who have no parents, no means of livelihood, and will not likely complete school. For a review of this issue, see: Kalanidhi, S.; Diane, C. *Reaching Out to Africa's Orphans*; World Bank: Washington, DC, 2004.

is building a management information system to keep track of critical skill losses and possibilities to fill those skilled positions. The republic is also creating educational scholarships to fund college degrees for critical staff.

Policy makers in the international community are implementing programs to raise awareness of and stimulate action around the AIDS issues given above, but these are still in their infancy. The UN Educational, Scientific and Cultural Organization now operates a website clearing house addressing *HIV/AIDS Impact on Education*; ILO offers a program, *HIV/AIDS and the World of Work*, which focuses in part on educational institutions and workforce issues; the World Bank has let contracts for studies in this arena;^[6] and organizations like the Association for the Development of Education in Africa have held conferences on the topic.^[10]

Major strategic approaches employed or at least being considered by national governments to combat AIDS may be clustered as follows:

- Information/data availability essential for planning, decision making, and evaluation.
- Capacity building to ensure effective short-term responses, maintaining the smooth functioning of the system.
- Constructive partnerships for generating a comprehensive effort to combat the spread of AIDS.
- Resource-intensive long-term initiatives to stabilize workforce supply and guarantee appropriate prioritization of development issues in Sub-Saharan Africa.

The strategies highlighted above consider two factors: implementation viability in political and economic terms, and their effectiveness as policy instruments to mitigate the pandemic's impact on teachers and civil servants.

INFORMATION/DATA AVAILABILITY

It seems reasonable to assume that AIDS impacts the education sector in a significant manner, given reports of teacher and student absenteeism and attrition. But there are less reliable data on, and considerable data gaps about, how the pandemic affects the education sector. Data extrapolated from aggregate sources—like census information—to analyze AIDS impacts overlook several critical issues, including the varying rates of HIV infections across space and working groups. For example, teachers and administrators in one district may be infected at different rates than in others, or primary teachers may be infected at lower rates per student as compared to secondary school teachers.

These differences require diverse policy responses if one wishes to successfully combat the pandemic. In Zimbabwe and South Africa, policy makers have begun putting into place local data gathering and exporting programs. The programs rely on monthly administrative—e.g., attendance records—and survey forms completed by teachers and staff to provide early warning signs of problems.^{[11]d}

Finally, to obtain accurate data, there must be a friendly workplace environment, free of discrimination against victims of the AIDS pandemic. Civil society organizations have held a series of “community conversations” to help mainstream discourse on AIDS and to encourage testing and treatment.^[7] In the interim, until better data are available, policy makers might create flexible capacity-building initiatives to help mitigate AIDS impacts.

CAPACITY-BUILDING INITIATIVES

A potentially effective capacity-building initiative involves training teachers in focused, demand-driven training workshops. The Mobile Task Team approach—funded by U.S. Agency for International Development (USAID) and implemented by the Health and Economics Research Division at the University of Natal, South Africa—conforms well to this strategy and might warrant replication. In fact, the USAID is now rolling out a similar initiative in Western Africa.^[7,10] Training workshops for workers in the education sector must be executed in a timely fashion, and tailored to local demand. Emphasis in these training sessions should be on facilitated reflection and application to increase their effectiveness.^[8] Otherwise, these training workshops will degenerate into drab sessions that induce “workshop fatigue” among participants because inundation by invitations from legions of vendors will set in, stifling these initiatives. This approach views the increase in diminishing

^dLack of computer equipment and Internet access, endemic to most of Africa, need not impede efforts to track and manage the AIDS epidemic as this initiative illustrates. The District-Level Education Management and Monitoring Information System (DEMMIS) program in Harare is one possible model to assist to manage and mitigate the impact of HIV/AIDS on education. By using inputs from survey forms and attendance records at educational institutions and the Ministry of Education, it provides early warning indicators and a decision support system management, enabling policy makers to manage and mitigate HIV/AIDS impacts. Analysts update DEMMIS every month, making it sensitive to changing local conditions. This helps measure program effectiveness in specific areas. Initial results have been encouraging, but the program has been in existence for a relatively short time and has only been tested in some provinces of Zimbabwe and South Africa. The program might be further analyzed to see how it can be improved and replicated elsewhere.

capacity of the education sector in Sub-Saharan Africa as spontaneous and unpredictable gaps in the system that need to be fixed rapidly. Some scholars have termed this phenomenon as the “Swiss cheese model.”^[7]

Another critical capacity-building initiative involves acquisition^e and distribution of ARV drugs as part of a work incentive package.^f Mere distribution of ARV drugs to educators is not adequate. A strong distribution and monitoring framework must be in place to ensure proper and effective usage of ARV drugs. Toward this end, Sub-Saharan African Ministries of Education and Health might develop a comprehensive and coordinated strategy modeled after programs offered by *Médecins Sans Frontières* (MSF), which offers ARV drugs only to people who have an elevated antibody count. Medicines Sans Frontiers, somewhat controversially, requires patients to bring along a companion as a “goodwill” demonstration of their willingness to talk about the disease with others.^[12]

CONSTRUCTIVE PARTNERSHIPS

Cross-sectoral partnerships among international aid agencies, national and local governments, private sector and public sector enterprises, non-governmental organizations, and civil society groups are critical in mitigating AIDS impacts on education. Only a sustained, comprehensive partnership approach can effectively combat AIDS, because the pandemic is spread across society: to deal with it, all segments of society—businesses, civic groups, religious groups—must embark on a joint crusade. In Botswana, a group of 40 AIDS organizations facilitated information sharing between different sectors—public and private—to improve service provision.^[13]

Teachers and administrators may wish to adopt best practices from the private sector's approach to protecting its workforce from the pandemic. Best practices can be adopted from the models like that of Debswana Diamond Company (Botswana).^[14] In 1991, the company created program coordinators at its mining

locations, establishing a comprehensive AIDS management workplace policy in 1996 and distributing ARV drugs to infected employees beginning in 2001.

Information technology (IT) companies and Ministries of Education might collaborate to automate information storage and retrieval.^[15] The personnel-based, public information system creates critical barriers to system functioning when key individuals—clerical and administrative—are absent, sick, or deceased. Growing awareness created partnerships between donor agencies like USAID and multinational corporations like Cisco, providing training to teachers and administrators in Uganda.

Training provided by IT companies might enhance social capital and skill levels in Sub-Saharan Africa, boosting economic growth and development prospects.^[15] Use of the Internet builds networks in ways that expand interpersonal relationships not always possible in face-to-face relationships. Nevertheless, IT usage can supplement or complement face-to-face relationships. Non-governmental organizations, like *Médecins Sans Frontières*, and international governmental organizations (IGOs), like the World Health Organization, can provide expertise to the Ministries of Education to help with ARV drug distribution and procurement.

Some have argued that national governments, international agencies, and NGOs in Sub-Saharan Africa might pool their resources to develop an AIDS vaccine. African universities and research laboratories can help in this endeavor. In 2004, G8 countries set aside funding for research, and South Africa has launched its own research initiative. Regardless, Ministries of Education must try to ensure cooperation across agencies on this front.

Partnerships between developed (North) and developing (South) countries are critical to mitigate HIV/AIDS impacts on education. Retired teachers from Northern countries can serve for 2 years in Sub-Saharan Africa to help replenish teacher ranks, a program modeled after the Peace Corps' “Crisis Corps” initiative in Swaziland. African universities also might team up with international aid agencies to create capacity-building initiatives, like the Mobile Task Teams.^g

Grant writing programs might also be developed for African universities to help local governments, civil society organizations, and schools access internationally available development funds. This is because an increasing number of donor aid funds require that recipients make a formal request for fund disbursement.

^eFor a review of issues surrounding ARV drug acquisition, see: Yolanda, T., Ed. *Battling HIV/AIDS: A Decision Maker's Guide to the Procurement of Medicines and Related Supplies*; World Bank: Washington, DC, 2004.

^fThis program is controversial, raising ethical questions such as why should teachers receive priority over any other group. Proponents apparently feel that teachers are critical to Africa's development and central in many ways to the progress of the pandemic, and therefore need to become part of the solution, not a continuing part of the problem.

^gSee USAID website at http://www.usaid.gov/locations/sub-saharan-africa/publications/docs/hived_brief.pdf for details of this program.

LONG-TERM STRATEGIES

The best long-term strategy might be to establish teacher training institutes across Sub-Saharan Africa. The most effective one to curb the ill-effects of AIDS on education is to create academies in key regions of Sub-Saharan Africa. These institutes would train teachers and administrators to narrow the gap between supply and demand in the education sector. They would also improve the capacity of the education civil service and administration to perform more effectively and efficiently.

Long-term national and regional strategies are required for development generally, as well as the education sector in particular, for ensuring the best possible prioritization of development issues and resource allocation in Sub-Saharan Africa. Countries like Uganda have also made AIDS a high priority in its development plan, and have seen great progress in reversing the epidemic. The New Partnership for Africa's Development (NEPAD)^h might want to lead on this front by creating a viable regional strategy to prioritize development issues and create a comprehensive regional strategy for education in the wake of the HIV/AIDS pandemic.

CONCLUSIONS

Strategies discussed here may not solve HIV/AIDS problems for teachers and administrators in the Sub-Saharan Africa. Lack of fieldwork and program/project evaluations constrains our ability to predict the likelihood of success for these strategies. Nevertheless, many observers believe that it is important to try to turn the pandemic, even if only through uncertain experimentation.

REFERENCES

1. Stanecki, K. *The AIDS Pandemic in the 21st Century*; Office of HIV/AIDS, Bureau of Global Health, U.S. Agency for International Development, 2004.

2. UNAIDS. *Report on the Global AIDS Epidemic*, Executive Summary; UNAIDS: New York, 2004.
3. Kelly, M.J. *Planning for Education in the Context of AIDS*; UNESCO/International Institute for Educational Planning: New York, 2000.
4. Bennell, P.; Hyde, K.; Swainson, N. *The Impact of the HIV/AIDS Epidemic on the Education Sector in Sub-Saharan Africa*; University of Sussex: Sussex, England, 2002.
5. Haacker, M., Ed.; *The Macroeconomics of HIV/AIDS*; International Monetary Fund: Washington, DC, 2004.
6. Goliber, T. *Exploring the Implications of the HIV/AIDS Epidemic for Educational Planning in Selected African Countries*; The Futures Group International: Washington, DC, 2000.
7. Mumuskshu, P.; Terry, F.B. *Mitigating HIV/AIDS' Impacts on the Civil Service and Teachers in Sub-Saharan Africa*; National Academy of Public Administration: Washington, DC, December 6, 2003; available at www.napawash.org.
8. AIDS Campaign Team for Africa. *Turning Bureaucrats into Warriors*; World Bank: Washington, DC, 2004; available at www.worldbank.org/afr/AIDS.
9. Desmond, C. *Human Capital and the HIV Epidemic in Sub-Saharan Africa*; ILO Program on HIV/AIDS: Geneva, June 2002.
10. Peter, B.-W. *HIV/AIDS Impact on Education in Africa*; Arusha, Tanzania, October 2001.
11. UNAIDS and Ministry of Education in Zimbabwe. *Direct Education Management Information System (DEMIS)*; Zimbabwe, 2003.
12. MSF (Doctors without Borders). *Demystifying ARV Therapy in a Resource Poor Setting*; 2003, available at www.msf.org.
13. UN Development Program. *The Impact of HIV/AIDS on Human Resources in the Malawi Public Sector*; UNDP: New York, 2002.
14. US Agency for International Development. *How Are African Businesses Responding?* 2003, available at www.debswana.com.
15. *Information Technology and Governance*, A Conference Sponsored by National Academy of Public Administration and the Ghana Public Services Commission, Accra, Ghana, May 30–June 5, 2004; available at www.napawash.org.

^hSee NEPAD website at <http://www.nepad.org> for a description of this organization.

Model Contracts to Reduce the Risks in Complex Information Technology Procurements

Michael Asner

Michael Asner Consulting, Surrey, British Columbia, Canada

INTRODUCTION

Model contracts are valuable tools. Their use can reduce the risks associated with complex contracts, identify issues for consideration and resolution during the negotiation process, and serve as models for similar procurements in other jurisdictions.

Contract negotiation is an important and critical element in dealing with suppliers. It is part of the procurement process. There is little point in including an incomplete or poorly drafted contract in the request for proposal (RFP) stage, or introducing it into the negotiation process. It is the contract that formalizes the intent and agreement of the parties and must therefore accurately document all the terms and conditions. The contract must be enforceable in court. It must also reflect the understanding and expectations of both parties.

In many RFP processes, it is the procurement professionals who have the responsibility of ensuring that the contract protects the agency, and reflects a shared understanding of the entire business deal. Traditionally, procurement officers have been at a disadvantage when dealing with suppliers. Few procurement officers are lawyers; in contrast, suppliers are very knowledgeable about protecting their rights and the nuances of the law. Often, procurement personnel are dealing with unfamiliar technology, and to make matters worse, many procurement teams have limited access to legal help during the contract development phase of procurement. Simply stated, procurement professionals are at a disadvantage.

accompanied by some training, a set of instructions, and possibly a web-based, fill-in template. It is more than simply a list of contract clauses.

Model contracts are only developed for commonly recurring complex purchases, e.g., the acquisition of a new computer system, software, or a systems integrator. Almost all model contracts are for information systems and technology because of the volume and frequency of technology-based RFPs, the high value of the procurement, and the inherent complexity of information technology (IT) acquisitions. They serve as a starting point for development of the specific and complete contract to be included in the RFP.

Most procurement personnel are not lawyers, nor are they subject experts knowledgeable about network protocols, e.g., systems integration or bridge construction. Over the years, many organizations have expended great amounts of time and energy to develop model contracts. The use of these contracts reduces the risks to buyers by identifying issues, risks, and remedies. “Home-grown” contracts are simply too risky, and they prolong negotiations, increase prices, and often abandon remedies which would be available to the agency.

In many jurisdictions, the procurement authority has prepared a set of model contracts designed to meet the majority of the needs of their agencies and departments. These needs are not unique to procurement in the United States; they exist in many other places including Australia, Canada, Hong Kong, and the United Kingdom. In Australia, one of the state governments has developed a set of principles for their model contracts. These principles can be applied in other jurisdictions. They believe that contracts should be based on the following:^[1]

- terms and conditions that are fair to all parties—the contract should incorporate a spirit of mutual trust;
- schedules and annexes that allow users flexibility in customizing contracts to suit their individual requirements;
- a format that is user-friendly and in “plain English” style;

OVERVIEW

Model contracts are an attempt by organizations to ensure that their rights are protected, without relying heavily on expert legal help for each new contract. A model contract serves as a template for developing a specific contract. There is a high initial cost associated with developing a model contract for use throughout an organization. Ideally, the model contract is

- a style and language suited to the end users of the products and services;
- consistency in terminology, definitions, and style;
- inclusion of the following elements:
 - a dispute resolution mechanism (e.g., appointment of adjudicator, referee, or expert),
 - performance incentives,
 - allocation of risks.

The last word on this topic goes to the State of Victoria, Australia, where its purchasing board cautions the agency procurement people about using standard or model contracts. They state that the use of the model contracts is not a substitute for:^[1]

- people thinking about, then clearly stating what they require;
- negotiation of terms to suit particular circumstances, where necessary;
- proper planning early in the tendering process;
- careful selection of the supplier;
- competent contract management; and
- creating and maintaining a working relationship with a supplier.

A BRIEF SURVEY OF SEVERAL JURISDICTIONS

A model contract is much more than a standardized set of terms and conditions which is the basis for a purchase of goods. It is complete, fully describing specific types of procurements such as computer applications software. Models differ in their level of detail, the amount of instruction and tutorial material provided, and the topics covered.

Many jurisdictions provide good examples of these model contracts and can be major sources of information. For example, the state of Victoria, Australia, has developed model contracts, a checklist of contract clauses, and a 35-page user guide.^[1] The state of Washington, U.S.A., has an 8-page guide that includes a checklist of clauses, and four model contracts for information technology.^[2] In 1999, California developed new draft model for IT contracts.^[3] All of these resources can help public procurement professionals in planning the RFP process and in negotiations.

In the remainder of this section, we will provide a brief survey of how model contracts are dealt with in several representative jurisdictions. All of these contracts deal with different aspects of information technology, reflecting the frequency, volume, and complexity of procurement acquisitions. Each of these examples is web-based, so that the information can

be centrally managed and, at the same time, easily accessible by people throughout the organization.

We begin with the **State of Montana**.^[4] This is the starting point for the development of a complete, robust, easy-to-use model contract facility. Their web site contains the minimum amount of data, a table of 84 “Standard Contract Language and Forms.” There are two types of entries. First, contract elements which should be used by all state agencies “in every bid, proposal, purchase order, or contract.” An example is the clause dealing with Access and Retention of Records. The second type of entry is contract elements, which are to be used as required. An example is the clause dealing with “alternate bids,” which is only used in certain types of procurements.

The system in the **State of New York** contains much more data than that in Montana.^[5] It is further along the road to the ideal model contract facility. The Procurement Services professionals in New York provide their users with a 7-page guideline to promote “a better understanding of the types of contracts available and a “blueprint” for effectively utilizing the many types of technology contracts used.” These guidelines deal with the different types of contracts and different procurement strategies. It provides guidance on establishing technology contracts, and on how to process a known requirement. While it does not contain model contracts, it does provide access to all current technology contracts, and their underlying RFPs.

The **State of California** has developed a set of model information technology contracts.^[6] To build a draft contract, a purchasing official simply uses an approved “Standard Agreement Form” and attachments. The attachments consist of a Statement of Work for the services required, and terms and conditions selected from five Model IT contracts:

- Information Technology General Terms and Conditions,
- Information Technology Purchase Special Provisions,
- Information Technology Maintenance Special Provisions,
- Information Technology Software License Special Provisions,
- Information Technology Personal Services Special Provisions.

According to the Procurement Division, the benefits of using the new model include the following:

- The use of one model for all IT goods and/or services purchases, whether using a formal bid exceeding \$500,000 or informal bids below \$500,000.
- Statewide standardization.
- Increased familiarity with the elements and provisions of the model.

Washington State^[7] provides another example of a model contracts facility for information technology. It is composed of a set of instructions and model contracts for equipment, software, purchased services, and personal services. Each contract, with its embedded instructions, is between 40 and 60 pages in length. The model contract for software, for example, contains 71 clauses, 3 schedules, and 2 exhibits. Obviously, not every clause will be required in every contract, nor will this contract deal with every issue related to a specific software acquisition. The instructions caution the agency that contracts should be created from the model plus other sources to meet specific needs.

The model contracts and instructions are available from the state's web site. The contracts contain color-coded instructions and fill-in-the-blanks so that agencies can create draft contracts without a separate set of instructions. While specific laws and legal requirements vary from jurisdiction to jurisdiction, these contracts can be used as a starting point—a checklist of issues—by procurement professionals outside the state of Washington.

The **Government of Australia** has set the standard for model contracts.^[8] It has developed a Government Information Technology and Communications (GITC) contracting framework. This was performed as a cooperative effort between industry and the Australian federal and state governments. The GITC web site contains extensive information useful in developing contracts, including the model contracts themselves. This version (Version 4) was developed to achieve the following outcome:

- Simplifying the contracting process for the purchase of IT-related products and services.

After developing their model contracts and user documentation, they made all of the information available via the Web. To increase the ease-of-use and access to this information, they implemented two innovations:

1. There is a web-based interactive program which anyone can use to develop a contract. You simply log on to the site and proceed to select contract clauses and input data. At the end,

you save your contract. The program provides significant content and functionality:^a

- Creating a new contract,
- Editing the contract details,
- Saving the contract,
- Printing the contract,
- Adding or removing categories from the contract,
- Opening an existing contract.

2. One can download a file, their “User Guide,” which is in fact their entire web-based model contracting system. As soon as this file is downloaded, the user can have access to their model contracting system through the user's own computer without having to access the Internet. Information includes:

- User Guide
- Frequently Asked Questions
- Contacts
- About GITC
- Head Agreement
- Terms and Conditions
- Appendices

The amount of information that this site provides is excellent, and the web site design makes it easy to access, locate, and use this information.

CONCLUSION

Model contracts are used increasingly by various parties. They are a good investment in any jurisdiction that purchases lots of information technology goods and services. With increasing public scrutiny, more litigation actions from aggrieved suppliers, fewer procurement staff, and decentralization of contracting, model contracts will become a major tool for imposing standards on the procurement process.

REFERENCES

1. Government Information Technology Conditions, Terms and Conditions Guidelines, Victoria Government Purchasing Board, www.vgpb.vvic.gov.au/polguid/guid2g.htm.
2. IT Portfolio Management, Policy, Standards and Guidelines, Department of Information Services, State of Washington, www.wa.gov/dis/portfolio.

^aThe GITC web site provides a mechanism for building contracts. With the introduction of GITC 4, this feature has been extended to provide the following additional functionality:

- Contracts can be saved (as XML) for later on-line editing.
- Contracts can be opened (uploaded) for editing.
- The products and services categories covered by the contract can be changed as soon as the contract has been created.

3. June 1999 New Draft Model I.T. Contracts, California Department of General Services, www.pd.dgs.ca.gov/acqui/itcmnew.asp.
4. State of Montana, Standard Contract Language and Forms, 26 pages, <http://www.discoveringMontana.com/doa/ppd/stst.htm>.
5. New York State Office of General Services, Procurement Services, Purchasing Memorandum, Sept. 18, 1998, 7 pages, http://www.ogs.state.ny.us/purchase/snt/CL301_0.htm.
6. California Department of General Services, Procurement, June 1999 New Draft Model I.T. Contracts, 4 pages, <http://www.pd.dgs.ca.gov/acqui/itcmnew.asp>.
7. Washington State, IT Portfolio Management, Policy, Standards and Guidelines, <http://www.wa.gov/dis/portfolio/>.
8. GITC 4 is version four of the Government Information Technology and Communications contracting framework. The web site containing this information is called “GITC 4” and can be found at: <http://www.gitc.fianance.gov>.

Models of Spatial Finance and “Voting with Your Feet”

Ray Gonzales

*Graduate Department of Public Administration, Rutgers University,
Newark, New Jersey, U.S.A.*

INTRODUCTION

This article explores the idea of housing choice as it was envisioned by Charles Tiebout. Centered around the idea that local government spending patterns influence the way people choose neighborhoods, his theory has done much to stimulate thinking and generate research in many disciplines. It is the intent of this piece to summarize and analyze that literature by examining extensions of the model, which include jurisdictional competition, capitalization, and the Leviathan Hypothesis. To assist the reader further a discussion of measurement issues, and their implications for interpreting research, is included.

SPATIAL FINANCE AND THE TIEBOUT MODEL

Spatial finance refers to the impact of taxpayer preferences on municipal fiscal characteristics. Charles Tiebout^[1] used spatial finance to speculate about rational choice determinants of neighborhood selection via the market mechanism of taxpayer mobility. The idea is that as the number of jurisdictions (e.g., municipalities) increases, the variability in the number of budgetary profiles also increases because financial decisions reflect the predilections of the respective constituencies. People gravitate toward, and are happiest in, those communities where spending patterns reflect their value choices (i.e., Tiebout sorting). Tiebout did not intend to reflect reality—just articulate the assumptions needed to develop a market theory of public choice in neighborhood selection. For such a theory to work, there were numerous assumptions to be made. Taxpayers would have to have complete mobility, knowledge, and disregard employment considerations during neighborhood selection; there would be a variety of communities from which to choose; for every pattern of community services, there could be no spill-over effects; there would exist an optimal community size (the greatest benefit at the lowest cost)—communities below the optimal size would seek to increase their numbers working towards an economy-of-scale, while communities above this size would seek to lower them; finally, and every jurisdiction should be limited by some fixed variable (e.g., geography). This is

the base theory. There is no mention of the size of government beyond the optimal size hypothesis or of the motivation of public officials, nor is the concept of competition among jurisdictions mentioned. These ideas and others would follow from these assumptions.

TIEBOUT EXTENSIONS

Tiebout's framework did more to stimulate theory and research than to explain the way society works. To date, the most comprehensive resource for articulating the conclusions on various themes in research is an article written by Dowding, John, and Biggs,^[2] reviewing over 200 pieces through the early 1990s. Consensus in the literature examining the many facets of Tiebout is scarce. Problems with method and an inability to adhere to the assumptions of the model have made drawing conclusions about concepts difficult.^[2] What is more, some arguments in favor of a Tiebout model have not held up to research findings. At best, the support for the base theory is mixed; the same holds for research examining residential contentment and Tiebout sorting.

Sorting is usually measured in terms of jurisdiction homogeneity—the extent to which people similar in type live together, figuratively speaking. Homogeneity can be operationalized in a variety of ways, including race, income, or preference for public good. Often, using methodological rigor as an interpreter, opinions on sorting remain cautious because some studies support hypotheses whereas others do not. The most that can be said for sorting is that jurisdictional cost–benefit issues are factors in making relocating decisions, but this does not approach the support needed to affirm sorting in general.^[3] Other work views school as a public good and comes to different conclusions about relationships with, or implications for, homogeneity.^[4–6] Some finds a tacit implication for sorting,^[4] showing an increase in the number of low-income households in states where decentralized public school financing is upheld.

Expanding Tiebout's theory to include a competitive element arose out of a need to overcome an obstacle described by Paul Samuelson^[7]—that a market analogy for the provision of public goods was not achievable due to the absence of a direct exchange mechanism. That is, a person paying for a service may

not be the only one experiencing the benefit and a person experiencing a benefit may not be the one paying for it. To get around this, researchers employing Tiebout assumptions conceptualized a competitive environment in which municipalities competed for resident resources. Almost everyone attributes this idea to Tiebout, but it is not clear how a competitive element can be introduced if local budgetary profiles are fixed, as his 1956 work maintains. Whatever the case may be, by relaxing the optimal size assumption and presupposing that the number of taxpayers is finite, researchers could hypothesize that municipalities would do their best to keep costs low via competitive tax rates and benefit levels high (to keep taxpayers satisfied with service and attract the greatest number of residents possible to that jurisdiction). Thus, a model of public service provision simulating a market environment was created. Today, Tiebout is synonymous with jurisdictional competition, but there is almost nothing in the research showing that governments compete with one another for residents, and what evidence is available is indirect or implied by the data.

The competitive implication carries with it a number of hypotheses for research in urban politics, public economics, and public administration. Using structural characteristics as predictors, one hypothesis held that large amounts of inter- and intrajurisdictional competition in a geographic area resulted in greater efficiency gains (i.e., decentralization hypothesis). Areas with fewer (and consequently larger) governments should produce public goods less efficiently than smaller jurisdictions. The review provided by Dowding, John, and Biggs^[2] suggested that this has been found to be true in the most general sense. Through the early 1990s, there were slight variations on this conclusion when exploring differences between single- and multi-purpose jurisdictions. One rationale for the finding linking jurisdictional size to efficiency was that as the number of jurisdictions increased, so did taxpayer ability to hold public officials accountable for services and policy choices. Unsatisfied residents “vote with their feet,” as mobility became a mechanism for accountability and efficiency. This has not been conclusively supported if you consider that research has shown Tiebout factors to be more important in the decision to enter a community than in the decision to leave. Refreshingly, work has shifted focus to more advanced stages of performance measurement by focusing on the effectiveness of public service provision,^[6] signifying an attempt to move beyond poorly operationalized efficiency pervading the research.

The tax competition literature possesses conflicting views on Tieboutian efficiency. One line of work says that competitive tax rates and regulation do not result in the efficient provision of public goods. Some believe that industry supplants residents as the focal point of

jurisdictional competition. Thus, communities alter tax policy and operating regulations to make themselves attractive to business. Others see the faults of tax competition in an argument similar to Theodore Lowi’s criticisms of interest group liberalism. Here, tax rates and regulations reflect the preferences of an influential minority, effectively replacing policy aimed at larger societal concerns with legislation benefiting the few. The quality of public goods falls, costs are externalized (e.g., neighboring communities are polluted), and property values are lowered (i.e., race-to-the-bottom). The alternative, supporting Tiebout and tax competition, contends that efficiency gains are associated with the competitive element in multi-jurisdictional taxation when examining the behavior of small groups.^[8] It is the suppressed tax rate that leads to greater efficiency and enhanced quality of public policy (i.e., race-to-the-top) because lower revenue streams force public officials to economize.^[9]

Another area of inquiry relating to taxation deals with an idea called capitalization. At the heart of capitalization is the distinction between a *tax base* and a *tax rate*. Crudely put, the tax base can be considered the value of property in a given community. The tax rate is the proportion or percentage of the property value assessed at tax time. If the tax rate is high, demand for houses should be low, and so should the tax base. If the quality of public service is high, there should be a corresponding demand for housing, a high tax base and a low tax rate. As an explanatory mechanism for Tiebout, capitalization does not do a good job, unless you know the effect of the tax rate and local services on property values (tax base).^[2] Schneider, using 1970s data, explored this effect and found that local fiscal policy affected the tax base.^[10] He also found that the tax base was negatively related to the tax rate. Although this suggests support for capitalization, caution should be exercised in accepting this finding because the coefficients speaking to the strength of relationship were quite low. New work looks at school finance reforms,^[5] arguing that centralizing local public school finance erases inequalities existing between rich and poor districts. Consistent with capitalization, property values and rents subsequently rise, and poor persons are unable to subsist in these municipalities.

A PROBLEM WITH MEASUREMENT

Conclusions about Tiebout derivations depend heavily on interpretation, knowledge of method, and the operationalization of variables. For example, in social science research there are primarily two levels of analysis—the micro and the macro. At the risk of oversimplifying the relationship by way of employing the distinction in a Tiebout framework, a microlevel slant

on research would take the individual as the unit of an analysis; meanwhile, a macrolevel approach would use the municipality as the unit of analysis. This is important because putting together a causal argument is easier when conducting research using microlevel operationalizations than when working at the macro level. This is important to the present discussion because the majority of the articles exploring Tiebout assumptions and implications measure concepts at the macro level, making causal arguments comparatively difficult. Rather than assess the processes associated with choice, most studies only look to see if Tiebout effects are visible. Thus, they look to a conclusion to make some judgment about the premises or the chain of events leading to that conclusion. This is known as affirming the consequent, a no-no in basic logic and reasoning in quantitative research.^a So why do we care? The conclusion that smaller municipalities provide services more efficiently than larger municipalities is an assertion supporting one aspect of the market view of public service provision. But it does so using research operationalizing variables at the macro level. This is where theory is important.

Does the fact that smaller jurisdictions provide public services more cheaply than larger jurisdictions reflect the competitive implication taken from Tiebout?

Or is it that larger municipalities tend to be more heterogeneous, economically speaking, and consequently have to spend more on welfare (i.e., appear more inefficient)? The point is, there are multiple competing explanations for a single conclusion, and there is no way to determine which is the correct one unless the level of analysis is lowered to that of the individual (the micro level). Efforts have been made to do this (see Ref.^[11] for an example), but for many research questions, there is little that can be done. One interpretation is grounded in pessimism about the nature of government. In this view, budgets grow because government has a monopoly on public service provision, creating an environment where self-interested bureaucrats have no incentive to minimize waste. Taxpayers cannot leave (the original market mechanism of mobility is removed in an area with few governments), expenditures are large, and service provision is inefficient. This is also known as the Leviathan hypothesis and an important extension of the Tiebout model from an administrative point of view.

THE LEVIATHAN HYPOTHESIS

The notion of a self-interested bureaucrat provides a microfoundation for the macrolevel assertion that

larger governments spend more than smaller governments. Some authors, employing a microfoundation, have shown that administrators do not maximize budgets^[12] or even go so far as to argue that budgets shrink.^[13] Still, many believe that it is the self-interested nature of bureaucrats working to increase annual programmatic allocations. The connection between this claim and support for Leviathan is not so clear when examining Niskanen's view. Niskanen, attempting to construct a theory of bureaucracy originating in the budgetary behavior of administrators, wrote that bureaucrats generally had budget maximizing tendencies, but that service provision was efficient. Later, he would modify his theory to include the idea that service provision was inefficient by saying that bureaucrats were no longer concerned with maximizing the budget overall,^[14] rather they were concerned with maximizing the *discretionary* budget. This means that bureaucrats are concerned with a part or *proportion* of the budget instead of the entire budget. When applying this distinction to the Leviathan hypothesis, things begin to fall apart.

Under the theory emphasizing proportions, the element explaining expanding budgets is missing because the bureaucrat is concerned with discretionary spending, not spending overall. It does not matter how big the budget is, as long as the proportion of discretionary spending increases. This is illustrated in devolutionary block grant funding (overall funding levels decrease in exchange for more discretion in terms of how to spend it). Furthermore, because service provision in the first model is efficient, there is no theoretical reason to believe that larger governments will be any more or less efficient than smaller governments in a Leviathan world. Both facets of the theory are needed to support the Leviathan hypothesis. The first version explains why budgets are big, and the second explains why there is waste. The problem with this is that the first and second views are not compatible with one another because service provision is efficient in one, but not the other. Brennan and Buchanan^[15] got around this by taking the perspective that any increase in amount of allocation results in an increase in surplus. Thus, a bureaucrat seeks to increase the allocation because the amount of money going toward discretionary spending will always increase, even if the proportion does not. Their theory still hints at the contradiction embedded in Niskanen via their concept of "surplus maximization" (p. 350), along with a host of potential problems relating to the cost of service provision. In the research, efficiency is operationalized in terms of per-capita expenditure or spending as a percentage of resident income, in most instances. This is a problem because it presupposes that the quality of service remains constant across jurisdictions. This has the effect of making low spending and low

^aIt is important to note that researchers are aware of the problem and qualify their claims to address this.

quality of service municipalities appear as the "better" service providers, when in fact service quality may be awful. The literature does a disservice to the Leviathan hypothesis by assuming away service quality and not factoring in redistributory expenditures. But Dowding, John and Biggs^[2] note that there is no redistributive element in competitive markets.

CONCLUSION

Rival explanations make it difficult to say why smaller governments provide service at cheaper rates than larger jurisdictions. This dilemma is indicative of many conclusions in the Tiebout literature because of the manner in which research is conducted. At the heart of this is the basis of causal reasoning; probably one of the most problematic facets of social science research. Even the most sophisticated econometric modeling will never exhaust explanations or account for unforeseen confounding variables, weakening the value of studies conducted with macrolevel data.

REFERENCES

1. Tiebout, C. A pure theory of local expenditures. *J. Polit. Econ.* **1956**, *64*, 416–424.
2. Dowding, K.; John, P.; Biggs, S. Tiebout: a survey of the empirical literature. *Urban Stud.* **1994**, *31* (4/5), 767–797.
3. Bratton, W.; McCahery, J. The new economics of jurisdictional competition: devolutionary federalism in a second-best world. *Georget. Law J.* **1997**, *86* (2), 201–278.
4. Aaronson, D. The effect of school finance reform on population heterogeneity. *Natl. Tax J.* **1999**, *52* (1), 5–29.
5. Dee, T. The capitalization of education finance reforms. *J. Law Econ.* **2000**, *43* (1), 185–214.
6. Hoxby, C. Does competition among public schools benefit students and taxpayers? *Am. Econ. Rev.* **2000**, *90* (5), 1209–1238.
7. Samuelson, P. The pure theory of public expenditure. *Rev. Econ. Stat.* **1954**, *36*, 387–389.
8. Wooders, M. *Multi-jurisdictional Economies, the Tiebout Hypothesis, and Sorting*, Proceedings of the National Academy of Sciences of the United States of America, Washington, DC, Sept. 14, 1999.
9. Perroni, C.; Scharf, K. Tiebout with politics: capital tax competition and constitutional choices. *Rev. Econ. Stud.* **2001**, *68* (234), 133–154.
10. Schneider, M. Local budgets and the maximization of local property wealth in the system of suburban government. *J. Polit.* **1987**, *49*, 1104–1116.
11. Teske, P.; Schneider, M.; Mintrom, M.; Best, S. Establishing the micro foundations of a macro theory: information, movers, and the competitive local market for public goods. *Am. Polit. Sci. Rev.* **1993**, *87* (3), 702–713.
12. Sigelman, L. The bureaucrat as budget maximizer: an assumption examined. *Public Budg. Finance* **1986**, *6* (1), 50–59.
13. Downs, A. Why the government is too small in a democracy. *World Polit.* **1960**, *12* (4), 541–563.
14. Niskanen, W. A reflection on bureaucracy and representative government. In *The Budget-Maximizing Bureaucrat*; Blais, A., Dion, S., Eds.; University of Pittsburgh: Pittsburgh, PA, 1991; 13–31.
15. Brennan, G.; Buchanan, J. Revenue implications of money creation under Leviathan. *Am. Econ. Rev.* **1981**, *71* (2), 347–351.

Models of the Policy Process

Thomas A. Birkland

Rockefeller College of Public Affairs and Policy, University at Albany, State University of New York, Albany, New York, U.S.A.

INTRODUCTION

The simplest definition of the term “public policy” is “the things government chooses to do or not do.” This very simple definition certainly raises more questions than it answers. How is public policy made? Who participates, inside and outside of government, to decide whether government should or should not act? Current models of the policy process seek to address these questions.

Since policy studies began as a distinct endeavor in the 1960s, two major categories of policy studies have emerged. First, the policy science and policy analysis tradition retains considerable influence and value in the study and practice of policy making.^[1,2] This research tradition is rooted in the analytic tradition that prizes doing “the greatest good for the greatest number” of people in a society. Policy analysis often considers whether a particular policy proposal will yield net societal benefits (although the definition of what constitutes a “societal benefit” is open to interpretation). A major critique of this tradition is that it focuses too much on the economic aspects of policies without admitting the complexity of politics in policy making.^[3] Policy process theories seek to understand how various elements of the sociopolitical environment interact with the structure and institutions of government to yield a set of policies. These theories are much more focused on the *politics* of policy, i.e., on the “who gets what, when and how” dimensions of politics famously outlined by Lasswell.^[4]

EARLY THEORIES OF THE POLICY PROCESS

In 1965, Easton^[5] conceived of politics and policy making as a *system* that contains inputs, outputs, and feedback. In Easton’s systems model, policies are the product of inputs, such as public demands. The system takes in these inputs, which are influenced by features of the policy environment, and then produces, as outputs, policies. Policies are laws, rules, and other authoritative statements of what government does or has chosen not to do. Thus systems models view policy as the product of many influences inside and outside of government (Fig. 1).

Systems models help isolate the key elements of policy making, but have traditionally left the workings of the system—the “black box”—underspecified. The challenge for social scientists has come in opening up this box to understand how groups, political parties, the public, and elected and appointed government officials work to promote or oppose public policies.

Similar to the input–output model is the classic “how a bill becomes a law” model. This model is an idealized process model of how the process works in Congress. It fails to extend on the basic constitutional and institutional rules to explain how or why particular ideas are translated into laws while others are rejected. It is therefore not a useful model for understanding the policy process overall.

The Stages Model of Policy Making

The “stages” or “textbook” model of the policy process was, for many years, the preeminent model. In the stages model, policy making proceeds in stages, from *issue emergence* and agenda setting to *implementation* and then to *evaluation* and *feedback* (Fig. 2), and shares many features of the systems model.

A main critique of this model is that it is advanced as a one-way model, in which the stages happen in order.^[6,7] In reality, policies are sometimes made without going through each step: some policies are quietly advanced without raising high on the agenda, and, in many cases, policies are not evaluated. In addition, the boundaries between stages of the policy process have never been particularly clear, when the stages are sometimes overlapping or intertwined, such as when implementation of a new policy causes problems to rise on the agenda to a level not seen during the agenda-setting phase. Still, the stages model has long helped social scientists to organize research around each stage of the process,^[8] such as agenda setting^[9–12] and implementation.^[13–15]

MODERN THEORIES OF THE POLICY PROCESS

As scholars realized that policy making is more complex than the stages model suggests, researchers

The policy making environment

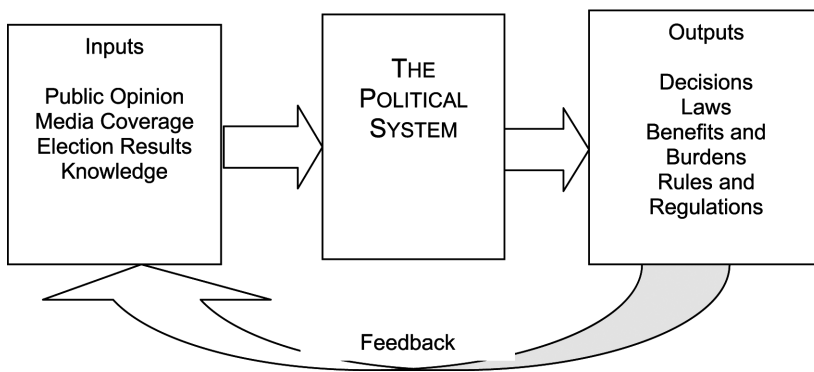


Fig. 1 A simplified input–output model of the political and policy system.

have developed increasingly sophisticated models of policy making. These theories include Kingdon's^[16] “streams” metaphor of public policy, Sabatier's Advocacy Coalition Framework,^[17] and Baumgartner and Jones's^[18] “punctuated equilibrium” model of agenda and policy processes.

Kingdon's Streams Metaphor

Kingdon built his idea of the “streams metaphor” by building on Cohen et al.'s^[19] description of decision making in complex organizations. Kingdon argues that policy issues gain agenda status and alternative solutions are selected when elements of three “streams” come together. Each of these three streams contains various individuals, groups, agencies, and institutions that are involved in the policy-making process. One stream encompasses the state of politics and public opinion (the *political stream*). A second stream contains the potential solutions to a problem (the *policy stream*). The third, the *problem stream*, encompasses the attributes of a problem and whether it is getting better or worse, whether it has suddenly sprung into public and elite (i.e., the key participants') consciousness through a focusing event, and whether it is solvable with the alternatives available in the policy stream.

Within any particular problem area, these streams run parallel and somewhat independently of each other in a policy area or domain until something happens to

cause one or more of the streams to meet in a “window of opportunity” for policy change. That trigger can be a change in our understanding of the problem, a change in the political stream that is favorable to policy change, a change in our understanding of the solvability of the problem given current solutions, or a focusing event that draws attention to a problem and helps open a window of opportunity. But the opening of the window does not guarantee that policy change will occur.

Kingdon provides a rich and multilayered description of policy making, from the early acceptance of new ideas about public problems to the active considerations of solutions as new public policy. However, one shortcoming of this model is that it does not describe the policy process beyond the opening of the window of opportunity.

The Advocacy Coalition Framework

Sabatier's Advocacy Coalition Framework (ACF) explicitly addresses two important features of the policy process: the role of groups and the importance of knowledge and learning in policy making. The pluralist tradition in political science argues that policy is made in a process of group competition. Sabatier understands, however, that dozens of groups working independently are unlikely to have as much power as they would have if they came together in larger groupings called *advocacy coalitions*.

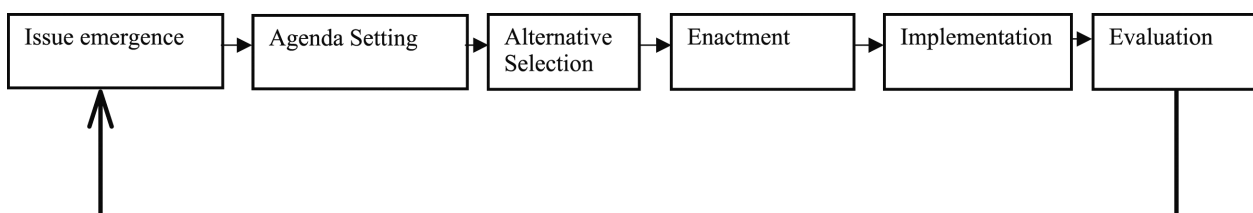


Fig. 2 The stages or textbook model of the policy process.

In the ACF, two to four advocacy coalitions typically form in a particular policy domain when groups coalesce around a shared set of values and beliefs. These coalitions engage in policy debate and compete and compromise with other coalitions. Some actors, called “policy brokers,” mediate competition between coalitions because the brokers are seeking particular policy outcomes or because they have an interest in promoting political harmony. Compromise is more likely when policies arise that do not threaten the advocacy coalitions’ core beliefs and values. Policy change is much less likely if polarization of advocacy coalitions is so great that there is no room for compromise in groups’ belief systems.

In the ACF, policy making is influenced both by “relatively stable” system parameters and by “dynamic (system) events,” with the interaction between the two promoting or inhibiting policy making. The stable parameters include the basic attributes of the problem area, the basic distribution of natural resources in the society, the fundamental cultural values and social structure, and the basic legal structure, which, in the United States, is the constitutional framework and judicial norms.

The dynamic features of the system include changes in socioeconomic conditions and technology, changes in systemic governing coalitions (partisan balance in the legislative or the executive branch, for example), and policy decisions and impacts from other subsystems. Change in the governing coalition corresponds to one example of change in the political stream in Kingdon’s model, while changes in socioeconomic and technological conditions influence the problem and policy streams. The activities of other subsystems can influence the policy, politics, and problem streams as their activities spill over into other policy domains. Like the streams metaphor, Sabatier’s ACF encompasses a variety of individual and institutional actors, and it views policy making as a process over years or decades. The ACF also considers the mechanisms for policy change (not simply the possibility for change, as in the streams metaphor) and more consciously encompasses the influence of implementation and feedback on the system.

Punctuated Equilibrium

Baumgartner and Jones borrow the concept of “punctuated equilibrium” from evolutionary biology to describe the policy process. They argue that the balance of political power between groups of interests remains relatively stable over long periods of time, punctuated by relatively sudden shifts in public understandings of problems and in the balance of power of the groups seeking to fight entrenched interests.

Key to their theory of equilibrium is the idea of the *policy monopoly*, which corresponds with the idea of policy subsystems. A policy monopoly is a fairly concentrated, closed system of the most important actors in policy making. Such a monopoly has an interest in keeping policy making closed because it benefits the interests of those in the monopoly and keeps policy making under some measure of control. Under the older “iron triangle” conception of groups in the policy process, this system will remain quite closed and quite stable for a long time. But Baumgartner and Jones argue that there are instances where the “equilibrium” maintained by policy monopolies will break down, greater and more critical attention to issues will follow, and rapid policy change will be the more immediate result. The policy monopolies themselves can break down or at least become more open-issue networks.

How do policy monopolies and their dominant construction of problems break down? First, greater media attention to an issue can begin to break open policy monopolies. Media attention to issues can grow when a small, but compelling or influential, group of people calls for change, to which other members of the policy community do not effectively respond. Baumgartner and Jones use the breakdown of the nuclear power monopoly to illustrate the effect of greater attention on a problem: in a mere 12 years, this domain was completely reshaped by the greater attention paid to the safety and constant problems that accompanied nuclear power.

This example illustrates how Baumgartner and Jones found that increased attention to a problem usually means greater *negative* attention to it. In this way, the “policy image” of various issues and policies can change. In the nuclear power case, the increased scrutiny of the industry began to break down the image of nuclear power as “the peaceful atom” creating power “too cheap to meter” to an image of danger and expense. This negative image was further re-enforced by the accident at the Three Mile Island nuclear plant in Pennsylvania in 1979.

Policy monopolies also break down as groups that seek access *to* and change *in* the policy-making process go “venue shopping” to find the best venue in which to press their claims. The media are one venue, and groups can seek to gain access to the courts or other units of government to gain access to policy debate. The reform of the congressional committee system and, most importantly, the increasing autonomy of subcommittees starting in the early 1970s have led to a greater number of venues in Congress for groups to find a sympathetic ear to influence policy making.

An important aspect of this way of thinking about policy is the long periods of stability followed by rapid change, and followed again by long periods of stability. In this way, Baumgartner and Jones argue that policy

change is not merely incremental and not always a state of constant flux: rather, policy remains stable, followed by a period of rapid change, then stability again, in sort of an S-shaped curve.

CONCLUSION

No one model of the policy process has yet been developed that can fully explain all the nuances and intricacies of the policy process. In addition, it is likely that the very complexity of politics and society will make a universal theory of the policy process impossible to achieve. But work continues on validating and improving on the models described in this article, with the goal of developing improved models to help us understand why public policy is or is not made.

REFERENCES

1. Weimer, D.; Vining, A. *Policy Analysis: Concepts and Practice*; Prentice Hall: Upper Saddle River, NJ, 1999.
2. Stokey, E.; Zeckhauser, R. *A Primer for Policy Analysis*; W. W. Norton: New York, 1978.
3. Stone, D. *Policy Paradox: The Art of Political Decision-making*; W. W. Norton & Co.: New York, 1997.
4. Lasswell, H.D. *Politics: Who Gets What, When, How*; Meridian Books: New York, 1958.
5. Easton, D. *A Systems Analysis of Political Life*; John Wiley and Sons: New York, 1965.
6. Nakamura, R.T. The textbook policy process and implementation research. *Policy Stud. J.* **1987**, 7 (1), 142–154.
7. Sabatier, P.A. Toward better theories of the policy process. *PS: Polit. Sci. Politics* **1991**, 24 (2), 144–156.
8. deLeon, P. The stages approach to the policy process: what has it done? Where is it going? In *Theories of the Policy Process*; Sabatier, P.A., Ed.; Westview: Boulder, CO, 1999; 19–32.
9. Cobb, R.W.; Elder, C.D. *Participation in American Politics: The Dynamics of Agenda-Building*; Johns Hopkins University Press: Baltimore, 1983.
10. Cobb, R.W.; Ross, M.H. *Cultural Strategies of Agenda Denial: Avoidance, Attack, and Redefinition*; University Press of Kansas: Lawrence, 1997.
11. Hilgartner, J.; Bosk, C. The rise and fall of social problems: a public arenas model. *Am. J. Sociol.* **1988**, 94 (1), 53–78.
12. Schattschneider, E.E. *The Semisovereign People*; The Dryden Press: Hinsdale, IL, 1975.
13. Nakamura, R.T.; Smallwood, F. *The Politics of Policy Implementation*; St. Martin's Press: New York, 1980.
14. Pressman, J.; Wildavsky, A. *Implementation*; University of California Press: Berkeley, 1973.
15. Van Horn, C. *Policy Implementation in the Federal System: National Goals and Local Implementors*; Lexington Books: Lexington, MA, 1979.
16. Kingdon, J.W. *Agendas, Alternatives and Public Policies*, 2nd Ed.; Harper Collins: New York, 1995.
17. Sabatier, P.A.; Jenkins-Smith, H. The advocacy coalition framework: an assessment. In *Theories of the Policy Process*; Sabatier, P.A., Ed.; Westview: Boulder, CO, 1999; 117–166.
18. Baumgartner, F.R.; Jones, B.D. *Agendas and Instability in American Politics*; University of Chicago Press: Chicago, 1993.
19. Cohen, M.D.; March, J.G.; Olsen, J.P. A garbage can model of organizational choice. *Adm. Sci. Q.* **1972**, 17, 1–25.

Models, Types and Attributes of

Jeffrey A. Weber

*Department of Political Science, East Stroudsburg University,
East Stroudsburg, Pennsylvania, U.S.A.*

INTRODUCTION

The term “model” suffers from numerous definitions that are often contradictory. Public administration and public policy, similar to social science, has adopted a wide variety of models: game models, trend models, probability models, casual models, network models, path models, and simulation models, to name but a few. Often, the term model refers to a representation of a specific set of relationships that are able to be empirically tested and provide some predictive value. Many times the term model is used interchangeably with theory.^[1]

Despite a wide variety of usages, the term model is separate and distinct from the term theory and should not be considered the same. A model is descriptive, while a theory is an explanation. In other words, a model does not answer the question “why” something occurs, while a theory provides a “why.” A model is a rendering of the specific events that comprise a phenomenon or a portion of a phenomenon. While a theory provides not only the events but also an explanation.

DEFINITION OF A MODEL

The many and varied uses of the term “model” have caused confusion in several fields. Therefore, some sense of order out of this confused state of affairs needs to be established. In order to clearly delineate a model from a theory, a model is defined.

A description which serves as an observable representation of a phenomenon or a portion of a phenomenon, which may be derived from a theory, or serve as a basis for the development of a theory.

This definition recognizes the descriptive nature of a model. Because a model is descriptive, its components are observable, while a theory is abstract, and its constructs are not observable. Furthermore, a theory serves to explain phenomena or phenomenon depending on the type of theory, but a model describes a phenomenon or a portion of a phenomenon. As such, it is possible to build a model from some of the constructs and their respective variables that comprise a

theory. Furthermore, a model could begin with variables that are later grouped into constructs, which, in turn, become the basis for the development of a theory.

Similar to a theory, models can be divided into different types. While theories are divided into a typology that reflects a theory’s scale and generalizability, a model’s type reflects its descriptive elements.

TYPES OF MODELS

Though there are no agreed upon typology or categorization of models, it is possible to discern from the descriptive elements of models four basic types:^[1] Object Models, Quantitative Models, Simulation Models, and Behavioral Models.

Object Models

An object model is a physical representation of an object. The most common version of this type of model would be a model car, plane, or ship that someone would assemble from a kit. When fully assembled, the model is typically a scaled representation of a larger object.

In public policy analysis, the variables in the model are depicted as scaled representations of the full-size objects. The interrelationships between the variables can be visually seen in the object, and then scaled measurements can be taken.

An example would be an architectural model of a community area, which could be used to help guide the implementation of a proposed community revitalization project. The model could be used to ensure proper infrastructure support, such as location and width of roads to support traffic patterns. Additionally, the model could be used to ensure fulfillment of all zoning ordinances.

Another example of an object model is found in the space policy arena. In determining the feasibility of landers and rovers on Mars, scaled models of the Martian surface and the rovers and landers were constructed. These object models permitted testing to be accomplished for feasibility and cost-benefit studies.

Quantitative Models

Quantitative models are some of the most popular types of analysis models. In quantitative models, the variables are depicted as numbers, and interrelationships between them are shown as mathematical formulas. Quantitative models are used in public policy analysis to model the interaction of a few variables. Examples can be found in most revenue or expenditure forecast models.

Simulation Models

A simulation model depicts the action of a phenomenon in a sensory manner. The most common form of simulation models are weather models, which provide us with hourly forecasts. In weather models, the different variables that comprise the weather (air pressure, moisture, temperature, etc.) and their interrelationships are depicted in a controlled moving fashion. By taking real-time measurements of the current state of the variables, the model then depicts the action that those variables will take as they interact, which serves as the forecast.

There are two primary differences between a simulation model and a quantitative model. First, a simulation model is a combination of mathematical language and symbolic language of computer program. Second, a simulation model involves a large amount of variables with interactions that have a high degree of complexity.

Behavioral Models

Behavioral models attempt to depict the numerous variables involved in human or animal interaction. These types of models can be constructed from reality as it is occurring. In other words, the researcher watches behavior of the subjects as the subjects are going about their normal routines. From these observations, a model is constructed.

Conversely, a set of contrived situations can be constructed in order to make observations of behavior in a controlled manner. These types of situations can involve social or group interactions within the context of a real situation. Or, situations and roles can be contrived that are isomorphic to reality.

ATTRIBUTES OF MODELS

Similar to a theory, a model has certain attributes. Boundaries are the first attribute of a model and, similar to a theory, a model's boundaries define the temporal and spatial characteristics. The temporal characteristics of a model refer to the time period depicted by the model. For example, an object model may depict a past event as an historical recreation, or may be used in a future sense, as here is what an object will look like. Spatial boundaries define the physical shape and dimensions of the model. The idea of a model having a shape and physical dimension is relatively easy to conceive for an Object model, but for a Quantitative or Behavioral model, it is a little difficult. An example for a Behavioral model would be the physical dimensions of where the role playing is occurring or the test group is being examined.

Variables are the second attribute of a model and are sensory entities. In other words, a variable is a component of a construct that has the quality of being present, or may be present in degree, as sensed by one or more of the five senses.^[2] Additionally, variables can be an attribute of a construct (previously defined) that is always present as a part of the construct. In other words, the variable is a part of the construct's identity. Conversely, a variable may be an attribute that is sometimes present, and if it is present, it exists in various degrees.

Interrelationships are the third and final attribute of a model. The interrelationships found in a model are more specific and detailed than those found in a theory.^[3] These interrelationships will depict some form of association or influence among the variables. Some may depict the influence as casual, while others may depict it as a possible association but not specify a cause-and-effect-type linkage.

REFERENCES

1. Miller, D. The roles of models in research design. In *Handbook of Research Design and Social Measurement*, 5th Ed.; Miller, D.C., Ed.; Sage Publications: Newbury Park, CA, 1991; 51–58.
2. Dubin, R. *Theory Building*, Revised Ed.; The Free Press: New York, 1978; 44 pp.
3. Bacharach, S.B. Organizational theories: some criteria for evaluation. *Acad. Manage. Rev.* **1978**, *14* (4), 500.

Moral Development Theory

Richard D. White Jr.

Public Administration Institute, Louisiana State University,
Baton Rouge, Louisiana, U.S.A.

INTRODUCTION

Moral development theory attempts to answer a fundamental question. How do people decide what is right and wrong? A branch of the philosophical study of ethics, or metaethics, moral development theory argues that all people in all cultures pass sequentially, invariantly, and irreversibly from lower to higher stages of moral reasoning; moral judgment is defined according to how an individual reasons rather than according to what the individual thinks; and values and ethics are developed as a result of the interaction between the person and the environment.^[1] In its most basic form, the theory rejects relativism and regards the affective domain as unimportant and irrelevant. However, the creator of the modern moral development paradigm, Lawrence Kohlberg (1927–1987), goes beyond the simple value-free framework and argues moral judgment *is* moral; it is not merely the application of cognitive skills or intelligence to moral questions or situations. Kohlberg insists on a close relationship between the level of moral development and procedural justice to insure a moral baseline for his body of theory. To him, moral development theory is grounded upon a Rawlsian concept of morality about how humans cooperate, how they manage conflicts, and what concepts of fairness they adopt.^[2]

Modern moral development theory is rooted deeply in the works of Piaget,^[3] McDougall,^[4] Baldwin,^[5] Mead,^[6] and Dewey.^[7] The theory owes much to Piaget, who sees the child as an active participant in the development of a moral philosophy through interaction with the environment. Piaget departs from psychoanalytic and social learning theories that view the child as a passive product of the environment and rejects the belief that children simply learn the norms of a group. Instead, a child autonomously develops through struggles to arrive at fair solutions.^[3,8] In the first of two stages, an impulse-governed, egocentric *morality of constraint* governs the child. Through interactions with other children, the child matures, de-centers, and enters a second, higher *morality of cooperation* stage. Intentionality, the child's ability to take into account intentions or motives in making judgments of praise or blame, is an important determinant. Piaget illustrates intentionality when he asks a

child to decide who is naughtiest: a child who knocks over 15 cups accidentally or a child who breaks one cup in an act of disobedience. Typically, children under nine base their judgment on the extent of material damage, while older children recognize the moral relevance of the intentions of the actor.^[3] Piaget assumes that moral development is innate, invariant, hierarchical, and culturally universal.

Also focusing on child behavior, Baldwin sees moral development as evolutionary. An infant first learns to distinguish persons from objects in the *projective* stage. Next, the child develops a sense of self in the *subjective* stage. In this stage, children visualize themselves as people among other people but as possessed of special feelings that they cannot observe in others. Finally, in the *ejective* stage, children develop a sense of otherness and think of others as having those same feelings they can detect in themselves.

From Dewey, Kohlberg derives his three levels of development (preconventional, conventional, and postconventional), and from Mead, the cognitive basis for role taking. While Piaget's theory includes elements of emotion and the affective domain, Kohlberg follows the exclusively cognitive approaches of Mead and Baldwin. Another early influence is McDougall, who describes moral development as a four-stage process: in the first, only pain and pleasure modify instinctive behavior; in the second, reward and punishment of society modify instinctive behavior; in the third, anticipation of social praise and blame control moral conduct; and in the fourth, man acts in a way that appears right regardless of praise or blame.^[4]

THE KOHLBERG MODEL OF MORAL DEVELOPMENT

Synthesizing the work of the earlier theorists, Kohlberg offers three levels of moral judgment, with each level representing a progressive shift in an individual's moral development (see Table 1). Each of the three levels contains two distinct stages. At the lowest, preconventional level, moral values reside in external happenings or physical needs rather than in persons and standards. The preconventional level is characteristic of young children who worry only about being caught

Table 1 Kohlberg's stages of moral development

Level	Self-perception	Stage	Value of life
Preconventional	Outside group	1. Obey or pay, punishment orientation. 2. Self (and sometimes others') satisfaction.	1. Confused with physical objects. 2. Instrumental to needs of possessor.
Conventional	Inside group	3. Win others' approval by helping them. 4. Law-and-order mentality; doing one's duty.	3. Based on empathy of family. 4. Based on legal rights and duties.
Postconventional	Above group	5. Respect individual rights; accept critically examined values. 6. Act with logically developed and universally accepted principles.	5. Life is universal human right. 6. Life is sacred.

and punished.^[1] In other words, *if I want it and I get it, then it is right*. Within the preconventional level are Stages One and Two. Stage One emphasizes an egocentric obeisance to superior power and a trouble-avoiding attitude. Stage Two represents an orientation whereby right action instrumentally satisfies one's own, and occasionally others, needs. Stage Two individuals conform only to obtain rewards, confuse the value of human life with the value of physical objects, and see human life solely as an instrument to their own satisfaction.

At the middle, conventional level, an individual recognizes a sense of responsibility to the group, the need for maintaining order, and admits that right is based on social inclusion or law. Within the conventional level are Stages Three and Four. Stage Three is an orientation toward pleasing others, avoiding disapproval of family members, and conforming to majority stereotypical images. Stage Four is an orientation toward doing one's duty, respecting authority, maintaining the given social order, and conforming to avoid censure by legitimate authorities and resultant guilt.

At the highest, postconventional level, an individual develops moral autonomy and acts according to universal ethical principles. Within the postconventional level are Stages Five and Six. Stage Five individuals conform to maintain the respect of the impartial spectator judging in terms of community welfare. Life is a universal human right. Stage Six individuals conform to avoid self-condemnation, with mutual trust, respect, and conscience directing their actions. Stage Six is an orientation to conscience or principles, not only to ordained social rules but also to principles of choice appealing to logical universality and consistency. Human life is sacred.^[9]

THE DEFINING ISSUES TEST

In 1971, the empirical application of Kohlberg's model expanded greatly when James Rest introduced the

Defining Issues Test (DIT). Responding to the need for a practical, validated, and objective method for assessing moral judgment, the DIT is a multiple-choice instrument that enables researchers to generate a quantified appraisal. Having been administered in thousands of studies, the DIT places Kohlberg's moral development theory on a more solid empirical footing.^[10] The test presents a subject with six moral dilemmas followed by a list of questions that a person considers in choosing what ought to be done in each dilemma. The subject considers each dilemma and ranks the importance of statements typical of the different stages of moral reasoning. The dilemmas comprise a variety of social moral issues, including three from Kohlberg's original moral judgment interview. Grading of the DIT yields a continuous number, the *P* index, standing for "principled morality," and represents a percentage that each subject places on Kohlberg's Stages Five and Six.^[10]

P index scores do not indicate a subject's worth as a person, or his loyalty, kindness, or sociability. Instead, the scores draw upon the conceptual frameworks by which a subject analyzes a social-moral problem and judges the proper course of action. According to Rest, the DIT assumes that people at different points of development explain moral dilemmas differently and have different perceptions about what is right and fair in a situation. While Rest supports Kohlberg's general theory of sequential moral development, he does not accept the rigid scheme of six distinct stages. Rest rejects the single-stage concept and suggests that people fluctuate within a developmental range.

CRITICISM OF MORAL DEVELOPMENT THEORY

While an established paradigm, Kohlberg's theory remains controversial. A serious criticism is that moral development theory does not represent how people make actual decisions when they face real-life ethical dilemmas. People typically hold one set of beliefs, but

these beliefs are different from the actual rules they adopt in practice, especially in situations where the risks are high. In response, Kohlberg argues a strong relationship between the way people think about moral dilemmas and how they actually act. Two meta-reviews of more than 100 studies using the DIT found two-thirds of the studies reporting a significant correlation between moral development level and behavior. Studies showing a correlation cover a wide range of behaviors, including juvenile delinquency rates, drug and alcohol abuse, spouse, conscientious objection, and opinions on abortion.^[11,12]

Another significant criticism is that the Kohlbergian model has an inherent sex bias. Carol Gilligan faults Kohlberg for neglecting a woman's perspective, erroneously suggesting that women are morally deficient because of their tendency to understand moral issues in emotional and personal ways, and assuming the impersonal male justice perspective over the female care perspective.^[13] Despite Gilligan's criticism, most moral development studies find sex differences to be trivial, with the preponderance of recent research suggesting that males and females base their reasoning on a combination of justice and care. Several DIT studies find no statistical difference in moral development level based on gender, while one meta-analysis of 56 moral development studies administered to over 6000 male and female subjects reports that at every age and educational level, females scored significantly *higher* than males.^[14]

Critics also argue the moral stage model gives no evidence for sequentiality, claim empirical research fails to confirm stage development, and call for wholesale abandonment of the theory. Others fault Kohlberg for overemphasizing higher postconventional stages and justice-based ethical reasoning, while underrating some equally worthy values at the lower stages, emphasizing the cognitive aspects of moral reasoning at the expense of important affective dimensions, and focusing solely on Western ideology and creating a cultural bias. Kohlberg and his followers devote considerable effort to counter the varied criticisms. Taking almost a decade to prepare, Kohlberg published a detailed, three-volume defense of moral development theory.^[15]

MORAL DEVELOPMENT RESEARCH AND PUBLIC POLICY

Public policy scholars turn to moral development theory to explain the ethical behavior of public officials, compare the moral development levels of public servants with other segments of society and between cultures, and correlate moral development levels with public policy variables. Primarily using the DIT, studies find moral development scores to be associated with political ideology and views toward the Equal

Rights Amendment, nuclear arms race, environmental ethical reasoning, and capital punishment.^[16,17]

Moral development theory provides insights that are useful for public managers who need to recognize that different levels of moral judgment exist and that situational variables such as education, age, life experience, and degree of autonomy may affect an individual's level of moral development. Public managers should recognize that most adults and institutions, including legislatures, make moral judgment decisions predominately at the conventional, Stage Four level. Empirical research finds little evidence of the political system employing postconventional, principled moral judgment.

A significant implementation of Kohlberg's moral development framework in the public policy arena uses a specially designed instrument, the Stewart-Sprinthall Management Survey (SSMS). The instrument features ethical dilemmas structurally similar to the DIT but comprises scenarios more typical to those experienced by public servants as they perform their professional duties. Applying the instrument to a large number of public managers, SSMS researchers confirm that most respondents place an emphasis on Kohlberg's conventional, Stage Four of moral development, suggesting that public officials tend to make ethical decisions primarily in terms of law and agency rules and avoid individual discretion.^[18,19]

An area where moral development theory has significant potential within the public policy arena is pedagogical. Kohlberg, Dewey, and Piaget argue a strong linkage between education and moral development and emphasize that the educational bedrock is to identify and act upon moral principles. Research provides evidence that ethics education intervention can advance a subject's moral development. A meta-review of 55 DIT studies of ethics courses concludes that programs designed to improve moral judgment through intervention treatments of about 3 to 12 weeks produce optimal improvements between pretest and posttest average scores. Pedagogical methods featuring peer discussion of moral dilemmas show positive, consistent improvement in DIT results, while discipline-oriented, information-laden courses on traditional academic topics seem to be less effective.^[10] The need for improved ethics education for public servants is well documented. Stressing the need for ethics training and education, scholars see the potential value of moral development theory in the public policy classroom and propose the creation of formal, but practical, educational programs that nourish moral reasoning and ethical behavior. They call for a seamless, context-relevant, and continuous connection between graduate and in-service professional training.^[19]

The Kohlberg model also enhances the understanding of the consequences of designing public

organizations in different ways and the consequences of staffing organizations with employees of differing moral development levels. Ideally, it would be desirable for all people to function at the highest level. However, the employment of individuals who exhibit the highest postconventional moral judgment may not be best for the organization. Postconventional thinkers are individuals who are highly autonomous, question orders, and resist performing duties that violate their own high standards. The military provides an extreme but illustrative example.^[9] Can a military unit be effective if soldiers question the morality of war and resist the lawful orders of their superiors? Can an assembly line operate efficiently if workers exhibit autonomous behavior? Individuals who make moral decisions at the conventional level may be desired for the traditional hierarchical, industrial organization, while postconventional individuals may be desired for the creative organization structured in a less-rigid hierarchy.

REFERENCES

1. Kohlberg, L. Moral stages and moralization: the cognitive-developmental approach. In *Moral Development and Behavior-Theory, Research and Social Issues*; Lickona, T., Ed.; Holt, Rinehard and Winston: New York, 1976.
2. Rest, J. *Development in Judging Moral Issues*; University of Minnesota Press: Minneapolis, 1979.
3. Piaget, J. *The Moral Judgment of the Child*; Kegan Paul: London, 1932.
4. McDougall, W. *An Introduction to Social Psychology*; Methuen & Company: London, 1908.
5. Baldwin, J.M. *Social and Ethical Interpretations in Mental Development*; Macmillan: New York, 1906.
6. Mead, G. *Mind, Self, and Society*; University of Chicago Press: Chicago, 1934.
7. Dewey, J. Selected writings. In *John Dewey on Education: Selected Writings*; Archambault, R., Ed.; Random House: New York, 1964.
8. Power, F. Moral development. In *Encyclopedia of Human Behavior*; Academic Press: San Diego, 1994.
9. White, R. Public ethics, moral development, and the enduring legacy of Lawrence Kohlberg: implications for public officials. *Public Integr.* **1999**, 1 (2), 121–134.
10. Rest, J. *Moral Development: Advances in Research and Theory*; Praeger: New York, 1986.
11. Blasi, A. Bridging moral cognition and moral action: a critical review of the literature. *Psychol. Bull.* **1986**, 88, 1–45.
12. Thoma, S.; Rest, J. Moral judgment, behavior, decision making, and attitudes. In *Moral Development: Advances in Research and Theory*; Rest, J., Ed.; Praeger: New York, 1986.
13. Gilligan, C. *In a Different Voice: Psychological Theory and Women's Development*; Harvard University Press: Cambridge, MA, 1982.
14. Thoma, S. Estimating gender differences in the comprehension and preference of moral issues. *Dev. Rev.* **1986**, 6, 165–180.
15. Kohlberg, L. *Moral Stages: A Current Formulation and a Response to Critics*; Karger: New York, 1983.
16. Emler, N.; Renwick, S.; Malone, B. The relationship between moral reasoning and political orientation. *J. Pers. Soc. Psychol.* **1983**, 45 (5), 1073–1080.
17. DeWolfe, T.; Jackson, L. Birds of a brighter feather: level of moral reasoning and attitude similarity as determinants of interpersonal attraction. *Psychol. Rep.* **1984**, 54, 303–308.
18. Stewart, D.; Sprinthall, N. Strengthening ethical judgment in public administration. In *Ethical Frontiers in Public Management*; Bowman, J., Ed.; Jossey-Bass: San Francisco, 1991.
19. Stewart, D.; Sprinthall, N. The impact of demographic, professional and organizational variables on the moral reasoning of public administrators. In *Ethics and Public Administration*; Frederickson, H., Ed.; M.E. Sharpe: New York, 1993.

Motivation

Evan M. Berman

Public Administration Institute, Louisiana State University, Baton Rouge, Louisiana, U.S.A.

INTRODUCTION

The term “motivation” is defined as that which prompts a person to act in a certain way, as in the statement: “Tom is motivated to make a lot of money.” The term is also used to characterize the level of commitment that describes a person’s actions toward a goal, such as “Mary is not motivated to do this work.” A consensus exists that motivation refers to a person’s internal state. Thus, the frequently used expression “I am motivating Joe” is linguistically incorrect. Managers cannot give motivation to others or do motivation to someone. They can only increase or decrease the level of motivation that is already present in others. The term “motivation” is most accurately used in connection with a specific task or goal, such as in the above expression concerning Mary.

Motivation has been the subject of many studies. The purpose of early clinical studies has typically been to identify the motivations of workers. This is of considerable interest to managers, because highly motivated workers are thought to be more productive, as their energies are better aligned with the needs or goals of their organization. Recent managerial studies investigate management strategies for better aligning the goals of workers with those of the organization.

The terms energy (vitality) and motivation are related but distinct. Motivation is that which prompts a person to act in a certain way, whereas energy is the amount of effort that is available to a person for acting in a certain way. It is widely believed that motivation brings forth energy, and also that healthy adults have sufficient energy to accomplish their goals; hence, the lack of energy is typically interpreted as indicating insufficient motivation to achieve some goal or complete a task. However, it should be noted that medical conditions (such as the flu or clinical depression, which recently is increasingly diagnosed among U.S. adults and adolescents), and exposure to prolonged stress (resulting in, for example, “burnout”) may produce a (temporary) vitality deficit. In these instances, managers have an interest in assisting workers to restore their vitality to normal levels, while they direct human energy, at whatever level it occurs, toward the purposes of the organization.

THE MOTIVATION OF WORKERS: ABRAHAM MASLOW

Human beings have a complex set of needs that motivate them. The purpose of Maslow’s^[1] classic study was to identify these needs. According to Maslow, the needs of U.S. adults are of hierarchical nature, shown in [Fig. 1](#). The general ideas are as follows: 1) adults are motivated to satisfy their needs; 2) an order exists among their needs, and lower-order needs must be satisfied before higher-order needs can be satisfied; and 3) lower-order needs that have been satisfied no longer provide motivation for workers, although they would if workers fall back to that level.

Maslow’s hierarchy of needs has become widely popularized, in part, because it reduces the complexity of human motivation to a few manageable categories, described below. However, a few observations can be made. First, later studies suggest that the hierarchical order is not always replicated for all cultures (for example, some cultures view social belonging as a higher-order need than self-esteem). Rather than focusing on the order, it is useful to focus on the categories of needs. Second, the categories are broad, and managers cannot know exactly what motivates workers. This is in part because workers may not always be clear about what motivates them, and because some workers may guard against telling managers truthfully. Sharing what motivates workers with managers is to provide managers with “hot buttons” that can be used at any time against workers; it is akin to handing managers a loaded gun. The implication is that managers and their organizations must provide a wide range of conditions that enhance the motivation of workers; although managers may not know what motivates each worker, a diverse package is likely to contain sufficient elements to ensure that the goals of each worker are aligned with those of the organization. Third, managers and their organizations need to “update” the package of conditions that appeal to workers’ motivations, because expectations and conditions are different today than in the 1950s. Motivations may also vary across ethnic and demographic groupings.

Maslow’s motivational needs are discussed as follows. Traditionally, the motivational need of “physiological well-being” involves having adequate food and shelter, whereas the security need is met by attaining a

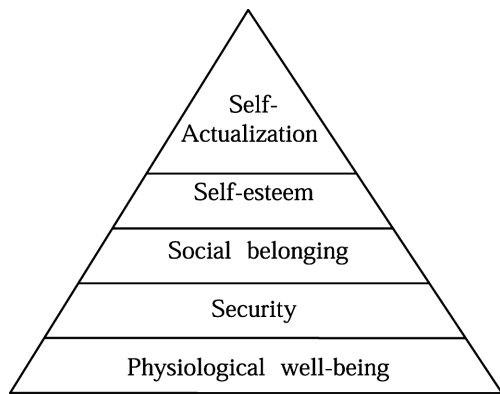


Fig. 1 Maslow's hierarchy of needs.

measure of job and retirement security. However, organizations have increasingly found ways to go beyond these motivational needs: some employers now provide exercise facilities, and they encourage workers to stop smoking. The rationale is that healthy workers are aligned with the organizational need for controlling health care and other employment costs. Also, many employers have revisited their retirement and health benefits to ensure that they are competitive and meet workers' needs, thereby ensuring the relevancy of these conditions as motivators. Some employers also have anti-workplace-violence programs, which may address security needs. Although it is unclear what percentage of workers regard any of these features as important job motivators, it may be a primary or secondary motivator for some, and a broad package of benefits is likely to contain some features that motivate many people. For example, retirement benefits are important for middle age and older workers.

It may be noted that physiological and security needs often are associated with money (salary and benefits). Studies show that money is a powerful motivator for workers with low salaries who struggle to pay their bills (that is, trying to meet their minimum levels of physiological well-being and security); few workers accept jobs when salaries are insufficient for that purpose. However, the importance of money tapers off among workers who regard themselves as being financially comfortable; for them, money is but one of several motivators, described below. For example, professional workers may feel that challenging work assignments are at least as important as money. These findings generally confirm that physiological well-being is a lower-order need. For workers who have generally satisfied this need, money, security, and physiological needs are necessary but insufficient to ensure a high level of motivation. For some workers, money remains an important motivator because of the ability to use money in satisfying other needs, such as status (discussed below), but not because of the need to meet basic physiological and security needs.

Some people have a need for friendly and close interpersonal relationships. People with a strong need for social belonging are motivated by harmonious working relationships. Coming to work is akin to being among friends; working with others is similar to enjoying the company of friends. They are also motivated to maintain harmonious working relationships. Strained personal relationships are demotivating, and such a workplace leaves a void of social belonging. In recent years, managers have been encouraged to increase a sense of teamwork and cooperation among employees, to ensure smooth working relationships, and to provide an opportunity for workers to meet their needs for social belonging. It is quite unclear what percentage of workers regard social belonging as an important need, whether managers are successful at providing conditions for social belonging, and whether workers experience sufficient social belonging in the workplace. A consensus exists, however, that employee picnics, casual Fridays, Christmas parties, and vague affirmations about the importance of teamwork are insufficient to satisfy the need for social belonging. McClelland^[2] also believes that affiliation is a basic human need, along with power and achievement (below). However, he notes that while a need for harmonious working relations is positive, too much need for such relationships can be negative, because it may prohibit open dialogue about performance, which may threaten the harmony of close interpersonal relationships.

Self-esteem is generally defined as holding oneself in high regard. High self-esteem is associated with getting recognition (praise) and positive feedback about one's actions, as well as positive self-talk. It should be noted that workers with a high need for self-esteem need not also have a strong need for social belonging; periodic praise from others about one's job performance may suffice. Management consultants frequently ask at seminars whether attendees (workers) perceive that they receive adequate recognition for what they do; typically, fewer than 10 or 20% of workers are reported to say that they do. This fact is disturbing, because it costs very little to recognize someone. An implication for managers is to find ongoing, new ways to recognize staff. Managers may believe that periodic salary raises or promotions suffice, but praise is more effective when it is immediate and directed toward a specific event. In this regard, some managers persistently provide praise and seek recognition for their staff from the organization.

Self-actualization occurs when workers are able to grow through their work. Although Maslow discusses self-actualization in terms of identity ("work become part of one's identity"), others discuss it as the need for achievement in ways that contribute to society and are consistent with one's need to develop. Achievement

(or accomplishment) is often an important human motivator. In this regard, the development and implementation of public programs is often an important achievement that allows for self-actualization among those managers who seek to improve society in this manner and relish the opportunity of developing their intellectual and political skills in new and oftentimes unpredictable ways, as they thrust themselves into the public realm. People with a high level of creativity are thought to have a high demand for such activities, and managers can enhance their motivation by providing challenging opportunities for achievement.

FURTHER PERSPECTIVES

Although Maslow's work identifies some of the most important motivating needs, other authors have identified others and provided important clarifications. David McClelland^[2] considers power to be an important source. The need for power is defined as the need to direct others and be in control of one's environment and destiny. Studies show that managers and executives often have a high need for power, far stronger than the need for social belonging or achievement. McClelland distinguishes between a positive and negative need for power. A positive need exists when managers want to accumulate power in order to improve society, including the conditions of their workers. A negative need for power exists when managers are driven by a need for personal dominance over others. Dictatorial behavior, including managers who are preoccupied with establishing and securing their fiefdom, is said to be motivated by a negative need for power. To satisfy the need for power, organizations typically reward managers with symbols and tools of power, such as the corner offices and higher salaries.

Herzberg's two-factor theory^[3] complements the above sources of motivation by distinguishing between motivators and demotivators (also called "hygiene factors"). Herzberg generally agrees that achievement, recognition, responsibility, and personal growth are important sources of motivation, but his empirical research shows that for a minority of workers, these are also sources of job dissatisfaction. For example, it may be that although self-actualization is an important motivator for some workers, it is not a motivator for all workers, and that these latter workers become demotivated when self-actualization is forced upon them. A lack of goal alignment occurs between the organization and the worker. Thus, although organizations generally need to provide a broad range of motivators, the way in which they couple them with individual workers may require some individual considerations: "one size" does not fit all (see "Psychological Contracts" below).

Herzberg also found that some factors are generally only regarded as demotivators that diminish worker motivation. In this regard, a majority of workers feel that policies, procedures, rules, and regulations demotivate them. For some workers, policies and administration do not demotivate, but they seldom, if ever, motivate. It is difficult to imagine being motivated by the prospect of dealing with rules and regulations, although they may turn some workers "on." Working conditions and relationships with supervisors also are viewed as demotivators. Good working conditions seldom motivate workers, but bad working conditions demotivate. The demotivational aspect of relationships with supervisors might be attributed to the fact that few people enjoy being subjected to power and controls, and being cajoled by others, no matter how considerately and tactfully this is done.

Herzberg also found that though salary and relationships with peers have mixed and moderate effects; they are moderate sources of motivation for some workers (compared with achievement, recognition, and responsibility) and moderate demotivators for others (compared to policy and administration or relationships with supervisors). Security and status were of negligible importance as either motivator or demotivator. Empirically, these factors seemed to matter little to most workers. (This is consistent with the above, when most workers already have their security needs met, when affiliation is assumed to be of moderate importance as a motivator to Americans who have socialized preferences for individualism, and when salary and status are viewed as derived from other needs, such as power.)

Victor Vroom^[4] notes that the way in which work is structured also affects motivation. Specifically, goals increase motivation when the reward associated with goal attainment is valued highly by the worker, when goal attainment is viewed as feasible, and when the expected effort that it takes to achieve goals is "reasonable" in relationship to the anticipated rewards. This is called expectancy theory. Others have expanded on this. For example, Locke and Latham^[5] note that specific goals increase performance more than abstract goals, that difficult goals accepted by employees increase performance more than simple goals, and that feedback leads to higher performance than no feedback. Feedback is also mentioned by others, especially in relationship to consequences. Skinner^[6] notes that behavior is shaped by its consequences: Positive consequences (including praise) reinforce behavior, and punishment (including negative feedback) suppresses behavior.

Finally, equity theory states that motivation is affected by how workers compare and evaluate their rewards relative to others, relative to policies and current practice, and relative to their standards as to what they feel they are worth. When workers feel that

they are being underrewarded, motivation decreases. When workers feel that they are being overrewarded, motivation increases. For example, when others are perceived as getting more than their fair share, worker motivation is diminished.

The implications for managers and human resource management are clear. Organizations need to provide workers with the following: 1) competitive salaries; 2) relevant benefits that address their needs for security; 3) meaningful rewards and recognition (and distribute them in equitable ways); 4) challenging (yet, achievable) assignments with opportunities for learning and skill development; 5) friendly and cooperative workplace relations; 6) assignments that allow workers to make a meaningful contribution to society; and 7) a sense among workers that they have meaningful control over their work environment (including rewards). At the same time, organizations need to minimize the demotivating effect of rules and regulations and negative supervisory relationships. The above also suggest that although these conditions should be present, individual workers vary in their needs and operationalization of any or all of these aspects. Thus, to further increase motivation, managers may wish to identify and respond to these individual variations.

PSYCHOLOGICAL CONTRACTS

The term “psychological contract” defines the unwritten understanding that an individual has with regard to the expectations, roles, and rewards that the worker has of the organization, as well as those that the organization has of the worker. The term is of recent vintage,^[7] though the concept appears in research in the 1970s. Psychological contracts are important to clarifying the needs of workers and aligning them with those of the organization, hence, increasing worker motivation. They are a tool for managing the alignment process.

Psychological contracts are unwritten understandings between two or more parties; they are not employment contracts and are not legally binding. The penalty for violating a psychological contract is worker withdrawal, that is, demotivation. Violation also undermines trust, which then becomes a barrier in establishing future productive working relationships between parties. Psychological contracts go beyond employment contracts, which typically provide

agreements about salary, benefits, grievance procedures, and working hours. Although these are important conditions that affect motivation, they do not clarify other conditions that also, and sometimes more strongly, affect worker motivation.

A psychological contract is established through a dialogue that is typically (though not necessarily) initiated by a supervisor or coworker. The purpose of the dialog is as follows: 1) identify what another person (say, an employee) is looking to get from the working relationship; 2) test that which the other person wants to get against that which the supervisor or coworker is willing and able to give; 3) establish what the supervisor or coworkers expect to get from the employee; and 4) establish what the employee is able and willing to give the supervisor or coworker. These elements are depicted in Fig. 2. Such a dialog can identify specific needs that are motivators, distinguish realistic from unrealistic needs or expectations that then might be eliminated through subsequent discussion, and establish equity by making sure that what is being asked from others is commensurate with what employees are willing to give, and vice versa.

A psychological contract is established when both parties agree to the cells shown in Fig. 2. A good psychological contract is one in which expectations are accepted by both parties, in which expectations are specific (unambiguous), and in which expectations and commitments are balanced. Examples of elements in psychological contracts often include keeping others informed, being willing to pitch in, listening to others, not bringing certain foods into work areas, receiving recognition, receiving training, and discussing assignments. Clearly, fulfillment of such expectations can increase motivation and productivity.

Psychological contracts should also contain the understanding that as conditions change, parties may need to renegotiate their understanding. Indeed, parties will want to discuss fulfillment of psychological contract, such as when one party is believed to be “not living up” to commitments. The moment of such insight is sometimes called a “pinch point.” Parties have a “right” to bring up the psychological contract when they feel that others are violating them. Often-times, a brief conversation is sufficient to address the sources of the pinch and to rectify it by providing a solution. It should be noted that in the absence of a psychological contract, it is often more difficult to bring up matters of discord.

	Expect to Get	Expect to Give
Worker		
Supervisor / Others		

Fig. 2 Psychological contract. (As adapted from Ref.^[8].)

Through psychological contracts, managers can gain additional insight into what individual workers expect and need. Psychological contracts complement the above sources of worker motivation by identifying additional sources of motivation. They do not substitute for the above sources of motivation.

REFERENCES

1. Maslow, A. *Motivation and Personality*; Harper, 1954.
2. McClelland, D. *The Achieving Society*; Van Nostrand, 1961.
3. Herzberg, F. *The Motivation to Work*; John Wiley, 1959.
4. Vroom, V. *Work and Motivation*; Wiley, 1964.
5. Locke, E.; Latham, G. *A Theory of Goal Setting and Task Performance*; Prentice Hall, 1990.
6. Skinner, B. *Beyond Freedom and Dignity*; Batham, 1971.
7. Rousseau, D. *Psychological Contracts in Organizations: Understanding Written and Unwritten Agreements*; Sage, 1995.
8. Osland, J.; Kolb, D.; Rubin, I. *Organizational Behavior*; Prentice Hall, 2000.

BIBLIOGRAPHY

1. Adams, J. Inequity in social exchanges. In *Advances in Experimental Social Psychology*; Berkowitz, L., Ed.; Academic Press, 1965.

Motives and Incentives: Rethinking Budget Theory in a New Institutional Context

William Earle Klay

*Askew School of Public Administration and Policy, Florida State University,
Tallahassee, Florida, U.S.A.*

INTRODUCTION

The institutionalized processes of budgeting confront administrators with pervasive disincentives. There is no method of productivity improvement known to business managers that is not also known to public administrators. Good public managers know ways to improve productivity, but they operate within institutions that discourage it. These disincentives exist primarily because past generations of legislators wanted to control executives more than promote productivity. Distrust of the executive is deeply rooted in our legislators' institutionalized behaviors, and these behaviors shape the incentive systems for public managers.

Failure to Spend an Appropriation

Agencies take great care to ensure that appropriations are fully spent before the end of a fiscal year. In classic studies, Anton^[1] and Wildavsky^[2] found that legislators were likely to cut agencies' budgets when they failed to spend all their earlier appropriations. Agency managers rationally respond to their incentive environment—they try to spend every possible penny. If money is likely to remain at the end of a fiscal year, managers make extra efforts to spend it. Even when agencies do not have to return unused funds to the general treasury, they fear that unobligated year-end balances will raise questions as to whether their programs are overfunded.^[3]

Disincentives and Fiscal Constraints

Even revenue scarcity does not end the disincentives for saving. When their governments face revenue shortages, savings that are created by conscientious managers are often shifted to support the programs of other, less cost-conscious, managers.^[4,5] Obfuscation and other defensive strategies seem logical when budgeters most reward those who stonewall well.

Cash Flow Planning

In private business, unforeseen cost increases or revenue shortfalls can cause a severe liquidity crisis and

failure, even when long-term prospects are good. Prudent stockholders, therefore, demand that their managers maintain cash reserves or an ability to borrow cash on short notice. Public managers face similar uncertainties. Service demands or costs (e.g., petroleum products) may exceed projected levels. Even though public managers also need contingency cash reserves, their requests for reserve funds, or efforts to save and retain money in reserve, are likely to be construed as “fat” to be cut. Requesting supplemental appropriations in the middle of a budget year can endanger an executive's career. Agency managers respond by trying to request enough in their annual budget to cover some contingencies. They then must spend all that is appropriated, even if higher costs or demands do not occur, simply to get funds to cover future contingencies.

“Pay for Performance”—An Empty Promise

Governments fail to provide incentives for productivity improvement.^[6] To remedy this, legislators occasionally enact “pay-for-performance” laws. Some laws offer higher pay only for top executives, others also promise salary enhancements to lower-level employees, in return for higher levels of performance.

Unfortunately, neither the U.S. Congress nor any state legislature has yet sustained a pay-for-performance system. Institutionalized legislative distrust, accompanied by desires to free money for other political priorities, have repeatedly derailed pay-for-performance initiatives. Contino and Lorusso,^[7] for example, described how public employees who cooperated to achieve greater productivity and savings were frustrated by the sweeping of the savings from their agency. Pay-for-performance laws usually require that auditors first verify agency cost savings. Auditors, though, reflect legislators' distrust of agencies and tend to become obstacles to the implementation of pay for performance.^[8]

In 1994, the Florida legislature enacted a law that promised financial awards to workers whose agencies produced verifiable cost savings. The Department of Revenue responded by effectively using quality management and reengineering techniques. When the

agency returned approximately \$9 million to the treasury, its director sought \$100 bonuses for half the agency's employees, totaling approximately \$250,000. The legislature failed to honor its promise, authorizing no bonuses in spite of strong revenue receipts. The state's largest business magazine accused the legislature of pulling "bait-and-switch" tactics on its employees.^[9] Florida is not unique; it is typical. Its institutionalized budgeting processes are quite similar to those of other states. Making pay-for-performance work, not just once but for years to come, requires an institutional capacity for sustained commitment that seems to be lacking among legislatures. No state legislature (nor the Congress) has yet demonstrated such a capacity.

Performance Data: Feared and Ignored

Believing that legislators will use performance data to punish but not reward agencies, managers become reticent to develop or provide it. A multistate comparison found that an "underlying current" of fear inhibits the development and use of performance data.^[10] Georgia failed to create a workable performance measurement system largely because agency managers feared that the central budget office would use the data against them.^[11]

When performance data are provided, elected officials (and their staffs) often pay little heed to it. A survey found that most states (42 of 46 reporting) collect information about agency performance but few (only 13) said that they actually used such data in making decisions.^[12] The situation appears no better in local governments. A survey of 510 local governments reported that 30% of those do service efforts and accomplishments (SEA) reporting. Actively promoted by the Government Accounting Standards Board, SEA reporting is an important development in revealing a government's productivity to multiple stakeholders. Only 5% of those doing SEA reporting (a meager 1.5% of the total respondents) indicated that the data played an important role in deciding those governments' budgets.^[13]

Politics, not Productivity, the Recipe for Success

In 1966, Davis, Dempster, and Wildavsky found that senior executives in the federal government "uniformly" believed that political skills were "more important in obtaining funds than demonstration of agency efficiency."^[14] The federal budget process clearly penalized the U.S. Securities and Exchange Commission by restricting its appropriations in the years following its best productivity.^[15] Connelly and Tompkins found that the performance data submitted by agencies in Missouri

had no statistical association with the governor's recommended spending changes. They concluded that performance data were used in a "dramaturgical" manner, its gathering giving "the impression of careful reasoning and active protection of the public treasury."^[16]

Analysts often distrust the data they get from agencies and, therefore, discount it. Even when analysts accept the validity of performance data, other political factors may push the data aside. There is no indication that this situation is changing. A survey of U.S. states revealed that 47 of 50 states had adopted budgeting systems that include the submission of performance data, but few states showed any link between performance information and actual appropriations.^[17] However, numerous studies have indicated that the agencies that are the most acquisitive in their political strategies tend to receive larger budget increases.^[18]

ECONOMICS-BASED THEORY NEGLECTS INSTITUTIONAL CONTEXT OF INCENTIVES

Sadly, a half century of successive economics-based budget reforms have been undermined by psychological and sociological naiveté. For the most part, these reforms have been rooted in the assumptions of scientific/economic rationality as described by philosopher Paul Diesing.^[19] Information is central to this form of rationality.

Budget Process Reform

Seeking to make budgeting more "rational," reformers have repeatedly sought to mandate enough information to allow decisions to be made according to calculations of marginal rates of return. A succession of process-oriented reforms like planning-programming-budgeting systems and zero-based budgeting sought to empower budgeters with the "perfect knowledge" of "economic man." Foolishly, those reformers failed to heed Herbert Simon's^[20] Nobel Prize-winning warning that economic theory must recognize that "perfect knowledge" is unattainable.

Even more modest reforms such as performance budgeting have never come to grips with the fact that they quickly overwhelm decision makers with details. No budget reform to date has squarely addressed the pressures and time constraints that confront busy legislators during sessions. Equally inexplicable is the fact that no budget reform has squarely addressed the problems of incentives. Little or no attention has been given to the disincentives that discourage managers from developing the information on which economics-oriented reforms so greatly depend. Disparaging politics, rather than seeking to understand it, generations of reformers in this tradition have

ignored the institutional complexities within which incentives structures emerge and operate.

Public Choice Theory

Incentives structures are central to public choice theories. This body of economic theory has not focused on budget reform but, in providing a theoretical basis for decentralization and privatization, it has enormous implications for all aspects of public financial administration, especially information systems, contracting, auditing, and budgeting. Public choice theorists declare that public managers are self-interested, materially motivated “agents” who possess various forms of advantages over the “principals,” taxpayers, and their elected representatives (although the latter are also agents).

Information is seen as the primary advantage that agents have over principals. Principals depend on agents to provide information about policy options and performance accomplishments. Agents, however, are presumed to manipulate information to their advantage. Consequently, much of public choice theory is concerned with reducing the adverse effects of these “information asymmetries.”^[21] Public choice theory concludes that principals must constrain agents via laws and penalties, hire auditors to monitor agents’ performance, and induce agents’ behavior via incentives. A complication arises, however, in that democratic governments almost never give their agents a significant share of the economic benefits that the agents might create for society (unlike private companies in which owners can share profits with hired agents in large amounts). Government incentives are modest, often non-monetary, and symbolic.^[22,23]

Recognizing that democratic governments cannot offer substantial economic inducements to public employees, public choice theorists prefer to abandon service delivery by public employees in favor of contracting to private for-profit companies. This, they believe, will introduce economies from competition and better harness incentives by allowing private companies to freely pay managers and workers according to their performance. If companies’ employees do not embody public service norms, however, their sole incentive is to maximize profit while delivering only the minimal quality of service needed to continue to compete for further contracts. Furthermore, a system based solely on the concept of bargaining and exchange “is poorly equipped to deal with exchanges between current and future citizens.”^[23]

Professional Norms Contradict Public Choice Assumptions

According to public choice theory, public employees are agent/bureaucrats who foster and enlarge information asymmetries in their own self-interest. Yet, for

at least a century, public sector professionals have fostered a series of reforms with the objective of reducing the information asymmetries! Early in the century, organizations such as the New York Bureau of Municipal Research (which started the first public administration training program in the United States) committed themselves to overcoming information asymmetries by providing financial and performance information to multiple groups of “principals,” including the residents of poor neighborhoods who were being poorly served. In recent years, the Government Accounting Standards Board, an authoritative body of public sector professionals, has been promoting Service Efforts and Accomplishments reporting. This reporting is based on the premise that conscientious professionals are obligated to measure and report both costs and accomplishments, through financial reports, to all stakeholders including elected officials, potential bond buyers, and taxpayers. Why do many public sector professionals repeatedly seek to remedy information asymmetries? There is no indication that public managers who do so will benefit economically. Their motives, therefore, cannot be explained by economics-based theory.

AN INSTITUTIONAL PERSPECTIVE ON MOTIVATION AND INCENTIVES

An “institutional perspective” acknowledges that motivation is multidimensional, shaped within complex social settings. The motives of public sector professionals who strive to remedy information asymmetries can only be understood in their institutional context. The legitimacy of all professions (to varying degrees) rests on their having a knowledge base rooted in science, so scientists’ belief in the free flow of information has become inherent to most, if not all, professions. The norm of stewardship—emphasizing that one’s personal worth is measured by how well others are served—is also taught in the socializing processes of public administrators (although perhaps not as well nor as explicitly as it should be). A commitment to the disclosure of information in the context of service to others, not economic self-interest, explains why public sector professionals have repeatedly sought to promote reforms that reduce information asymmetries.

Unfortunately, public administrators themselves often fail to take into account the institutional context of legislators. When the budget staff of the State of Texas promoted performance budgeting, they hoped it would make the legislative process more analytical. They were disappointed when legislators’ behaviors did not change appreciably.^[10] Even careerists, it seems, have repeatedly failed to understand the institutional nature of legislatures.

Acknowledge the Institutional Nature of Legislatures

Legislatures are venerable institutions with powerful socialization processes that date back to the Magna Carta and the American Revolution. Legislatures socialize their members into a different form of rationality—not inferior, but different—from that into which scientists (including economists) are socialized. “Economic/scientific” rationality seeks clarification and systematic comparison of objectives to optimize resource use. “Political rationality” is more concerned with the viability of a decision-making process. Building support, while averting future opposition, is paramount. Clarifying objectives is irrational when it increases conflict.^[19] More than anything else, the repeated failures of budget reforms must be attributed to reformers’ failure to acknowledge the institutional realities of legislatures. To remedy these failures, prescriptive theory should help legislators to behave in ways that encourage and reward desirable behaviors in agencies.

Needed: Institutionally Savvy Descriptive Theory

Budget reforms have been sociologically and psychologically naive. Unfortunately, the wonderful descriptive work of such scholars as Fenno^[24] and Wildavsky^[2] fell from favor; displaced by a generation of effort to statistically model such concepts as incrementalism. A more inclusive institutional framework is now needed to gain understanding of how budgeters and other financial decision makers actually behave. Miller’s^[25] proposal to use anthropologically oriented methods to observe how financial decision makers cope with ambiguity, frame their identity, and make sense of things is an important step toward building institutionally sensitive theory. So also is James Perry’s^[26] proposal of a theory of motivation that seeks to understand the behaviors of public officials within a sociohistorical context. The motivations of public officials can only be understood in the context of their socialization and membership in institutions.

Given legislatures’ reluctance (or inability) to sustain pay for performance, it seems wise to explore how public managers’ non-economic motivations might be tapped. For example, public managers desire greater discretion—that is, more leeway in using their judgment. An ingenious proposal to tap the motive force of discretion, while carefully holding managers accountable for their actions, was once proposed by three experienced federal administrators.^[27] It was not a proposal from self-serving/budget-maximizing bureaucrats. Discretion-oriented reforms would require public managers to become more openly

accountable and less able to hide behind information asymmetries.

Professional Norms

Where public managers are effectively socialized to internalize values of stewardship, service, and freedom of information, value seeking on behalf of the public is more likely to occur. Public managers who seek personal fulfillment through service to others will act differently than managers who only believe in economic motivation. Reformers who seek to promote better use of information and legislators who depend on information supplied by “bureaucrats” should encourage public service professionalism when it shows itself to be dedicated to the remedying of information asymmetries. Unfortunately, attention to professional norms has been woefully absent in both descriptive and prescriptive public financial management theory. The Progressive tradition, including its emphasis on the motive and constraining force of professionalism, should be relearned and emphasized.^[28]

Conclusion: Behavioral Interventions Essential to Reform

Any budget reform seeks to alter established behaviors. Many of these behaviors are deeply rooted legislative routines. Surprisingly, no reform to date has included carefully fashioned behavioral strategies for altering institutionalized behaviors. Not recognizing the rationale or staying power of institutionalized behavior patterns, reformers have unwittingly sewn the seeds of failure. Within social sciences, a large body of theory has been developed throughout the twentieth century regarding the altering of established behaviors. Organizational development, with its many strategies for intervention to achieve organizational change, is a well-established area of practice. Yet, with few exceptions,^[29,30] theorists have not recommended any methods of organizational development to enhance the prospects of budget reforms.

Institutional change requires sustained commitment from top leaders, but many legislatures, especially those with term limits, experience frequent leadership changes. Elected officials are unlikely to attend the lengthy meetings that planned organizational change requires. Consequently, senior staff members in budget offices and appropriations committees (who have often resisted budget reforms as much as their elected bosses) should be the leaders targeted in behavioral change interventions. Their roles as repositories of institutional memory, as well as trusted advisors, give senior staffers an institutional status from which to “champion” behavioral change on a sustained basis.

If top staff members can be assisted to learn how to modify routines to better use analysis, they are in a unique position to help their elected officials change the decision processes. One desirable change would be learning to substitute performance-oriented reviews for traditional inputs-oriented hearings. Another would be learning to make necessary political trade-offs without detailed line item formats. Still another would be to fashion ways to make legislators aware of the disincentives that they create for public managers. One thing, however, is certain. No change in the deeply rooted routines of legislatures (and elected chief executives) will be possible without a sophisticated understanding of the institutions within which budgeting is done. With institutional insight, though, it may someday be possible to institutionalize positive incentives for good public administration.

REFERENCES

1. Anton, T.J. *The Politics of State Expenditures in Illinois*; University of Illinois Press: Urbana, 1966.
2. Wildavsky, A. *The Politics of the Budgetary Process*; Little, Brown and Co.: Boston, 1964.
3. U.S. General Accounting Office. *Spending Patterns of the Departments and Agencies of the Federal Government*, GAO Report PAD-80-34; GAO: Washington, DC, 1979.
4. Levine, C.H. More on cutback management: hard questions for hard times. *Public Adm. Rev.* **1979**, 39 (2), 179–183.
5. Straussman, J.E. A typology of budget environments: notes on the prospects for reform. *Adm. Soc.* **1979**, 11 (2), 216–226.
6. U.S. General Accounting Office. *Increased Use of Productivity Management Can Help Government Control Costs*, GAO/AFMD-84-11; GAO: Washington, DC, 1983.
7. Contino, R.; Lorusso, R.M. The theory Z turnaround of a public agency. *Public Adm. Rev.* **1982**, 42 (1), 66–72.
8. Jarrett, J.E. *Improving Productivity through Monetary Incentives*; Council of State Governments: Lexington, KY, 1981.
9. Berry, J.F. Bait and switch tallahassee style. *Fla. Trend* **1997**, 40, 110 (February).
10. Broom, C.A.; McGuire, L.A. Performance based government models: building a track record. *Public Budg. Finance* **1995**, 15 (4), 3–17.
11. Lauth, T.P. Performance evaluation in the Georgia budgetary process. *Public Budg. Finance* **1985**, 5 (1), 67–82.
12. Jordan, M.M.; Hackbart, M.M. Performance budgeting and performance funding in the states: a status assessment. *Public Budg. Finance* **1999**, 19, 68–87 (Spring).
13. O'Toole, D.E.; Stipak, B. Service efforts and accomplishments reporting in local governments. *J. Gov. Financ. Manag.* **2001**, 50 (1), 36–40.
14. Davis, O.; Dempster, M.; Wildavsky, A. A theory of the budgetary process. *Am. Polit. Sci. Rev.* **1966**, 60 (3), 529–547.
15. Warren, R. Bureaucratic performance and budgetary reward. *Public Choice* **1975**, 24, 51–57 (Winter).
16. Connelly, M.; Thompkins, G.L. Does performance matter? a study of state budgeting. *Policy Stud. Rev.* **1989**, 8 (2), 288–299. quote p. 296.
17. Melkers, J.; Willoughby, K. Budgeters' views of state performance-budgeting systems: distinctions across branches. *Public Adm. Rev.* **2001**, 61 (1), 54–64.
18. Douglas, J.W. Budget reform theory: a guideline for successful budget reform. *Int. J. Public Adm.* **2000**, 23 (11), 1967–1996.
19. Dising, P. *Reason in Society*; University of Illinois Press: Urbana, 1962.
20. Simon, H.A. Rational decision making in business organizations. *Am. Econ. Rev.* **1979**, 69 (4), 493–513.
21. Dixit, A.K. *The Making of Economic Policy: A Transaction-Cost Politics Perspective*; The MIT Press: Cambridge, MA, 1996.
22. Wilson, J.Q. *Bureaucracy: What Government Agencies Do and Why They Do It*; Basic Books: New York, 1989.
23. March, J.G.; Olsen, J.P. *Democratic Governance*; The Free Press: New York, 1995.
24. Fenno, R. *Power of the Purse: Appropriations Politics in Congress*; Little, Brown and Co.: New York, 1966.
25. Miller, G.J. *Government Financial Management Theory*; Marcel Dekker, Inc.: New York, 1991.
26. Perry, J.L. Bringing society in: toward a theory of public-service motivation. *J. Public Adm. Res. Theory* **2000**, 10 (2), 289–327.
27. Boynton, R.P.; Medina, W.A.; Covello, L.S. How you always wanted to manage but were not allowed to try. *Bureaucrat* **1977**, 6, 131–151.
28. Miller, G. Above politics: credible commitment and efficiency in the design of public agencies. *J. Public Adm. Res. Theory* **2000**, 10 (2), 289–327.
29. Klay, W.E. An incentives and behavioral approach to productivity improvement. In *Dealing with the Problem of Budget Disincentives*; Sherwood, F.P., Ed.; Florida State University: Tallahassee, 1986; 12–41.
30. Forrester, J.P.; Adams, G.B. Budgetary reform through organizational learning: toward an organizational theory of budgeting. *Adm. Soc.* **1997**, 28 (4), 466–488.

Municipal Securities

Mark D. Robbins

University of Connecticut, Storrs, Connecticut, U.S.A.

Bill Simonsen

Center for Policy Analysis and Management, University of Connecticut, Storrs, Connecticut, U.S.A.

INTRODUCTION

Municipal securities are the subset of financial instruments that represent the debts of state and local governments, their agencies and authorities. The power to create municipal obligations lies in each state's constitution and statutes. Only the power to issue debt on behalf of the federal government is proscribed by the U.S. Constitution. Local governments maintain the ability to borrow and tax at the pleasure of their state governments, which retain all powers necessary to set the conditions and limits on these obligations. Markets for municipal securities exist primarily at the state and national levels. International (foreign denominated) municipal securities are issued by U.S. governments, but represent a very small share of the market.

Investors purchase bonds entitling them to future principal and interest (also called coupon) payments. Interest earnings from state and local government bonds are exempt from federal income taxes. States with income taxes typically exempt their own bonds from their own income tax. The effect of these exemptions is to decrease the interest rate necessary to offer investors in order to persuade them to purchase the bonds. In practice, this amounts to a subsidization of government borrowing provided from the federal (and state) governments. Governments use the subsidy to finance their capital projects, while maintaining low borrowing costs.

TYPES OF REPAYMENT SECURITY

Municipal bonds are the most common government security. These government obligations must be backed by a security pledge that demonstrates to investors the source of revenue that will be used to repay the debt. All else being equal, the price of bonds falls as this security pledge appears weaker and less certain. The strongest pledge is the full faith and credit pledge

that assures bond holders that the government will use its full taxing power to raise the revenues necessary to repay the debt. If that taxing power is unlimited, the security is referred to as a general obligation unlimited tax bond. Many governments operate under state or local tax limitations or other authority that limits their ability to raise taxes. General obligation limited tax bonds carry a full faith and credit pledge that is limited by such a revenue-raising restriction. Both forms of general obligations may face statutory or constitutional debt limitations or requirements for voter preapproval.

The classification of revenue bonds is used to designate the group of securities supported by a specific revenue stream rather than general tax support. These securities are paid for by government enterprise proceeds such as toll revenue or utility charges. Because they do not obligate taxpayers directly, they typically do not carry legal requirements for citizen approval.

Revenue bonds come in various forms, the most common of which are enterprise revenue bonds. These are single-party bonds in which the issuer dedicates its own specific enterprise revenues to repay the debt. An example would be bonds issued by a sewer authority that uses rate payments as the security to repay the bonds.

Lease revenue bonds are more complex structures involving more parties. Lease-backed bonds are repaid from the proceeds of operating leases—typically for buildings leased by government agencies and authorities. In these cases, an authority is created to build and operate facilities that are leased to a larger government or government agency. The government agrees to make lease payments, and the lease payments (or the promise of future lease payments) are used as security for the bonds. This security pledge is considered riskier than a dedicated revenue source because the payments are subject to the continued participation of the lessee government. Other restrictions may apply, including non-appropriability in which governments are forbidden from making multiple-year financial commitments

and contracts without statutory authority or voter permission.

Certificates of Participation offer investors an interest in the lease proceeds resulting from the payments to governments in exchange for the use of real property, equipment, and vehicles. These are riskier still because governments retain the right to halt appropriations for such lease payments at any time and because investors own a share of the lease proceeds as opposed to a specific repayment pledge. Certificates of Participation are often issued by government association lease pools that operate on a revolving fund basis to finance the capital purchases (typically vehicles and equipment) of participating jurisdictions. The pool is used to finance the purchases, and then lease proceeds are used to pay back the pool and cover the payments to investors. In this way, governments can access the capital market for the financing of relatively small capital items. This can be particularly attractive to smaller governments that would not otherwise access this market.

There are other types of tax-backed obligations, such as special assessment bonds, where taxes are pledged only for certain geographic areas. Tax increment bonds are obligations where only the increment in assessment for a certain geographic area after a specified point in time is subject to taxation for bond repayment.

State and local governments may also issue notes. Notes represent the short-term obligations of governments, typically (but not exclusively) maturing in 1 year or less. A common need of governments is to obtain interim financing for a project for which other revenue sources are anticipated. Tax anticipation notes and revenue anticipation notes are issued and secured by future tax or other revenue receipts, whereas bond anticipation notes are issued pending the conclusion of a long-term bond sale. Hybrids, such as tax and revenue anticipation notes, are also issued. Because of their short term, notes are considered riskier obligations than longer-term bonds and because investors are subject to greater interest rate risk and liquidity concerns.

Tax exempt commercial paper (TECP) is an alternative to notes and typically carries shorter maturities. Tax exempt commercial paper is typically purchased by commercial banks and institutions. The proceeds might be used by governments seeking to finance the construction phase of a project without incurring long-term debt until its completion, or seeking to delay the sale of longer-term obligations in anticipation of lower interest rates. Like notes, TECP is governed by legal authority separate from that of municipal bonds and results in higher relative borrowing costs on issuers.

Moral obligations state the good intentions of a government—without promising a specific revenue

source to repay a debt. This pledge indicates the government's recognition that they are morally, if not legally, obliged to cover the default of subsidiary agencies and authorities, without issuing a full faith and credit pledge and its accompanying voter approval requirement. The phrase was first added to bond covenants in the 1960s, when New York State bond counsel John Mitchell sought to make that state's issues more attractive to investors. This evasion of voter approval was apparently intentional by design. When asked decades later whether he created moral obligation bonds to bypass the referendum and initiative petition processes, Mitchell reportedly replied, "That's exactly the purpose of them."^[1] Rating agencies typically overlook the moral obligation pledge, basing their judgment on the creditworthiness of the specific issuer and the strength of the project revenue stream.

Additional forms of municipal securities include repurchase agreements and reverse repurchase agreements in which governments purchase or sell securities under the condition that the securities will be purchased or repurchased at a later time. These are primarily used as sources of short-term borrowing (or investment) by local governments.

ISSUANCE PROCESS

In addition to the government that creates and sells securities, there are many additional participants in the issuance process (Fig. 1). The participants include the issuing government, the underwriter or banking firm that purchases and remarkets the bonds, and the investors who buy the bonds from underwriters. Also involved are financial advisors, bond counsel, and members of the ratings and insurance firms that evaluate the issue prior to its sale.

Most governments are small and issue few if any municipal bonds in a given year. As a result, the internal

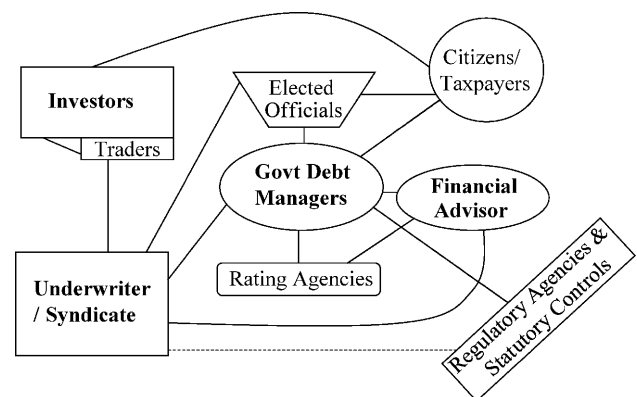


Fig. 1 Participants in the municipal securities issuance process.

capacity of a government to successfully oversee all aspects of these transactions is limited. In part to compensate for this, governments employ a series of expert consultants to guide them through the process. The financial advisor is one of these consultants, and may be hired to help with all or some subset of the steps of issuance from structuring the bond issue to overseeing the sale. There are two types of financial advisors—*independent financial advisors* and *underwriting financial advisors*. Independent financial advisors have no fiduciary interest in bond sales and work exclusively in the interest of issuing governments, helping to structure, size, and time the bond sale so that the government gets the best (highest) possible price for its bonds. Underwriting financial advisors work purchasing and reselling municipal bonds, in addition to providing governments with expertise about the bond sales process. In some states, underwriting financial advisors are permitted to purchase the bonds of the governments for whom they provide advice.

Each government security issuer obtains a legal opinion from bond counsel prior to the sale of their securities. These attorneys are hired to certify that the sale is conducted in accordance with all applicable statutory and constitutional authority, and that the interest payments received by the bond holders will be exempt from (state and) federal taxes. Bond counsel may also be involved in creating the official statement—offering the securities for sale, establishing the mechanisms for holding the proceeds of the bond sale, and determining the maximum allowable earnings jurisdictions may legally accrue on these proceeds pending their use.

Once an issue has been approved for sale, the issuer will select an insurer (if any) and a ratings firm (if any). Both will examine the condition of the jurisdiction and make a determination of creditworthiness. This determination will affect the rating assigned by the firm and the premium charged by the insurer. Once in place, the ratings and insurance signal investors about the riskiness of the issue.

Issuers must choose a method to sell their bonds. They may select a competitive or negotiated sale, a private placement, or an online auction. Private placements occur when a government issuing a small amount of bonds sells them directly to one bond holder, typically a bank or financial institution. Larger sales typically require the use of an underwriter to act as an intermediary, buying the bonds from the issuer and reselling them to investors. In the case of a negotiated sale, an issuing government selects an underwriting firm, sometimes through the use of a request for proposal. The underwriter premarkets the bonds by lining up investors in advance and then negotiates with the government about the interest rate that it will have to pay. When selling bonds

competitively, issuing governments solicit sealed bids from as many underwriters as possible. The bids show the interest rate that the underwriting firm is willing to accept to purchase and remarket the bonds. The underwriter presenting the lowest interest cost wins the right to purchase the bonds. In both competitive and negotiated bond sales, a single underwriter (or a syndicate composed of several underwriters) purchase(s) the entire bond issue, reselling the individual bonds to investors later.

In November 1997, Pittsburgh, Pennsylvania, initiated the first online auction of (new issue) municipal bonds. This system is a form of competitive sale that allows issuing governments to coordinate a bond sale by collecting bids over the Internet. One of the innovations of this method of sale is that investors may bid on as many or as few of the bonds as they like, thus opening the bidding process to parties (e.g., individual investors) that seek to purchase less than the entire bond issue and potentially bypassing underwriters altogether. Since that first sale through July 2001, MuniAuction, the company pioneering the technology behind this innovation, has concluded more than 2300 online sales. This method continues to grow in prevalence.

The responsibilities of the government do not end after the sale is consummated. Issuing jurisdictions are required to establish and maintain a regular payment mechanism to assure that bond holders receive interest and principal when they become due. This responsibility is typically assigned to a registrar payee—a firm specifically in the business of monitoring and completing such transactions. Governments with outstanding municipal bonds are required to make regular disclosures to bond holders and the public regarding events that may materially affect their financial condition and ability to repay the obligations. This responsibility is typically satisfied by the submission of annual financial statements to a nationally recognized municipal securities information repository, which disseminates the information. It is the responsibility of the issuer to disclose in the same fashion any interim event that may materially affect repayment ability.

RATINGS AND CREDIT ENHANCEMENT

The default rates of municipal securities is generally considered to be low compared with other sectors of the financial markets. The truth of this maxim depends on the sector of the tax exempt market that is being observed. Fitch IBCA reports a cumulative default rate of 0.26% for general-purpose tax-supported bonds for the period from 1979 to 1986. This compares with a corporate bond default rate averaging 2%.^[2] Industrial

development bonds (used to extend tax-free borrowing to private sector corporations in an effort to promote economic development) carried a default rate of 14.89%—a default rate exceeded only by corporate securities rated BB or worse. The Tax Reform Act of 1986 limited the circumstances under which governments could issue industrial development bonds, and the volume of such financings (and associated defaults) has subsequently decreased. Total defaults have consequently decreased as a percentage of total issuance across all municipal bond sectors from an average of 1.49% (1979–1986) to 0.87% (1979–1994).

Governments selling bonds and other securities pay rating agencies to evaluate their creditworthiness and publish a rating that is made available to investors selecting the securities that they will purchase. In some cases, governments will purchase more than one rating to provide further certification of the value of their bonds. Jurisdictions that want to gauge their ratings in advance may purchase ratings not subsequently released to the public. These are referred to as “preliminary ratings.”

The municipal bond rating industry is robust, with three prominent firms rating most of the issuance volume. This presents a conundrum. Given the low overall default rates of municipal securities and the failure of rating agencies to anticipate some notorious defaults (e.g., Washington Public Power Supply System, Orange County, California), the value of a bond rating system to help investors to predict default risk might appear nominal in this market. So why do governments continue to pay to be rated? The power of the rating mainly lies in the collective psychology of investors, not only in some intrinsic value of the ratings themselves or their ability to signal default. Ratings are used as a quality heuristic by investors. As long as some investors continue to rely on ratings, they remain significant determinants of bond prices. Issuers care about ratings because investors pay attention to them, and investor interest determines the amount of money municipalities can raise when selling their bonds.

Bond ratings also have an unintended, but positive, influence on government management. Bond ratings are viewed broadly as the independent evaluation of the quality of a government's management practices, as well as its economic and financial condition. The Aaa rating (the highest obtainable) is coveted by public managers as a sign of their superior performance. A private system of regular independent evaluation of governments has thus evolved that managers and the public pay attention to, and that may help point governments toward improved performance.

Many municipalities purchase insurance for their bonds. Bond insurance pays interest and premium payments in the case of a bond default. The premium for

the insurance is paid in full at the time of the closing of the bond sale. Because many municipal bond issues are refunded (refinanced) prior to their final maturity date, it is common for issuers to pay for more insurance than they actually use. Nevertheless, issuers purchase bond insurance because insuring an issue guarantees a Aaa rating and is expected to produce higher bond prices. Issuers already obtaining Aaa ratings without the use of insurance sometimes purchase additional ratings in an attempt to certify to investors the unanimity of good opinion about the quality of the repayment security.

UNDERSTANDING BORROWING COSTS

Citizens, government analysts, and managers are concerned about the costs that governments face when borrowing money. Because of the number of parties and provisions involved when issuing municipal securities, these costs are not always transparent. The relationship between various security features and the costs that they impose can be complex and unanticipated. Much research has examined the determinants of interest costs for municipal bonds, and this work has developed some consistent lessons about these costs and their determinants.

Several characteristics tend to increase the cost of borrowing, all else equal. These factors include longer maturity lengths, borrowing very large or very small amounts of money, the presence of a call provision (allowing issuers to pay off bond holders prior to maturity), poorer ratings or the absence of ratings, the complexity of the financing and security pledge, issuance during volatile or high interest rate markets, and sale through negotiation. Factors expected to lower interest costs for borrowers include shorter maturity lengths, the presence of a put provision (allowing investors to demand redemption prior to maturity), the use of credit enhancement such as bond insurance, issuing general obligations as opposed to revenue bonds, and selling these bonds through a competitive (multiple bidder) sales process.

CONCLUSION

Governments borrow money for various reasons, including the need to maintain liquidity, to obtain interim financing, and to support the construction of major capital projects. The selection of a particular type of municipal security needs to match the needs of the issuer at a particular moment in time, but each type of security possesses certain attributes that make

it more or less attractive to governments and investors. These attributes, and the mechanics of the bond issues themselves, combine to present governments with a series of trade-offs when selecting a financing mechanism. Government managers need to understand these trade-offs when pursuing project financing.

REFERENCES

1. Marlin, G.; Mysak, J. *The Guide to Municipal Bonds: The History, the Industry, the Mechanics*; American Banker/Bond Buyer: New York, 1991; 24 pp.
2. Bond Market Association. *Municipal Bond Defaults*; 2001; Online: <http://www.bondmarket.com/Research/defaults.shtml>, download date: 08/02/01.
1. Fitch IBCA. Municipal Default Risk. In *Public Finance Special Report*; 1999 [Mimeo].
2. Grant Street. *MuniAuction Over \$1 Trillion in Par Amount*; 2001; Online: <http://www.grantstreet.com/perl/bond.pl>. Download date: 08/04/01.
3. Public Securities Association. *Fundamentals of Municipal Bonds*, 4th Ed.; Public Securities Association: New York, 1990.

BIBLIOGRAPHY

1. Fitch IBCA. Municipal Default Risk. In *Public Finance Special Report*; 1999 [Mimeo].
2. Grant Street. *MuniAuction Over \$1 Trillion in Par Amount*; 2001; Online: <http://www.grantstreet.com/perl/bond.pl>. Download date: 08/04/01.
3. Public Securities Association. *Fundamentals of Municipal Bonds*, 4th Ed.; Public Securities Association: New York, 1990.

Narrative Inquiry

Sonia M. Ospina
Jennifer Dodge

*Robert F. Wagner Graduate School of Public Service, New York University,
New York, New York, U.S.A.*

INTRODUCTION

Narrative inquiry is a unique approach to qualitative, interpretive research that draws on diverse interpretive theoretical traditions. It directs attention to narratives as a way to study an aspect of society, and is not only about studying written or visual texts, but finding meaning in stories people use, tell, and even live. Narrative inquiry can be located within a constructionist epistemology and an interpretive theoretical perspective, and it has specific features that make it a unique interpretive approach. What distinguishes it from other methods is the focus on narratives and stories as told, implicitly or explicitly, by individuals or groups of people. For public administration research, this often translates into an emphasis on the meaning of administrative experience or practice communicated through stories.

Within this broad orientation, three distinct approaches to narrative inquiry can be identified: narrative as language, narrative as knowledge, and narrative as metaphor.^[1] While these approaches share basic assumptions, they draw on distinct theoretical perspectives that emphasize one of three core assumptions about the nature of narrative. Together they provide a rich array of methodologies to understand administrative experience and practice.

FOUNDATIONS OF NARRATIVE

Narrative inquiry has its roots in “constructionist epistemology,” a theory of knowledge that suggests that we know the world not by objectively observing an external reality, but by constructing how we understand it with others.^[2,3] Proponents pose that all meaningful reality is socially constructed. By implication, narrative inquirers do not claim to document “reality,” but to capture individual interpretations of reality, as well as shared social constructions among a given community. Hence, narrative inquirers usually adopt an interpretivist, rather than positivist, theoretical perspective.

Interpretive research aims to understand the dynamics of social phenomena rather than explaining it causally or predicting it. In contrast to the positivist

goal of identifying and testing causal relationships to predict behavior, interpretive research illuminates meaning to build, test, and elaborate theory that emphasizes the intentions, beliefs, and values of social actors. Those following this tradition assume that meaning can be better unveiled and understood through experience and practice, always mediated through language and narrative.

Its focus on narratives and stories as told, implicitly or explicitly, by individuals or groups of people, distinguishes narrative inquiry from other interpretive approaches. While much discourse and content analysis are about studying written or visual texts, narrative analysis is about finding meaning in the stories people use, tell, and even live. Three basic assumptions are at the core of narrative inquiry: First, narratives convey meanings—intentions, beliefs, values, and emotions—that reflect situated social reality, rather than an “objective reality” *per se*.^[4] Second, narratives carry practical knowing that individuals gain through experience^[5] so that creating and passing on knowledge requires crafting and telling stories from experience.^[6] Third, narratives are both *shaped by* individuals for their own purposes and they become forces that shape human beings, giving meaning to the social worlds they inhabit.^[3] This means that narratives are constitutive.

MULTIPLE FACES OF NARRATIVE INQUIRY

The relative weights that researchers give to each of the three assumptions of narrative inquiry translate narrative theory into very different research practices, with implications for data collection, analysis, interpretation, and data presentation.^[1] Each of these three approaches is built on a primary narrative assumption with implications for its view of how people understand and experience the world. Similarly, each serves a different research purpose, and therefore privileges different methods (see [Table 1](#)). These three approaches represent ideal types of narrative inquiry, each highlighting one assumption over the others, while still drawing on all three.

Narrative as language gives priority to the assumption that narrative is a medium of expression, thus

Table 1 Three approaches to narrative inquiry

Narrative as ...	Primary assumption	Implications of primary assumption for how people experience the world	Purpose of the inquiry	Methods
Language	Narrative is a medium of expression	People <i>create and use</i> stories to communicate meaning	To understand some phenomenon from the perspective of the person/people experiencing it	Systematic Comparative Looks for common patterns
Knowledge	Narrative is a way of knowing	People <i>think and know</i> through stories Narratives contain knowledge	To surface tacit knowledge or share theories in use To enhance practice, or draw lessons from practice	Comprehensive Looks for and elaborates exemplars
Metaphor	Narrative is a symbol of deep structures of meaning	People <i>are born into or enter</i> (i.e., are socialized into) existing stories or storied institutions	To unveil implicit shared meanings To offer alternative interpretations of accepted views	Deconstructive Looks for taken-for-granted structures of meaning

highlighting the social nature of language and its function for making interaction possible. Drawing on phenomenological perspectives,^[2] this view emphasizes that people create narratives to convey meaning about their experience. Narratives are thus useful for studying how people understand a topic, for example, organizational change.^[7] Narratives help researchers understand actors' experience or at least the version of experience that the narrators choose to convey. This way, narratives illuminate something about the constructed reality they reflect and about the experience and intentions of those engaged in it.

An important implication of this assumption is that people are *purposeful* social agents who *create and use* stories to communicate meaning to themselves and others.^[4,7] The researcher attempts to understand experience from the perspective of those enacting it, drawing on meanings and arguments embedded in the actors' words. Since language reflects individuals' understanding of reality, the researcher gets a glimpse of the social world, as filtered by the meaning-making process of those who experience it.

Scholars interview informants to collect stories that serve as windows to their world. These are tape recorded and transcribed to fully and accurately capture what was said, including non-verbal cues like pauses. Analysis involves surfacing participants' sense making about the world. Systematic comparative analytical techniques distill form, content, and structure of stories, reflecting a qualitative orientation that looks for patterns of experience across stories and sites. In public administration, analysis often focuses on the narratives of multiple people with common experiences. Scholars assume that we learn something new by shifting focus from structural or behavioral aspects of public institutions to the ways in which actors' make sense of their experience in governance.^[8]

Feldman and Skoldberg represent this approach.^[7] By generating stories from open-ended interviews in two U.S. city administrations, they provide an account of how administrators made sense of change processes, thus illuminating political decision-making and organizational change. They used a special type of narrative analysis, called rhetoric, which involves dissecting the form and structure of arguments made in stories to uncover the meanings behind organizational realities.

The second approach, *narrative as knowledge* also privileges individuals' interpretations of their experiences, but focuses on their immediate practical use. It prioritizes the second assumption, that narrative is *a way of knowing*: people use stories to draw knowledge explicitly from their lived experience. This view emphasizes the potential of storytelling to generate understanding and practical learning.^[9] Bruner, for example, argues that we construct the world through narrative, and describes storytelling as one of two complementary ways to order experience.^[6] Logical proof, tight analysis, argumentation, and hypothesis-driven discovery contrasts with the narrative mode which focuses on good stories that convince through lifelikeness. Lifelikeness lends practical insights by allowing us to compare others' experience to our dilemmas, thus drawing useful lessons.^[10] Because people think and know through narratives, narrative inquiry offers a unique tool to learn about practical knowledge.

This approach has several common features, though connections to various theoretical heritages produce different applications. First, narrative as knowledge draws attention to the learning embedded in stories about practice. Second, it assumes that these forms of knowing produce insights from which to draw generalizations to enhance future practice,^[11] and can be accessed through practitioner reflection.^[12] The immediate goal is to illuminate tacit knowledge or

“theories-in-use” implied in stories and embedded in the accounts of practice,^[12] to ultimately draw lessons that enhance practice.

In this approach, rather than doing systematic comparative analysis of bounded stories, the analyst creates a comprehensive story of experience, highlighting extraordinary successes or failures, or simply insights about practice. Explicit tools help gain access to the insights embedded in day-to-day practices, to make them available for others.

In one way of implementing this approach to narrative, reflective practitioners tell their own story^[11,12] or analysts draw from written or verbal accounts of those reflections as a way to build theory from practice. Another strand of this approach relies on action-oriented forms of inquiry such as action research, action learning, action science, and appreciative inquiry.^[10] Through conversation, scholars help practitioners’ understand and act on their own environment. For some, the inquiry focuses on problem identification and action planning, giving priority to practical reasoning over instrumental reasoning, and seeking to integrate theory and practice.^[9] Another variation combines storytelling with observations and documentation, to produce an overarching account with plot, characters, and chronology. The new narrative is the product of the analyst’s sense making around stories told and heard over time, distinct from the typical case study, a historical narrative, or a traditional ethnographic account. An example is Behn’s work on the implementation of large public management reform efforts, where he draws lessons for leadership from the way various actors tell about their perspective of the experience, combined with his own interpretation of additional documents and narratives.^[13]

Both narrative as language and narrative as knowledge privilege informants’ interpretations of immediate experience. The third approach attends to the influence of broader institutional contexts on people’s ability to fully understand the complexities that inform their experience.

Narrative as metaphor emphasizes the assumption that narratives are constitutive. In this view, notions of self, existence, identity, and the like are “named” through language, and begin to feel “real” in their consequences. People take these notions for granted as objective and pre-existing realities. Over time, this process becomes automatic and usually unconscious.^[2] Moreover, subjective interpretations and associated actions are influenced by historical and institutional contexts that already offer manufactured meanings, manifested in discourses (like “the American dream”), in institutions (like the health system), or even in practices (like performance evaluation). People contribute to the (re) creation of these external “institutions of meaning” that then stabilize an “order of things.”^[14]

These institutions of meaning can be read as “expressive statements” or “texts” with a particular grammar, syntax, and structure that send normative messages and regulate social life, similar to what postmodernists call the “grand narratives” of a society.^[15] Drawing from hermeneutics, critical theory, and poststructuralism, researchers in this approach view narratives that apparently reflect straightforward social practices as metaphors capturing deeper meanings about the social order. For example, the way office space is assigned in an organization is a metaphor for structures of power. Emphasizing narrative as metaphor helps the researcher attain deep understanding of practices, organizations, institutions, or systems.

An important implication of these assumptions is that individuals are born into or enter (through socialization) existing stories or storied institutions. As human creations, these institutions gradually take on a life of their own and affect people without their awareness. Hence, the primary goals of this approach are to unveil the powerful but often invisible meanings embedded in institutional life and to identify suppressed or competing narratives thus proposing alternative interpretations.

In contrast to the microperspective of the first two approaches to narrative, this approach adds a macroperspective that links the immediate experience of social actors to broader institutions of meaning. The analyst chooses methods to critique and deconstruct texts, rather than taking them at face value, and tries to do an analysis that considers, in a recurrent way, meanings that are abstract and concrete, general and particular, evident and hidden. Analysts search for and re-examine claims to authority embedded in the text, and treat their own point of view as part of the reality under scrutiny.^[15]

Scholars of public affairs often use this approach to explore the nature of politics and policy making.^[16] Yanow, for example, studied how policy meanings were communicated to various audiences as the legislation to create the Israel Corporation of Community Centers (ICCC) was designed and implemented, and during its early years of existence.^[15] She argues that the difference between the intentions behind legislation and the final implementation cannot be understood simply as inadequate legislation or poor policy implementation but as part of a larger story about national identity.

CONCLUSION

Narrative inquiry is a tradition that is beginning to shift from the margins to the mainstream in public administration research. Unpacking the three approaches to

narrative inquiry clarifies the logic of this mode of interpretive research and illustrates its potential for illuminating the social dynamics of public institutions. They offer a rich set of resources from which to draw in designing interpretive inquiry. Clarifying their theoretical assumptions and methodologies can help those interested in designing narrative studies to advance interpretive public administration research. This type of inquiry complements traditional positivist research, contributes to make the field more pluralistic, and deepens our understanding of the processes, structures, and dynamics that constitute the theory and practice of public administration.

ACKNOWLEDGMENTS

We would to acknowledge the contribution to these ideas of our colleague, Erica Gabrielle Foldy, who has been a close partner with us in thinking about and doing narrative research.

ARTICLES OF FURTHER INTEREST

Postpositivist Perspectives in Policy Analysis, p. 1530.
Narrative Policy Analysis, p. 1289.

REFERENCES

1. Dodge, J.; Ospina, S.; Foldy, E.G. Integrating rigor and relevance in public administration scholarship: the contribution of narrative inquiry. *Public Admin. Rev.* **2005**, *65* (3), 286–300.
2. Berger, P.L.; Luckmann, T. *The Social Construction of Reality*; Random House: New York, 1966.
3. Gergen, K.J. The social constructionist movement in modern psychology. *Am. Psychologist* **1985**, *40* (3), 266–275.
4. Riessman, C.K. Narrative analysis. In *The Qualitative Researcher's Companion*; Huberman, M.A., Miles, M.B., Eds.; Sage: Thousand Oaks, CA, 2002; 217–270.
5. Clandinin, D.J.; Connelly, F.M. *Narrative Inquiry: Experience and Story in Qualitative Research*; Jossey-Bass: San Francisco, 2000.
6. Bruner, J.S. *Actual Minds, Possible Worlds*; Harvard University Press: Cambridge, MA, 1986.
7. Feldman, M.S.; Skoldberg, K. Stories and the rhetoric of contrariety: subtexts of organizing change. *Culture Organization* **2002**, *8* (4), 275–292.
8. Maynard-Moody, S.; Musheno, M. *Cops, Teachers, Counselors: Stories From the Front Lines of Public Service*; The University of Michigan Press: Ann Arbor, MI, 2003.
9. White, J.D. *Taking Language Seriously: The Narrative Foundations of Public Administration Research*; Georgetown University Press: Washington, DC, 1999.
10. Forester, J. *The Deliberative Practitioner: Encouraging Participatory Planning Processes*; The MIT Press: Cambridge, MA, 1999.
11. Schall, E. Notes from a reflective practitioner of innovation. In *Innovation in American Government: Challenges, Opportunities and Dilemmas*; Altschuler, A.A., Behn, R.D., Eds.; Brookings Institution Press: Washington, DC, 1997; 360–377.
12. Schon, D.A. *The Reflective Turn: Case Studies in and on Educational Practice*; Teacher's College Press: New York and London, 1991.
13. Behn, R.D. *Leadership Counts: Lessons for Public Managers From the Massachusetts Welfare, Training, and Employment Program*; Harvard University Press: Cambridge, MA, 1991.
14. Robichaud, D. Narrative institutions we organize by. In *Narratives We Organize By*; Czarniawska, B., Gagliardi, P., Eds.; John Benjamins Publishing Company: Amsterdam and Philadelphia, 2003; 37–53.
15. Yanow, D. *How Does a Policy Mean? Interpreting Policy and Organizational Actions*; Georgetown University Press: Washington, DC, 1996.
16. Schram, S.F.; Neisser, P.T. *Tales of the State: Narrative in Contemporary U.S. Politics and Public Policy*; Rowman & Littlefield Publishers, Inc.: Lanham, MA, 1997.

Narrative Policy Analysis

Emery Roe

Public Policy, Mills College, Oakland, California, U.S.A.

INTRODUCTION

Policy narratives are the arguments, scenarios, and other accounts about how policy and management have, can, and should proceed in ways that enable decision makers to take action, be they policy makers, their analysts, or others. More formally, policy narratives are scenarios (stories and arguments) that stabilize decision making for issues of high complexity, uncertainty, incompleteness, and conflict. Each narrative has a beginning, middle, and end (or premises and conclusions, if offered as an argument), and revolves around a sequence of events or positions in which something is said to happen or from which something is said to follow.

There are, of course, all manner of narratives in public policy. Some are really non-stories, in that they do not have their own beginnings, middles, or ends. Standard examples of non-stories are: (i) policy critiques that rebut arguments without telling their own stories and (ii) circular arguments whose beginnings, middles, and ends are by definition difficult to parse. Other narratives come to us as condensed stories. There are adages: The best is the enemy of the good; the opposite of good is good intentions. There are analogies: The abortion of a human zygote no more destroys life than a burning Home Depot destroys houses.

OVERVIEW

Most narratives in public policy, however, do have beginnings, middles, and ends. All such narratives, moreover, come with counternarratives; in semiotic terms, a story is marked by what it is not. Where critiques and circular arguments on their own serve only to undermine a policy narrative rather than replace it, a counternarrative offers up an alternative story or argument with its own beginning, middle, and end. You say *x*; I say counter-*x*; and unsurprisingly, both assertions, and many in between, subsist and thrive alongside each other for issues of uncertain, complex, incomplete, and disputed merits.

Policy narratives, though different in what is taken to be their merits, can share very similar structures. What crisis narrative do I have in mind? The earth

releases gases into the atmosphere that are then triggered by sunlight and lead to droughts, tidal waves, storms, and other natural disasters. No, not global warming, but rather Aristotle's theory of comets. Furthermore, crisis narratives are frequently linked. Africa for many is the seriatim crises of epidemics, droughts and refugees, underdevelopment, desertification and deforestation, failed states, declining incomes, and always-on famine. Or closer to home: Take a city to which you would not move. What's the problem there? Unaffordable housing, crime, inadequate public education, too much traffic, the inner city, too much growth, or too little of the right kind: all these or more?

While narratives are ubiquitous and importantly so in politics, policy, and management, you do not find "policy narrative" as an index entry in our leading textbooks on how to analyze, manage, and make decisions in politics and policy. The irony is that the policy analysis textbooks are themselves the stories we tell our people. We may disagree over whether or not "Rand Corporation begat . . .," "Planning, Programming and Budgeting System (PPBS) begat . . .," or "economics (no, public administration!) begat policy analysis, which begat implementation analysis, which begat public management, which begat the new governance theory . . .," but these creation myths are constantly retold in our graduate schools of public policy and management.

Myths? The American policy analyst and manager will, should he or she have a graduate career, come across the story that policy analysis rose from the ashes of failed public administration programs. Our European counterparts, in contrast, know—repeat—know that public administration begat policy analysis and that public administration has always been the welcoming parent to its prodigal children, be they now calling themselves implementation, public management or the new governance. Most graduate schools do not treat such foundation stories as myths—and understandably so. If they were "simply" the stories we told our people, then what sense would there be to our claims that these, rather than those, are the better methods for policy analysis and public management? If our house is built on sand, what purchase is there in saying one remodeling lasts longer than another? Without the link between PPBS and what

we have learned since, how would we justify an eightfold, sixfold, or fourfold path to policy analysis? After all, who can doubt that analysis begins in problem definition and ends . . . and ends in what, precisely?

For its part, narrative policy analysis begins and ends in stories we tell to make sense of and act upon issues that are often uncertain, complex, incomplete, and conflicted at the same time. The issues are uncertain when causal processes are unclear or not easily comprehended. They are complex when the issues are numerous, varied, and interdependent than before. They are incomplete when efforts to address them are interrupted and left unfinished. They are conflicted when different people can and do take different positions on them precisely because of the uncertainty, complexity, and unfinished business. Policy narratives are one major way to underwrite and stabilize our decision making in the face of any such amalgam of contingencies.

As many of the issues, which policy analysts and public managers are asked to address, are ill-structured problems, the stepwise policy process they conventionally summon to guide and justify their course(s) of analysis and action must itself be narrativized. It should also go without saying that the terms, “uncertainty, complexity, incompleteness, and conflict,” are themselves cast as narratives to differing degrees. Take something as central to social science as “interdependence.” The concept is notorious for its lack of agreed-upon empirical measures. Undertake a thought experiment then. What if instead of interdependence we had all along been talking about the Taoist notion of resonance? Instead of things being interdependent, they resonate with each other. Instead of being interrelated, things move and act in response to and in concert with each other. Even the slightest divergence in a foundation narrative can ramify through the development and application of that narrative thereafter.

DEFINITION

Exactly how does its own narrative go when it comes to narrative policy analysis? The terminology has varied and the process is by no means as mechanical as the following passage reads, but four steps are involved in a narrative policy analysis. First, the analyst identifies the dominant policy narrative(s) for the issue in question. Second, the analyst identifies the narratives that run counter to the dominant account. Third, the analyst compares the dominant policy narratives with the counternarratives to determine if there is metanarrative that can reconcile or accommodate these multiple, conflicting perspectives, without in the process slighting any of the oppositions. The last step is to determine if and how a metanarrative recasts the issue

in such a way that it is more tractable to conventional policy analytical tools of microeconomics, statistics, organization theory, and law and public management practice than were the original narratives, which the metanarrative embraces. If we are told that the policy narrative is one where $a \rightarrow b \rightarrow c$, can you imagine a counternarrative, where $(\text{not-}a) \rightarrow (\text{not-}b) \rightarrow c$ or where $a \rightarrow b \rightarrow (\text{not-}c)$? The conditions under which both the narrative and the counternarrative hold at the same time constitute the metanarrative.

As such, narrative policy analysis begins each analysis not with the conventional question in policy analysis, “What’s the problem?” but rather with “What’s missing?” What am I, an analyst or manager, not seeing or understanding? What is there in this ill-structured issue that could explain how the opposing positions could all be the case at the same time? What are the initial conditions, which, if I knew them better, connect the contrary elements in the narratives so that they are no longer mutually exclusive?

Note what has not been said. It is not claimed that all policy problems are wicked and all management ill-structured. It is not claimed that there is no such thing as power and that everything is relative. It is not claimed that there will always be a metanarrative, that it will always be just one metanarrative, and that it will always recast the problem more tractably. It is not claimed that narrative policy analysis is the only or the best way to treat amalgam issues of high contingency. Such policy issues require triangulation from very different directions, and narrative policy analysis is only one of the tacks that can be taken, especially when policy narratives and counternarratives fuel the issue.

EXAMPLE

Readers scarcely need to be told that the policy narrative in favor of the economic globalization of trade is a dominant scenario in numerous decision-making arenas, as witnessed by the many calls for trade liberalization, deregulation, and privatization to expand economic growth. As this is a highly contested area, there are counternarratives to the globalization narrative. Focus here on the opposing arguments put forth by environmental proponents of sustainable development. This opposition can be roughly divided into two camps: a “green” counternarrative and an “ecological” one.

The green counternarrative assumes that we have already witnessed sufficient harm to the environment owing to globalization and thus demands taking action now to restrain further globalizing forces. This counternarrative is confident in its knowledge about the causes of environmental degradation as they relate to

globalization and certain in its opposition to globalization. In contrast, the ecological counternarrative starts with the potentially massive but largely unknown effects of globalization on the environment, which have been largely identified by ecologists. Here enormous uncertainties over the impacts of globalization, some of which could well be irreversible, are reasons enough for no more globalization. Where the green counternarrative looks at the planet and sees global certainties and destructive processes definitively at work, those ecologists and others who subscribe to the ecological counternarrative know that ecosystems are extremely complex, and thus the planet must be the most causally complex ecosystem there is. The ecological counternarrative opposes globalization because of what is not known, while the green counternarrative opposes globalization because of what is known. The former calls for more research and study and invokes the precautionary principle—do nothing unless you can prove it will do no harm. The latter says we do not need more research or studies to do something—take action now—precisely because we have seen the harm.

In practice, the two environmental counternarratives are conflated. It is often argued, for instance, that agricultural biotechnology should not be promoted until we know it will not cause environmental harm—and anyway, you can't trust biotech companies, can you? Conflation, however, in no way reconciles the substantive contradictions between the opposing environmental narratives.

Which is the better counternarrative? Clearly, the ecological one is preferable over the green version, if only because it takes uncertainty and complexity seriously throughout, while its green counterpart goes out of its way to deny that any dispositive uncertainties and complexities are at work. Does this mean the ecological version is the better environmental counternarrative? That depends. Plainly, the ecological counternarrative is itself open to all manner of critique simply because of its reliance on the precautionary principle, e.g., I can no more prove a negative than I can show beforehand that application of the precautionary principle itself will do no harm in the future. That said, from the narrative policy analysis framework, critiques against the precautionary principle or for that matter against any environmental counternarrative are of little decision-making use, if they are not accompanied by alternative formulations that better explain how and why it makes sense to be environmentally opposed to globalization.

Is there at least one metanarrative that explains how it is possible to hold, at the same time and without being incoherent or inconsistent, the dominant policy narrative about globalization, its ecological and green counternarratives, and other accounts,

including critiques of both? Readers may have different metanarratives in mind, but the one I am most familiar with is the metanarrative about who claims best to steward the planet. The proponents of globalization and the proponents of the environmental counternarratives disagree over what is best for the environment, but both groups are part and parcel of the same technomanagerial elite who assert they know what is good for us. According to this metanarrative, we need free trade, the precautionary principle, more research, or whatever because we—the unwashed majority—cannot effectively steward our resources on our own with what we now have. While we may bristle against the condescension of these elites—who elected any of them to steward the earth?—it is patent that we must look for other counternarratives and metanarratives if we are to get out of the bind of trying to decide between the current candidates discussed so far.

Where are those other candidates? Clearly, part of the problem is that the counternarratives and the metanarrative just discussed are in no way specific enough to guide real-time decision making. For if ecosystems and the environment matter as much as each counternarrative argues, then they must matter to us, now, right now. Not tomorrow, not over the longer term of “sustainable plans for [name the resource],” but now. It is right now that decision making requires ecologists, environmentalists, and others to tell the operators in the control rooms of our large water supplies and electricity systems, for example, whether to open that water gate and save the manatee or close it and dry out the already endangered habitat; bring online that older generation plant to keep the lights thereby increasing air pollution; close down that pump to save the delta smelt but risk breaching urban water quality standards; or some equally urgent trade-off requiring choice. Yet neither counternarrative nor the metanarrative just discussed is detailed enough to answer these kinds of questions about real-time trade-offs.

CONCLUSIONS

Fortunately, entirely new fields are arising out of real-time decision making across our critical infrastructures and from these fields arise new policy narratives. Ecologists are increasingly in and around the control rooms making just such decisions, where “real-time ecology” is knowledge emerging out of the involvement of these ecologists (broadly writ) in the day-to-day, if not hour-by-hour, operation of our large water and hydro-power supplies. The involvement arises not out of the environmental counternarratives just discussed, but primarily because of specific legislative and regulatory

mandates for endangered species and habitat protection. Moreover, the knowledge is typically narrativized as advice to system operators by ecologists based on the latter's case-by-case analyses, their recognition and anticipations founded on patterns perceived in a run of cases, and their localized contingency scenarios for cases that have not yet occurred but might well do so—all applied “just-in-time” for management purposes. It is to these fields we have to look for new narratives, and these managers—be they control room operators or villagers—are very much the marginalized voices to whom *Narrative Policy Analysis* recommended we give much more attention.

ACKNOWLEDGMENTS

I thank all of those who have found the narrative policy analysis useful. Keying in “narrative policy analysis” into a Google search indicates far more sources than I could review for this short article.

BIBLIOGRAPHY

1. Roe, E. Narrative policy analysis for decision making. In *Handbook of Decision Making*, Goktug, M., Ed.; Marcel Dekker: New York, 2006; pp. 607–626.

National Security Policy

Eric K. Leonard

Shenandoah University, Winchester, Virginia, U.S.A.

INTRODUCTION

The events of September 11, 2001 dramatically altered the landscape of American national security. However, in many ways, this traumatic event may have clarified a flailing policy agenda, while reinvigorating the unilateralist position that is such a prevalent part of traditional U.S. national security policy. This article will examine the transformation of national security policy from the post–Cold War to the post–9/11 era. The primary method for examining this change will be the presidential doctrines/strategies that guided American national security policy during this period of history.^a This article will conclude by analyzing the parallels between the early Cold War doctrines and the current National Security Strategy of the Bush administration. Although it appears that both exhibit similar understandings of how to engage the world, one primary difference exists—the employment of preemptive strikes.

POLICY OPTIONS

Within the realm of U.S. national security doctrine, the policy options available to U.S. decision makers tend to fall into two categories: isolationist and internationalist.^b The isolationist option is probably the oldest form of national security policy dating back to George Washington's Farewell Address. In this address, Washington urged American policy makers to "steer clear of permanent alliances with any portion of the foreign world" because of the fact that "Europe has a set of primary interests which to us have none or very remote relations." The only rationale for extending foreign relations is "commercial interests." As for "political connections," Washington warned future policy makers against such ties. In short, the isolationist policy option advises that American decision makers avoid entangling relations with foreign nation-states. By heeding such prudent advice, Washington believed that the result would be a greater sense of national security.

Internationalists propose a more engaged form of national security policy. This policy option gained popularity as the international arena evolved from a disaggregated set of political entities to a more interdependent, globalized world. Internationalism promotes a national security policy agenda in which the United States actively pursues its national interest via involvement, not isolation. However, when discussing this policy option, one must also understand that several methods exist for achieving its goals. The two most prominent methods are unilateralism and multilateralism.

Unilateralism proposes that the United States promote its national security by pursuing its *own* agenda on its *own* terms. Any concern for allies and their interests, legal principles, and/or morality are all subservient to American national interest as defined by American policy makers. Thus a unilateralist policy agenda understands the need to engage the world, but always according to American interests and needs. There is no room for pandering to other global entities—whether they are nation-states or international organizations. Cooperation only occurs when those with whom the United States is cooperating accept the U.S. conditions for engagement. In short, a unilateralist perspective states that national security is best protected by a policy that engages the world according to U.S. national interest with no regard for compromise.

Multilateralists discuss engaging the world through more cooperative means, with cooperation meaning compromise between competing national interests. Advocates of this position recommend establishment of and adherence to international law, cooperative arrangements, and the defining of interests via a cosmopolitan perspective as opposed to an autonomous understanding of national interest. The resulting policy agenda would entail high levels of cooperation among nation-states and international organizations as a method for solving global problems, including issues of national security. Thus this perspective advocates the use of the UN as a forum for national security decision-making processes, along with a belief that international legal institutions are a quintessential tool for pursuing one's security interests. Consequently, multilateralists believe that cooperative/collective means are the best method for protecting American national security.

^aRef.^[1] is an excellent source of analysis concerning these doctrines.

^bRef.^[2] provides a detailed examination of these terms.

HISTORICAL OVERVIEW

Historically, U.S. national security policy has dabbled in all three of these policy options. Most scholars consider the pre-World War II era a time of isolationism, with moments of unilateralism (such as the Spanish-American War).^c However, with the outbreak of World War II, and, more importantly, the attack on Pearl Harbor, U.S. policy makers realized that an isolationist policy was simply not an option anymore. The world had become too interrelated for a rising hegemon to wall off its borders and exist in peace. Thus the United States began the modern period of internationalism.

In general, the Cold War appeared to be a time of unilateralism, with U.S. doctrine dominating any “cooperative” arrangements that may have existed. In other words, cooperative institutions did exist, but they existed on U.S. terms and were controlled by U.S. policy. American policy makers pursued their definition of national interest as they saw fit, with little regard for competing interests. But the Cold War period was more or less conducive to this type of environment. This was a time of ideological war in which the national security interests of two superpowers, the United States and the Soviet Union, dominated the global arena. Thus this power struggle not only defined American national security policy during the Cold War, but also the national security policies of the majority of the global community.

With the end of the Cold War, the United States enters a period of confusion concerning its national security policy. The Evil Empire is gone, the world appears safe from the scourge of global communism, and the liberal democracy stands alone as the victor. In such an environment, how does the global hegemon define and pursue its national security interests?

CLINTON'S POST-COLD WAR POLICY

With the onset of the post-Cold War era, President William Jefferson Clinton found himself searching for a national security directive. With the collapse of the Soviet Union, the United States lost its main adversary and the primary threat to its national security. As a result, the Clinton administration needed to redefine national security interests in an era devoid of liberal-communist ideological conflict. In September of 1993, President Clinton's first national security adviser,

Anthony Lake, established an outline for a new national security agenda.^[3] According to Lake, the new security strategy must entail an active engagement in the world “in order to increase our prosperity, update our security arrangements and promote democracy abroad.” This engagement would entail an attempt to: 1) strengthen the world's market democracies; 2) foster and consolidate new democracies and market economies where possible; 3) counter aggression and support the liberalization of states hostile to democracy; and 4) help democracy and market economies take root in regions of greatest humanitarian concern. According to Lake, the justification for such a policy was that, “our own security is shaped by the character of foreign regimes.” Thus it is imperative that the United States actively pursue a policy of engagement and enlargement.

This policy of enlargement had an air of multilateralism, or as UN Ambassador Madeline Albright characterized it, “assertive multilateralism.”^[4] According to the Clinton administration, American security was predicated on the proliferation of market democracy. If this ideological perspective took hold in all regions of the world, America's national security interests would be protected.^[5] However, this meant that the United States must actively pursue a policy of enlargement through the means of engagement. Isolationism was simply not an option in an era of globalization.^[6] The pursuit of national security now encompassed a broader definition of security than during the Cold War era, thus providing the Clinton administration with justification for a more active foreign policy agenda.

Somalia was the first and last real test for assertive multilateralism. Despite success in quelling the famine in Somalia, the catastrophic events of October 3, 1993 altered the direction of national security policy again. On that day, Somali rebels killed 18 American soldiers and then dragged their bodies through the streets of Mogadishu. As a result of the botched American mission and a public outcry against the American presence in Somalia, President Clinton began to rethink the notion of an expansive national security agenda. The result was Presidential Decision Directive 25 (PDD-25). This document, while still maintaining an internationalist foundation, initiated a period of more controlled multilateralism bordering on unilateralism.

PDD-25 limited the engagement policy to certain situations. According to this document, American intervention will only occur if the following conditions are present: 1) there must be an identifiable U.S. interest at stake in the operation; 2) the mission must be clearly defined in size, scope, and duration; 3) there must be sufficient resources, along with the political will to complete the mission; 4) there must be a tangible exit strategy in place; 5) the cost of U.S. involvement in multilateral peace operations must decrease.

^cThis statement on American isolationism is a generalization. True American internationalism did not take hold until the WWII era; however, moments of internationalism certainly permeated U.S. policy prior to the WWII period. One example of this internationalism is the Spanish-American War.

The first attempt at a national security policy predicated on multilateralism appeared to have fallen by the wayside. Although President Clinton maintained the rhetoric of a multilateralist, the practice was much more unilateral.^d Some prime examples of this unilateral pursuit of security are the following: the failure to engage in a multilateral military operation in Rwanda; the lack of support for the formation of a permanent International Criminal Court; the failure to pay UN arrears; the slow reaction to the decade-long crisis in the former Yugoslavia, among others. All of these situations are prime examples of the Clinton administration acting in a manner that benefits U.S. interests, but with little or no regard for the interests of others in the global community. Any intervention was to take place on U.S. terms with no room for debate or compromise. Consequently, one can classify the Clinton administration as striving for a multilateral security agenda, but ultimately falling into a more unilateralist one.^e

GEORGE W. BUSH AND SEPTEMBER 11TH

George W. Bush's initial stance on national security bordered on isolationism, not internationalism.^f At the outset of his term, Bush appeared to heed the advice of George Washington by publicly admonishing almost every multilateral arrangement that was under negotiation. Bush rejected the Kyoto Protocol, the land mines convention, and the small arms treaty; President Bush also unsigned the Rome Statute for the International Criminal Court and withdrew from the Anti-Ballistic Missile Treaty. The only form of multilateral arrangement that President Bush supported was economic/capitalist ones. If an international cooperative arrangement did not favor U.S. economic interests, then his administration felt it prudent to turn inward. However, the events of September 11, 2001 changed everything concerning U.S. national security policy.^g

With the tragic events of 9/11, the Bush administration realized the need to engage the world in a more

fundamental manner. The attack on the United States demonstrated that national security in the 21st century is not realized by isolating one's self. In fact, isolationism is not an option because of the globalization process. President Bush had to become an internationalist, but what means would he choose to pursue this policy of engagement?

In September 2002, President Bush published *The National Security Strategy of the United States of America*. In this document, President Bush outlines not only the means to national security, but also the enemy the United States is engaging and the values the United States must uphold. The explicit rhetoric of this document smacks of a multilateralist tone. For example, in the introductory section, the Bush doctrine states:

we [the United States] do not use our strength to press for unilateral advantage. We seek instead to create a balance of power that favors human freedom.

Again, in the opening section, the doctrine states:

We are also guided by the conviction that no nation can build a safer, better world alone. Alliances and multilateral institutions can multiply the strength of freedom-loving nations. The United States is committed to lasting institutions like the United Nations, the World Trade Organization, the Organization of American States, and NATO as well as other long-standing alliances.

However, the multilateral rhetoric is not about a cosmopolitan agenda of cooperation and compromise. Instead, the multilateral rhetoric extends only to situations in which the United States is in full control of the mission/organization. Thus the United States is committed to institutions such as the United Nations (as stated in the doctrine), but this is because of the fact that the United States has veto power over the UN's actions. This is the form of multilateralism the Bush administration favors—multilateral arrangements that not only favor U.S. interests, but also are controlled by U.S. power. The post-9/11 situation in Iraq is a prime example of U.S. unilateralism in action.^[12] The initial engagement of Iraq came in the form of multilateral efforts. The United States allowed the UN to play a role in the situation by collectively agreeing that the formation of a weapons inspection team was beneficial to the situation. However, when this team failed to find any weapons of mass destruction, the United States decided to intervene according to its own agenda.

At this point, the Bush administration disengaged the Iraq debate from the UN, in particular, the UN Security Council, because of the contentious nature of the topic. The United States knew that France,

^dThe National Security Strategy of Engagement and Enlargement [Ref.^[7]] document is one example of the constant struggle to find a middle ground between multilateral and unilateral methods.

^eThe Clinton administration did engage in some multilateral endeavors (such as leading the NATO airstrikes during the Kosovo crisis); however, if one is going to generalize about the administration's overall record on national security, unilateralism appears to be the primary method for achieving their defined goals.

^fSee Refs.^[8,9] for an understanding of Bush's initial national security agenda.

^gRefs.^[10,11] provide excellent analyses of the post-9/11 policy options.

Russia, and, possibly, China would veto any resolution authorizing a military intervention. Consequently, the United States chose to pursue its national security policy in a unilateral fashion, with the international community of states either standing by our side or getting out of the way. Some nation-states chose to stand by the United States and support our action (the coalition of the willing), but this ad hoc alliance of states was inconsequential to the action the United States was going to take. Bush had decided to invade, and no foreign power was going to alter this agenda. In the words of secretary of defense Donald Rumsfeld, "The mission must determine the coalition; the coalition must not determine the mission."

In the post-9/11 era, the United States begins a path toward unabashed unilateralism. In many ways, the events of September 11th have assisted U.S. policy makers in defining national security. During the Clinton years, the administration was left searching for a purpose for its national security agenda. There was no evil empire or ideological enemy. Instead, Clinton asserted that the promotion of free market liberal democracy would solidify American national security. However, in the post-9/11 era, a new focus/enemy exists—terrorism.

PREEMPTION AS A POLICY OPTION

As with the Soviet Union and its communist ideology in the immediate post-WWII era, U.S. national security policy now has a purpose—to combat terrorism. In the words of President Bush:

Defending our Nation against its enemies is the first and fundamental commitment of the Federal Government. Today, that task has changed dramatically. Enemies in the past needed great armies and great industrial capabilities to endanger America. Now, shadowy networks of individuals can bring great chaos and suffering to our shores for less that it costs to purchase a single tank. Terrorists are organized to penetrate open societies and to turn the power of modern technology against us.^h

The question that now looms is how to combat the enemy. During the Cold War years, the primary method for protecting American security interests was deterrence. In the post-9/11 period, the Bush administration has begun the use of preemptive strikes.ⁱ According to the Bush administration, because the enemy no longer fits the traditional nation-state

model, the old rules of engagement no longer apply. Deterrence was a plausible option when engaging other sovereign states that exerted a relative balance of power with the United States. However, the United States, or any other nation-state, does not have the ability to deter the current enemy (non-state terrorist organizations). Consequently, the use of force in a preemptive and preventive manner is necessary. Engaging in a policy of deterrence would simply result in future terrorist attacks like September 11th. The only way to prevent such tragedies is to strike first—as in Iraq.

The use of preemptive force is a new form of national security policy. Never before has the United States publicly called for the employment of a first-strike policy. But according to the Bush administration, the current era of "shadowy networks of individuals," whose use of weapons of mass destruction could perpetrate another September 11th or worse, necessitates the use of preemption. The Bush administration has altered the course of U.S. national security policy, but according to this administration, that is because the enemy has changed. A new enemy necessitates the initiation of new methods. The result is that preemption appears to be the future of national security policy.

CONCLUSION

The post-Cold War era is clearly a dynamic era full of rapid change. The national security policies of the two presidential administrations bear this fact out. With the onset of the post-Cold War era, President Clinton found himself having to redefine American national security. The ideological conflict between the Soviets and the Americans had disappeared and left U.S. policy makers with no clear security agenda. Clinton first initiated a multilateral policy agenda, but the conflict in Somalia quickly altered the method for obtaining and the definition of national security. The Clinton administration found it necessary to abandon much of the multilateral practices they had wanted to initiate for a more unilateral policy. However, even with this change toward unilateralism, the definition of what constitutes a national security threat was still unclear.

The events of September 11th provided President Bush with a specific enemy to center his national security policy around. Terrorism became the new threat to the American people, and in the vein of the Truman Doctrine and NSC-68, the Bush doctrine established a roadmap for how to protect the American people and eradicate the threat. The major difference between the post-WWII policies and the post-9/11 policies is notion of preemption. No longer does the United States wait to be attacked. President Bush, by his

^hRef.^[13] introductory section.

ⁱRef.^[14] contains an expansive discussion of the Bush doctrine's policy agenda and the premise of preemptive action.

actions in Iraq, has made it clear that when his administration perceives a viable threat to national security is in existence, the U.S. military will act. This sets the United States down an unprecedented course in security policy, one whose future, and the future of American national security, remains shrouded in doubt.

REFERENCES

1. Oliver, J.K. *The Foreign Policy Architecture of the Clinton and Bush Administration*. White House Studies, Wntr, 2004.
2. Patrick, S. Multilateralism and its discontents: the causes and consequences of US ambivalence. In *Multilateralism and US Foreign Policy: Ambivalent Engagement*; Patrick, S., Forman, S., Eds.; Lynne Rienner Publishers: Boulder, CO, 2002; 1–46.
3. Lake, A. *From Containment to Enlargement*; Speech given at the School of Advanced International Studies, Johns Hopkins University, September 21, 1993.
4. Albright, M. “Myths of peace-keeping.” statement before the subcommittee on international security, international organizations, and human rights of the house committee on foreign affairs, Washington, DC, Dispatch **1993**, 4, 464–467.
5. Brinkley, D. Democratic enlargement: the Clinton doctrine. *Foreign Policy Spring* **1997**, 106.
6. Berger, S.R. A foreign policy for the global age. *Foreign Aff.* **2000**, 79 (6), 22–39.
7. Clinton, W.J. *A National Security Strategy of Engagement and Enlargement*; White House: Washington, DC, 1996.
8. Hirsh, M. Bush and the world. *Foreign Aff.* **2002**, 81 (5), 18–43.
9. Rice, C. Promoting the national interest. *Foreign Aff.* **2000**, 79 (1), 45–62.
10. Smith, S. The end of the unipolar moment? September 11 and the future of world order. *Int. Relat.* **2002**, 16 (2), 171–183.
11. Haass, R. *Defining US Foreign Policy in a Post-Post-Cold War World*; Department of State: Washington, DC, 2002. <http://www.state.gov/s/p/rem/9632.htm>.
12. Nye, J.S., Jr. U.S. power and strategy after Iraq. *Foreign Aff.* **2003**, 82 (4), 60–73.
13. Bush, G.W. *The National Security Strategy of the United States of America*; White House: Washington, DC, 2002.
14. Ikenberry, J. America’s imperial ambition. *Foreign Aff.* **2002**, 81 (5), 44–60.

Natural Resource Policy: Institutional Aspects

Amy R. Poteete

Department of Political Science, Concordia University, Montreal, Quebec, Canada

INTRODUCTION

Policies for natural resource management contend with a number of challenges. The complexity of natural resource systems places a premium on scientific knowledge and strains the cognitive capacity of policy makers.^[1,2] Natural resource systems rarely correspond to political jurisdictions. Interdependencies across jurisdictions present administrative, strategic, and distributional challenges.^[3] While the scale of many resource systems increases the importance of cooperation across jurisdictions, spatial variability in resource conditions creates a contrary pressure for local management.^[4] Difficulties in limiting use present incentive problems even in the absence of problems related to complexity, scale, and local variability. How natural resource systems are conceptualized, the choice of management goals and time horizons, the scale of jurisdictions, and patterns of accountability all have important distributional consequences. This entry elaborates on scientific uncertainty, issues of scale and exclusion, and spatial variability as they affect natural resources and draws out the implications of each issue for natural resource policy.

SCIENTIFIC UNCERTAINTY, COGNITIVE LIMITATIONS, AND EXPERTISE

Each natural resource (e.g., trees and water) is part of a natural resource system (e.g., forests and rivers). Natural resource systems have multiple interdependent components, so that the condition of any resource depends on interactions with other parts of the system. Thus, the condition of trees in a forest depends on the condition of soils, groundcover, and animals, among other things. Interactions within a resource system can be quite complex, playing out over decades and operating through mechanisms that are not readily apparent. The complexity of natural resource systems presents cognitive, distributional, and democratic challenges.

Complexity poses a cognitive challenge whether one assumes full or bounded rationality. Fully rational decision makers maximize their utility by assigning values (utility) to each outcome and choosing the policy (or strategy) that yields the highest valued

outcome. When each policy yields more than one possible outcome, the capacity for rational decision making depends on the ability to assign probabilities to each outcome. If this is possible, decision makers can calculate their expected utility by summing the value of each possibility multiplied by the probability that it will occur. It is then possible to maximize expected utility.

The complexity of natural resource systems raises doubts about the applicability of expected utility theory to this policy area. Natural resource policies often involve significant risks, both because of the high degree of uncertainty about interactions within these complex systems and because decisions—or non-decisions—may lead to irreversible ecological changes. Expected utility calculations assume risk neutrality. There is considerable evidence, however, that most people avoid risky choices in favor of certain options or the status quo.^[5] Interactions in natural resource systems are often sufficiently complex that it is not possible to identify all possible outcomes or assign probabilities to outcomes. Under such conditions, calculation of expected utility becomes impossible.

Models of bounded rationality emphasize limited cognitive capacity and scarcity of attention. Limited searches and heavy reliance on cognitive shortcuts such as standard operating procedures generate suboptimal policy choices regardless of policy area. The interaction of bounded rationality with the complexity of natural resource systems has at least two important implications. First, unless a natural resource system shows dramatic signs of collapse, signals of resource strain may not catch the attention of decision makers until degradation is quite advanced. Second, the consequences of imperfect policy choices reverberate through natural resource systems in ways that may be partly or wholly irreversible. Thus, the shortcomings associated with bounded rationality are amplified in natural resource policy.

Scientific complexity also increases the influence of experts in the policy process.^[6] Experts are supposed to increase the rationality of policy making by expanding the information base and reducing uncertainty.^[6] But expertise exacerbates information asymmetries inherent in any delegation of decision making. The limited ability of non-expert supervisors and politicians to evaluate expert advice increases the scope for expert discretion and decreases capacity for democratic

accountability. Different experts prioritize different values: ecological integrity, sustainability, economic returns, human welfare, etc. Expert advice may deviate from democratically expressed preferences, whether by favoring ecological integrity over human welfare or by favoring some uses over others.

SPATIAL VARIABILITY AND (DE)CENTRALIZATION

Dramatically different microclimates develop in mountainous areas in response to aspect (e.g., southern or northern exposure) and elevation. In arid rangelands, pasture quality varies spatially and temporally with scattered rainfall patterns. The complexity represented by spatial and temporal variability challenges centralized policy making and implementation. Standardized national policies will be excessively restrictive in some times and places yet grossly inadequate in others, producing wildly divergent outcomes. In principle, locally variable policy can limit waste and degradation by adapting to local conditions. Context specific policy must be very detailed or allow considerable local discretion. Shortages of attention and administrative preferences for standard procedures exacerbate the inherent challenges of collecting and processing the tremendous quantities of information required for context-variable policy. Central decision makers rarely have the information or understanding required to identify all possible situations and devise detailed prescriptions for each.^[4] Furthermore, the viability of centralized management is limited by enforcement problems, especially when exclusion from a resource is difficult.^[7]

Perceived failures of centralization have contributed to interest in decentralization of natural resource management. Local decision makers were assumed to have a better sense of local resource conditions, a better understanding of how local natural resource systems respond to various interventions, and a better capacity to enforce management rules.^[3,4,8] Even under favorable conditions, local decision makers do not always have a full understanding of interactions within natural resource systems, particularly over longer periods of time. Moreover, the effectiveness of decentralization depends on a minimal degree of local consensus about

the nature of the management problem and how it should be addressed, supported by legitimate authority to make and enforce decisions about natural resource management. As discussed in the next section, these conditions do not always obtain.

INTERDEPENDENCIES AND JURISDICTIONS

Management of many natural resources involves interdependent decision making. Unless coordinated, the decisions by one actor or jurisdiction may undermine the effectiveness of decisions by others or simply shift problems from one location to another. Or actors may reject an otherwise acceptable policy because the benefits will be shared with others who are perceived as not doing their fair share. The interdependent nature of natural resource systems creates collective action problems. Because interdependency often spans jurisdictions, it also challenges their delineation.

Collective action is hindered by coordination problems or the threat of free riding. The nature of the resource influences the kind of collective action problems that emerge. As illustrated in Fig. 1, resources (or economic “goods”) may be characterized along two dimensions: subtractibility and ease of exclusion. Subtractibility, also referred to as rivalry in consumption, indicates that use of the resource by one actor reduces its availability for use by others. Resource use encompasses non-subtractible consumption (e.g., viewing a sunset) as well as subtractible consumption (e.g., fishing). Subtractible goods present problems of depletion; non-subtractible goods cannot be depleted, by definition, but their value may be lowered through crowding. Ease of exclusion depends in part on the physical nature of the resource and technologies that influence the capacity to monitor use. Combinations of these two dimensions yield a fourfold classification: Private goods are subtractible but exclusion is easy; club or toll goods are non-subtractible goods for which exclusion is easy; common pool resources or goods combine difficult exclusion with subtractibility; and public goods feature difficult exclusion and non-subtractibility.

Externalities—side-effects of actions that affect somebody other than the actor—represent an important

	Exclusion easy	Exclusion difficult or costly
Subtractible (rivalrous in consumption)	Private goods	Common-pool goods
Non-subtractible (non-rivalrous in consumption)	Club or toll goods	Pure public goods

Fig. 1 Types of goods.

subtype of non-excludable good. Externalities may be positive, as when tree planting on a hillside improves water quality at the foot of the hill, or negative, as when water pollution affects people downstream. Positive public goods, common pool resources, and externalities tend to be undersupplied because the producer does not enjoy the full benefits; negative public goods, common pool resources, and externalities tend to be over-supplied.

When exclusion is easy, resource management requires the design of institutions to reinforce exclusion and limit either extraction or access. When exclusion is not easy, resource management must address free riding as well as levels of use or access.

Natural resource policies that build on ease of exclusion seek to match the boundaries of rights and responsibilities to the boundaries of the resource system. In essence, these policies redefine property rights within jurisdictions or create new jurisdictions. Thus, privatization is supposed to reduce land degradation by ensuring that the land owner bears the full cost of land use, whether in the form of grazing pressure, timber production, or the sustainability of agricultural techniques. The same premise lies behind the creation of special jurisdictions for the management of river basins and aquifers that overlap with several subnational or national political jurisdictions. Pricing policies also involve the formalization and redefinition of rights and responsibilities for resources and the consequences of actions that were not previously priced. The creation of new formal rights and associated markets for those rights can be seen in carbon markets and similar programs.

Appealing in the abstract, institutional solutions can run into fatal political problems. Alterations to property rights within a single jurisdiction have distributional implications that may hinder otherwise desirable reforms. Natural resource systems encompass a variety of resources that are used or valued by different individuals and groups in different combinations and in different ways. Subsistence and commercial users of forests, for example, value different tree species and give different priority to trees and non-timber resources. Others value forests for their influence on the water table, soil conservation, habitat protection, or for their very existence. To take another example, cattle may be raised for beef markets or kept for ploughing, for milk, or as a store of wealth. The complexity of natural resource systems complicates policy making not only because these systems are difficult to understand, but also because policy choices often have significant distributional consequences.

Different management goals suggest different management strategies. Policies that promote particular management strategies favor some interests and values over others. Although the well-to-do may enjoy

more influence over policy-making processes, problems of exclusion limit the enforceability of institutions that ignore opposing interests. Institutional solutions also become more difficult when they compete with or attempt to displace prior institutions, as discussed below.

Coordination problems arise when multiple solutions exist and solutions have different distributional implications or actors lack a mechanism for selecting the same solution. Unless communication is a problem (e.g., among pastoralists or other mobile populations), distributional conflict is more likely to obstruct coordination.

Free riding becomes a problem when exclusion is difficult. Although everybody might benefit from provision and maintenance of public goods, each actor would prefer to enjoy benefits without bearing the costs of the public good. Likewise, even if nobody wants common pool resources to be depleted, each actor would prefer others to bear the costs of preserving the resource. Within a single jurisdiction, individuals or households may face free-riding problems in managing a common water source, forest, grazing area, park, or other such resource. When common pool resources cross jurisdictional boundaries, incentives to free ride influence the behavior of jurisdictions and their officials as well as the individuals within each jurisdiction. Thus, countries that share a river basin or forest might wish to limit overall consumption yet resist restrictions on their own use. Even if the respective governments agreed on a formula for distributing limitations on use, the limitations might not be strict enough to sustain the resource system. Moreover, precisely because exclusion is difficult, governments may be unable to effectively restrict resource use by individuals and firms even if they agree to do so.

Institutional design can compensate for problems of exclusion, if reliable monitoring and enforcement of rules restricting access to and use of common pool resources can be achieved.^[9] Cohesive social relations, high levels of trust, and common interests in the shared resource facilitate effective monitoring and enforcement. Conversely, divergent understandings of the resource system, competing interests in the shared resource, and poorly coordinated or competing institutional arrangements exacerbate difficulties in exclusion. Problems of divergent understandings and interests may arise among local resource users or between jurisdictions. When jurisdictions are territorially nested (e.g., a locality in a region in a country), coordination across the various levels determines whether the jurisdictions reinforce or undermine management strategies at each level.^[3] Free riding by jurisdictions as well as coordination problems must be addressed when resources cross jurisdictional boundaries.

Special-purpose jurisdictions dedicated to natural resource management have been proposed for various resources, including aquifers, river basins, watersheds, forests, and wilderness areas. Where general-purpose jurisdictions reflect administrative, historical, or cultural boundaries, special-purpose jurisdictions attempt to encompass an entire resource system.^[10] The performance of special-purpose jurisdictions depends on relations with general-purpose jurisdictions with which they overlap but rarely coincide. When special-purpose jurisdictions are created by officials of general-purpose jurisdictions frustrated with their inability to manage shared natural resource systems, the two sets of jurisdictions are likely to reinforce each other. An example might be a river basin commission consisting of national governments. Whether the special-purpose jurisdiction adopts a sustainable management regime depends on bargaining among members. The threat of sabotaging implementation strengthens the position of each member, limiting the scope for acceptable policies but decreasing the probability of coordination or free-riding problems.

When special-purpose jurisdictions coexist with other authorities that have no input into their creation or operation, competition between jurisdictions can interfere with resource management. Donor agencies and national governments have created special-purpose committees for community-based natural resource management, for example, without involving local or regional authorities. Indeed, the creation of new special-purpose committees sometimes attempts to circumvent government bodies that are perceived as inefficient, corrupt, insufficiently committed to sustainability, or otherwise limited in their capacity for resource management.^[11] Officials associated with general-purpose authorities frequently view special-purpose committees with overlapping jurisdictions as competitors. The ability of special-purpose jurisdictions to achieve anything at all depends not only on dynamics of internal decision making, but also on the authority and legitimacy of competing institutions and the level of hostility between institutions. Authority and legitimacy flow from both formal and informal institutions and resources, limiting the capacity of donors and national decision makers to overcome local antagonism through the introduction of new formal institutions.

CONCLUSIONS

Several features of natural resources exacerbate more general policy-making challenges. Complex interactions within natural resource systems interact with organizational behavior to limit the capacity of policy makers to detect resource degradation before it becomes

severe. The complexity of natural resource systems also increases the cognitive challenges of developing appropriate policies for sustainable resource management. Shortcuts in decision making and action that deal with cognitive limitations become more costly because changes in natural resource systems may be irreversible. When scientific experts are brought in to help address cognitive challenges, the degree of complexity in natural resource systems exacerbates more general problems of democratic accountability in situations involving delegation of authority. Spatial and temporal variation in natural resource conditions further challenges centralized policy making and limits the effectiveness of standardized policies. Efforts to address variability in resource conditions require either highly detailed policies or decentralization of authority. Highly detailed policies further aggravate cognitive challenges, while decentralization will not occur unless policy makers give up some of their own authority.

Natural resources vary in degree of subtractibility and the difficulty of exclusion. Different combinations of these features present different types of challenges for making and implementing natural resource policy. When exclusion is not a problem, policies can focus on management of extraction levels. Debate will center on management goals, the appropriate level of overall extraction, and related distributional issues. When exclusion is a problem, policy must address monitoring and enforcement problems as well as management goals and extraction levels. Both coordination and free-riding problems must be addressed. These issues play out among individual resource users, between overlapping jurisdictions, and between political jurisdictions and resource users.

ARTICLES OF FURTHER INTEREST

- Administrative Discretion*, p. 58.
- Democracy and Public Policy*, p. 538.
- Knowledge Utilization: Issues and Recommendations*, p. 1097.
- Market Failure, Theory of*, p. 1203.
- Public Choice Theory*, p. 1618.
- Public Goods*, p. 1628.
- Public Policy Formation Under Conditions of Ambiguity*, p. 1634.
- Special District Governments*, p. 1788.

REFERENCES

1. Poteete, A.R.; Welch, D. 2004. Institutional development in the face of complexity: developing rules for managing forest resources. *Hum. Ecol.* **2004**, 32 (3), 279–311.

2. Wilson, J. Scientific uncertainty, complex systems, and the design of common-pool institutions. In *The Drama of the Commons*; Ostrom, E., Dietz, T., Dolsak, N., Stern, P.C., Stonich, S., Weber, E.U., Eds.; National Academy Press: Washington, DC, U.S.A., 2002; 327–359.
3. Berkes, F. Cross-scale institutional linkages: Perspectives from the bottom up. In *The Drama of the Commons*; Ostrom, E., Dietz, T., Dolsak, N., Stern, P.C., Stonich, S., Weber, E.U., Eds.; National Academy Press: Washington, DC, U.S.A., 2002; 293–321.
4. Ostrom, E. *Governing the Commons: The Evolution of Institutions for Collective Action*; Cambridge University Press: New York, U.S.A., 1990.
5. Quattrone, G.A.; Tversky, A. Contrasting rational and psychological analyses of political choice. *Am. Polit. Sci. Rev.* **1988**, 82 (3), 719–736.
6. Haas, P.M. Introduction: Epistemic communities and international policy coordination. *Int. Org.* **1992**, 46 (1), 1–35.
7. Hardin, G. The tragedy of the commons. *Science* **1968**, 162 (3859), 1243–1248.
8. Baland, J.-M.; Platteau, J.-P. *Halting Degradation of Natural Resources: Is There a Role for Rural Communities?* Oxford University Press: New York, U.S.A., 2000.
9. Gibson, C.C.; Williams, J.T.; Ostrom, E. Local enforcement and better forests. *World Dev.* **2003**, 33 (2), 273–284.
10. Hooghe, L; Marks, G. Unraveling the central state, but how? Types of multi-level governance. *Am. Polit. Sci. Rev.* **2003**, 97 (2), 233–243.
11. Gibson, C.C. *Politicians and Poachers: The Political Economy of Wildlife Policy in Africa*; Cambridge University Press: New York, U.S.A., 1999.

Naturalistic Decision Making

Anna C. Goldoff

*Department of Public Management, John Jay College of Criminal Justice,
New York, New York, U.S.A.*

INTRODUCTION

This article details an important new addition to decision-making theory and practice, the model known as “naturalistic decision making.” Naturalistic decision making (NDM) is the study of real people making crisis decisions. Its focus is on the use and usefulness of experience in making complex decisions in uncertain and constantly shifting environments.

Naturalistic decision making emphasizes the highly rated consequences, time pressures, lack of information, and incomplete goals that characterize most political, administrative, and public policy environments.^[1] It is deeply rooted in classical works in decision theory, political science, and public administration, especially in works done by Charles Lindblom and Herbert Simon.^[2,3] Its most prominent proponent today is the researcher Gary Klein, who reformulates NDM into his own Recognition-Primed Decision Model.^[4]

Naturalistic decision making is said to be the representation of a true paradigm shift in decision theory away from the scientific/classical/rational approach. This article defines naturalistic decision making; elaborates on its roots in political science and the public administration literature; and demonstrates how it fits into a new paradigm of decision-making theory. This new paradigm has evolved as an organic alternative to the classic linear rational model of economic theory.

Detailed elements and analysis of the naturalistic decision-making model will be offered here, along with new directions in the field and recommendations for its practical implications and implementation. This article ends with a summarizing note and conclusion on the impact and importance of naturalistic decision making.

OVERVIEW

Naturalistic decision making summarizes a paradigm shift in decision-making theory, veering the field far away from the classical ideal, the rational-comprehensive (or rational actor model) or analytical decision-making model (ADM).^[5] The rational model (based on economic “choice” theory) thrives in controlled, laboratory-like settings with fixed, concurrent alternatives. It stresses unlimited time and resources

in order to choose a “right” answer or choice.^[6] Cost-benefit analyses, probability theory, and decision trees are all used by the decision maker to find a correct policy or administrative action.

Since the 1970s, cognitive psychologists, political scientists, and policy analysts have been criticizing rational choice as the ideal method for politicians and public administrators to follow.^[7] Building on Barnard’s and Simon’s work^[8] on more realistic decision making in large organizations, the field of naturalistic decision making has been born.^[1] It incorporates attributes of incremental theory (Lindblom) and of Simon’s bounded rationality^[2] (to be discussed in more detail below).

Naturalistic decision making has been built on empirical studies of how real people in crisis decisions really make decisions.^[4] It emphasizes the roles played by experience, intuition, quick assessments, and expert judgment, rather than time-consuming formulas and decision trees. It is based on the idea that time is of the essence in most political and public policy decision making. Aspects of the model also emphasize the important roles played by training and the case study method in improving managerial capacities to decide successfully.^[9] Thus, it has profound theoretical and practical (technical) consequences for the field of decision analysis.

Recommendations from the naturalistic decision-making researchers also fit into the notion that a paradigm shift has occurred—that a so-called “New Paradigm” has emerged to counter the rational approach as a decision-making ideal.^[10] This new paradigm embraces change rather than runs away from it; is organic (one with its environment); recognizes pattern recognition and deduction (looking at the whole or big picture); moves decision makers away from “tunnel vision.” It accepts quantitative tools and fact-based models like risk analysis and strategic management but says that they are alternative tools, not substitutes for the intuition, experience, and judgment of a skilled manager or leader.

This new paradigm (of which naturalistic decision making is a part) fits in more with fluid, flatter, task force or task-based organizations that are harder to manage than the old bureaucratic, “chain of command” hierarchical entities (Weber).^[11] As government, the

private sector, and non-profit organizations try to adapt to fast-paced technological, informational, political, and environmental changes, they will be pressed to adopt the naturalistic decision-making model. Their decisions have more costly impacts, and the time and resources they have to make them in have been telescoped. So, a model that emphasizes fast-paced, crisis decision-making and which is based on concrete, provable training methods to improve decision makers should be much in demand.

THEORETICAL BACKGROUND/HISTORY

Naturalistic decision-making as a model of decision analysis arises out of a simple bifurcation in political science and public administration theory, what Lindblom calls the “root versus branch” theory of policy making in his classic essay, “The Science of ‘Muddling Through’.”^[12] Graham T. Allison calls the classical model of crisis decision making the “rational actor” approach, used by the analyst when he structures decisions or events as the purposive choices of consistent actors (the root method).^[12] He says that “Classical ‘economic man’ and the rational man of modern statistical decision theory and game theory make optimal choices in narrowly constrained, neatly defined situations.”^[12] He says that, “In economics, to choose rationally is to select the most efficient alternative, i.e., the alternative that maximizes output for a given input or minimizes input for a given output.”^[12] The rational decision maker faces a given problem and has a given set of alternatives and consequences from which to choose.^[12] The simple rational solution is to “select the alternative whose consequence is preferred . . . ranking each set of consequences in order of preference.”^[12] Allison sees the basic concept of rational action thus as goals and objectives, alternatives, consequences, and choice.^[12]

Incrementalists say, “whoa,” not so fast. Choice and decision making is not that easy. Instead of maximizing alternatives, real-life decision makers “satisfice” or do “good enough.”^[13] They use decision-making dodges or shortcuts like “the next chance,” “feedback,” “remediality,” “seriality,” and breaking decision-making bottlenecks.^[13] This results in incrementalism, policy-making that:

1. Concentrates the policy maker’s analysis on familiar, better-known experiences.
2. Sharply reduces the number of different alternative policies to be explored.
3. Sharply reduces the number and complexity of factors the policy maker has to analyze.^[13]

Herbert Simon calls this more realistic approach to decision making “bounded rationality” as opposed to

the comprehensive rationality of the classical model. He says that real-life decision makers factor (or split up) complex problems; they satisfice, they search, they are motivated by uncertainty avoidance, and they create organizational repertoires to deal with complex problems.^[12] Thus, Charles Lindblom concludes that “the piecemealing, remedial incrementalist or satisficer may not look like an heroic figure. He is nevertheless a shrewd, resourceful problem-solver who’s wrestling bravely with a universe that he is wise enough to know is too big for him.”^[13] Naturalistic decision making is a natural offshoot of Lindblom’s wise incrementalist. It is an approach that is based on empirical study of real-life crisis decision-makers in many different fields. These key decision makers satisfice rather than maximize alternatives; they use remediality, they try out serial alternatives; they break up bottlenecks; they factor complex problems; and most importantly, they make decisions based on experience, not cost-benefit analysis or the construction of elaborate decision trees. They almost never compare concurrent alternatives. Naturalistic decision making is the twenty-first century progeny of mid-twentieth century incrementalism.

NATURALISTIC DECISION MAKING

In *Designing TDG’s—A Tactical Decision Games Workbook—How We Decide*, the U.S. Marine Corps talked about the “classical (analytical) decision making model (ADM)” and said that for decision making to be rational or systematic, it has to follow the format below:

Concurrent comparison of multiple options (my emphasis) . . . identify all the possible options, analyze all these options, assign a value to each aspect of each option . . . and choose the option with the highest aggregate value . . . this is known as multi-attribute utility analysis.^[14]

This, as previously noted, is Lindblom’s “root” method and H. Simon’s “comprehensive rationality.” It is a method that is time consuming, as doing a complete multiattribute analysis on complicated policy, political, or military crisis decisions is going to take you a lot of time. Time is usually of the essence in crisis situations, and if it is, use of this approach is going to be minimized in time-sensitive, complicated situations.

“The analytical model requires a high level of certainty and accuracy of information.”^[14] If the future is uncertain and the context unknown, and “if information is missing or unreliable . . .”^[14] again, the use of an ADM is inhibited. Third, “reasoning power is essential to analytical decision-making, but *experience and judgment are not* (my emphasis).”^[14] Theoretically, a novice or an expert will come to the same conclusion after subjecting the problem to

the rational-comprehensive process. Intangible factors like “*judgement, intuition and insight*” (my emphasis)^[14] are not necessary and, in fact, are excluded, because they are not quantifiable.

What is naturalistic decision making? As we have seen, classical theory relies on “economic man,” a rational decision maker who maximizes (or optimizes) decisions in carefully controlled, neat and tidy situations. Decisions are made in a closed system, operating in a perfect world, with perfect, “all-seeing” knowledge and unlimited time and resources. Proponents and researchers know that this is not how decisions are made in the real world. Naturalistic decision making offers a contrasting paradigm consisting of:

1. Ill-structured problems (not artificial, well-structured problems).
2. Uncertain, dynamic environments (not static, simulated situations).
3. Shifting, ill-defined, or competing goals (not clear and stable goals).
4. Action/feedback loops (not one-shot decisions).
5. Time stress (as opposed to ample time for tasks).
6. High stakes (not situations devoid of true consequences for the decision maker).
7. Multiple players (as opposed to individual decision making).
8. Organizational goals and norms (as opposed to decision making in a vacuum).^[15]

As mentioned above, Gary Klein is the driving force who pioneered the naturalistic decision making model. In his book, *Sources of Power—How People Made Decisions*,^[4] Dr. Klein sums up 13 years of field studies that debunk the rational model and support, instead, NDM. He says that the rational approach emphasizes human limitations, making it seem as most key decision makers are constantly falling short. Klein says this is not true at all, that his empirical studies of “firefighters . . . pilots, nurses, military leaders, nuclear power plant operators, chess masters . . .”^[4] and others have demonstrated the exact opposite to be true. These studies have relied heavily on interviewing and extensive participant observation by Dr. Klein and his staff. Key, heavily stressed individuals in crisis situations almost never use the rational approach, in the form of raising and evaluating several or many concurrent alternatives.^[4] Key decision makers do not rely on what he calls “conventional sources of power . . . deductive logical thinking, analysis of probabilities, and statistical methods.”^[4] Instead, they use four things—“the power of intuition, mental simulation, metaphor, and storytelling.”^[4] They use these things in a context that is realistic and typical. Klein studied people under time pressure, experienced decision makers who were making high stakes decisions

with inadequate information, unclear goals, and constantly changing conditions and procedures.^[4] Of course, this describes most political or policy decisions as well.

Klein’s team found that in case after case, people do not use the rational approach—they have no time.^[4] They do not use two-option strategies, they “don’t compare anything.”^[4] In studying five commanders at work, he found that they did not “compare options” or “wrestle with choices.”^[4] Instead, when faced with complex, emergency situations, commanders “came up with a good course of action from the start”^[4] based on their experience. Klein stated that “their experience let them identify a reasonable reaction as the first one they considered, so they did not bother thinking of others.”^[4] He called this a skillful use of strategy—“recognition-primed decision-making.”^[4]

Klein’s decision makers tended to evaluate several options, but not concurrently—they did this one at a time. Also, they evaluated and rejected options in a matter of minutes. He related it to Simon’s notion of alternative generation and search, in short, of “satisficing,” as alluded to above.^[4]

To sum up, Klein’s decision makers like Herbert Simon’s did not search to maximize or find the best choice, they just turned up options until one worked.^[4] Also, rather than constructing analytical decision trees, real-life crisis decision makers used mental simulations to discover whether or not their preferred strategies would work.^[4]

Caroline Zsombok looked at themes in NDM research^[1] and found that in the past decade, NDM researchers have focused on the “big picture:” situational assessment and awareness; expertise; and cognitive task analysis.^[1] The latter, she said, “as opposed to behavioral task analysis is to understand the decision requirements . . . behind an experienced person’s job or task performance.”^[1] The impact, implementation, and themes for training decision makers are the topic of the next section.

NEW DIRECTIONS

Naturalistic decision making as a field promises to make better decision makers in large organizational settings. It states that successful crisis decisions are made due to the mature judgement, intuition, and experience of key decision makers. This translates into a focus on expertise. How can training move the novice through several stages of proficiency, to becoming an “expert” in the field? Because situational pattern recognition, mental modeling (simulations), and cognitive task analysis are part of the approach, researchers have focused on the case study approach in training and implementation. Constant repetition of various alternative decision-making problems, under time-stressed

(and incomplete information) conditions is required, much like the experience of a new resident in a hospital emergency room. Soon, the novice develops pattern recognition skills, after seeing the same set of symptoms again and again. Mental simulation of possible “outcomes” (and causes) leads him to develop shortcuts to the best possible solution, under the circumstances. Emergency room physicians do not have the time to develop elaborate alternatives and scenarios. The first and most basic solution to come to mind is usually tried and hopefully works. Cognitive task analysis also emphasizes the sharing of information and intention among decision makers, as timely communications are crucial in many real-life crisis and political and organizational tasks. Also, real-life organizational and public decision making tends to be made by teams, not single individuals. If teams are making the decisions, communications flows are key to making successful decisions. In this way, naturalistic decision making is in sync with other New Paradigm theories like chaos theory.^[16]

Training to improve “naturalistic decision makers” must make use of real-time scenarios and uncertain and dynamic situations (just as in the real world). Decision makers have to be trained to multitask (again, as in the real world, real-life problems do not present themselves one at a time). Immediate “thinking-aloud” verbal expressions of the task just completed can help higher-level managers and researchers to deconstruct actual decision-making processes. Levels of proficiency can be benchmarked through the number of successful solutions gained in these simulations and training sessions. In short, naturalistic decision-making researchers are saying that expertise and intuition can be “taught” or force-fed by constantly facing the decision maker with realistic conditions and constant problems to be solved.

CONCLUSION

Naturalistic decision making is a new field in decision-making theory and practice. It is a theoretical outgrowth of incrementalism and of H. Simon’s concept of “bounded rationality.” Its focus is on real-life decision making that has to take place in uncertain environments that are constantly changing with ill-defined or competing goals. Decision makers are charged to make these decisions under the stress of little time and clouded information, rather than in the perfectly realized world of goals, alternatives, and resources that represent the rational (actor) or choice model. The naturalistic decision maker does not know or cannot predict the future but can use mental simulation to find the best solutions to the problem that will “satisfice” or be good enough.

Naturalistic decision makers also work in teams in large bureaucratic organizations, so theories of individual choice are not as applicable. Decision making is not made in a vacuum but is influenced and effected by a complex environment.

Naturalistic decision-making research has relied heavily on the case study method, participant-observation, and key person interviews in crisis/emergency management-type professions. Its initial conclusions are that more empirical research is needed, but better decision makers can be made through a combination of repetitious and realistic case simulations.

REFERENCES

1. Zsombok, C.; Klein, G. *Naturalistic Decision Making*; Lawrence Erlbaum Associates: Mahwah, NJ, 1997; 13–432.
2. Lindblom, C. The science of muddling through. *Public Adm. Rev.* **1959**, 79–88, Spring.
3. Simon, H. *Administrative Behavior*, 2nd Ed.; The Free Press: New York, 1957; 259 pp.
4. Klein, G. *Sources of Power—How People Made Decisions*; MIT Press: Cambridge, MA, 1999; pp. 1, 3–6, 12, 16–18, 20, 21, 330.
5. <http://www.mcu.usmc.mil/staff/dec2.htm> (accessed January 2001).
6. <http://tuvok.au.af.mil/au/awc/awcgate/stratdm.htm> (accessed May 2000).
7. <http://www.dodccrp.org/copch10.htm> (accessed January 2001).
8. Barnard, C. *The Functions of the Executive*; Harvard University Press: Cambridge, MA, 1938; 322, See Ref. [13].
9. Zachary, W.; Ryder, J. Hicinbothom. Cognitive task analysis and modeling of decision-making in complex environments. In *Decision-Making Under Stress: Implications for Training and Simulation*; Cannon-Bowers, J., Salas, E., Eds.; American Psychological Association: Washington, DC, 2001; 1–20.
10. Goldoff, A.C. Decision-making in organizations: the new paradigm. *Int. J. Public Adm.* **2000**, 23 (11), 2017–2044.
11. Weber, M. Bureaucracy. In *From Max Weber: Essays in Sociology*; Gerth, H.H., Mills, C.W., Eds.; Oxford University Press: New York, 1946; 196–233.
12. Allison, G.T. *Essence of Decision*, 1st Ed.; Little Brown and Co.: Boston, 1971; pp. 28–30, 71–72, See also Ch. 1.
13. Lindblom, C. *The Policy-Making Process*, 1st Ed.; Prentice Hall: Englewood Cliffs, NJ, 1968; 24–27, Herbert Simon’s idea.
14. <http://www.mcu.usmc.mil/staff/dec2.htm> (accessed January 2001), p. 2.
15. Orasanu, J.; Connolly, T. The reinvention of decision-Making. In *Decision-Making in Action: Models and Methods*, 1st Ed.; Klein, G., Orasanu, J., Calderwood, R., Zsombok, C.E., Eds.; Ablex: Norwood, NJ, 1993; 3–20.
16. Wheatley, M. *Leadership and the New Science*, 1st Ed.; Berrett-Koehler: San Francisco, 1994; 101–120.

Net Assets and Fund Balance

Dean Michael Mead

Governmental Accounting Standards Board, Norwalk, Connecticut, U.S.A.

INTRODUCTION

Financial position—the status of a government’s assets, liabilities, and residual balances at a given point in time—is undoubtedly a staple of governmental financial analysis. The commonly accepted approach to evaluating financial position is to compare a measure of a government’s residual balances—its *net assets* or *fund balance*—with a measure of its annual financial activity.

Net assets and fund balances, therefore, may be among the most basic and important information reported in a government’s financial statements. This entry explains the nature of net asset and fund balance information and the central role it plays in governmental financial analysis. The first two sections describe the basic requirements under generally accepted accounting principles (GAAP) for state and local governments in the United States for reporting net assets and fund balance. The succeeding section then describes how this information is incorporated by financial analysts into assessments of governmental financial health.

NET ASSETS

State and local governments that prepare their annual external financial reports in conformity with GAAP are required to present four financial statements that could be described as statements of financial position. Three of these financial statements—the *government-wide* and *proprietary funds statements of net assets* and the *statement of fiduciary net assets*—are prepared using an accrual basis of accounting and economic resources measurement focus. This means that these statements report all of a government’s assets, both financial and capital, as well as all of its liabilities, both short term and long term.

In simple terms, net assets are the difference between assets and liabilities. In other words, if a government’s assets were used to repay its liabilities, net assets would be what was left over—they are literally the government’s assets *net* of liabilities. In the private sector, this is called retained earnings or owner’s equity. From a governmental perspective, net assets are resources that the government can use to provide services.

Government-Wide and Proprietary Funds Net Assets

The issuance of GASB Statement No. 34, *Basic Financial Statements—and Management’s Discussion and Analysis—for State and Local Governments*, substantially revised the basic model for financial reporting by subfederal governments in the United States.^[1] Statement 34 added new financial statements, prepared on a full accrual basis for the entire governmental entity, to the existing fund financial statements. One of these new accrual statements, a statement of net assets, reports the assets, liabilities, and net assets of a government. The information in the statement is typically presented in four columns: information related to *governmental activities* (typical governmental services supported primarily by taxes and other general revenues); *business-type activities* or BTAs (those services operated like private enterprises, charging fees intended to recover costs); the *total primary government* (governmental and business-type combined); and *discretely presented component units* (public benefit corporations and other legally separate entities for which the primary government is financially accountable) (see [Table 1](#)).

A unique feature of governmental accounting is its combination of consolidated information for the government as a whole and disaggregated information for specific portions of the government. These individual portions or *funds* are reported in three groups of financial statements: *governmental funds*; *enterprise funds*; and *fiduciary funds*. The proprietary funds are reported using the same full accrual accounting as the government-wide statements. Proprietary funds include *enterprise funds*, which are largely synonymous with BTAs, and *internal service funds*, which track the finances of activities that provide services to the government itself, such as a central garage or supply function. The proprietary funds financial statements show the *major* or the most significant enterprise funds separately and aggregate the remaining funds in a single column.

Net assets are reported in three categories at both the government-wide level and for the proprietary funds—*invested in capital assets net of related debt*, *restricted*, and *unrestricted*. The title of the first category is self-explanatory: the net book value of a government’s long-lived assets (i.e., the original cost

Table 1 Sample net assets section from a government-wide statement of net assets

Net assets	Primary government			Component units
	Governmental activities	Business-type activities	Total	
Invested in capital assets, net of related debt	\$123,456,789	\$76,543,210	\$199,999,999	\$15,151,151
Restricted for:				
Capital projects	34,567,890	—	34,567,890	4,343,343
Debt repayment	2,345,678	4,321,098	6,666,776	—
Recreation programs	5,678,901	—	5,678,901	—
Other purposes	1,234,567	2,109,876	3,344,443	—
Unrestricted	3,456,789	10,987,654	14,444,443	1,212,212
Total net assets	\$170,740,614	\$93,961,838	\$264,702,452	\$20,706,706

less accumulated depreciation) minus outstanding debt used to finance the purchase, construction, or rehabilitation of those assets. This feature of residual balance reporting is unique among accounting standards, signaling the importance of capital assets, particularly infrastructure assets such as roads and sewer pipes, in the provision of governmental services.

Restricted net assets represent resources that are legally limited to a particular purpose. Net assets can be restricted in three ways:

- By external parties, such as other governments, donors, or grantors.
- By constitutional provisions.
- By enabling legislation.

Enabling legislation, as it is used in Statement 34, is not run-of-the-mill legislation. Statement 34 specifically requires that it must authorize the government to raise an entirely new revenue source and bind it to a particular purpose with a legally enforceable restriction, rather than *earmark* existing resources. This feature is also unique to state and local governments. In designing it, the GASB viewed these characteristics of enabling legislation as constituting “an unbreakable compact with the resource providers that the resources would be used for the promised purpose only.” (See, paragraph 18 of Ref.^[2].) In other words, the government may not have been able to obtain the tacit agreement of the taxpayers to raise these resources if it did not promise to use them in a manner acceptable to the taxpayers. Furthermore, these taxpayers or others external to the government have the power to compel the government to use the resources only for the purposes specified in the enabling legislation.

Unrestricted net assets are all of the resources not in one of the other two categories. These resources can be used by a government for any purpose. A government

may disclose in the notes to the financial statements that it has tentative plans regarding how it will use its unrestricted resources, but it is not allowed to display such *designations* on the face of the statement.

Lastly, if a government’s net assets include resources from permanent endowments or permanent fund principal amounts, then a further distinction is made between net assets that are *expendable* and *non-expendable*. Nonexpendable resources are required to be retained in perpetuity—they can be invested and the earnings spent, but the nonexpendable resources themselves cannot be spent.

Fiduciary Net Assets

The fiduciary funds track activities in which a government acts as a fiduciary for resources that belong to other organizations or persons. Unlike the governmental and proprietary funds, fiduciary fund finances do not appear in the government-wide financial statements because these resources do not belong to the government and cannot be used to provide services. The most common example of a fiduciary fund is a *pension trust*. They also include *private-purpose trusts*, *investment trusts*, and *agency funds*.

The fiduciary funds do not use the three categories of net assets described above. Rather, the difference between plan assets and plan liabilities is captioned *net assets held in trust for pension benefits* or other purposes. Agency funds, which hold resources temporarily before they are distributed to other governments or persons and organizations external to the government, show no net assets at all. The assets of agency funds are exactly equal to the amount that must be passed along. Agency funds are often used when a government collects taxes on behalf of other governments; the receipts are held in the agency fund

until they are disbursed to the various governments, but at no time do those resources belong to or are available to the government collecting them.

FUND BALANCE

The aforementioned accounting and reporting takes place in a full accrual environment. The governmental funds, however, utilize a *modified accrual accounting basis* and *current financial resources measurement focus*. As a result, these funds (which are roughly synonymous with the governmental activities of the government-wide statements) focus on the short-term finances of the government—essentially, the year following the date of the financial statements. The governmental funds include the *general fund* or primary operating fund, *special revenue funds* (which track resources legally restricted to particular purposes), *debt service funds*, *capital projects funds*, and *permanent funds*. As with proprietary funds, the governmental funds statements show the most significant funds individually in their own columns and aggregate the remaining funds into a single column (see Table 2).

The counterpart to the statement of net assets—the governmental funds balance sheet—includes only current financial resources. These include cash and other assets expected to be liquidated within a year, and liabilities that are expected to be satisfied within a year. There are no long-term investments or capital assets; neither are there any long-term debts. The difference between assets and liabilities in the balance sheet is *fund balance*.

The focus of fund balance is on *availability* of resources. It is of particular interest to finance officers, legislators, and service advocates seeking to allocate resources to areas of need and favored projects. In

addition to new revenues raised during the year, expenditures can be financed with available resources that have been accumulated from prior years. Fund balance communicates the amount of resources available and the purposes for which they are available to be used.

Fund balance is reported in two components, reserved and unreserved. Two factors may result in a reservation of fund balance. First, resources may be in a nonexpendable form, such as inventory. Second, resources that have been “legally segregated for a specific purpose” are reported as reserved fund balance. Legal segregation includes the circumstances that lead to a restriction of net assets, but is broader: There are actions that would reserve fund balance, such as legislative earmarking, but would not restrict net assets.

It is important to note that the second kind of reservation interacts with fund reporting, which is intended to demonstrate that resources are limited to or intended to be used for a specific purpose. In the general fund, any legal limitation on the use of resources will be shown as a reservation of fund balance. However, the other governmental funds already communicate that the resources they contain are intended for a particular use. The resources in a debt service fund, for example, are supposed to be used to repay debt. Therefore, in order for a legal segregation to result in reserved fund balance in one of the other governmental funds, it must be for a purpose that is *more specific than the purpose of the fund itself*. For instance, if a government has resources that can only be used to rehabilitate school buildings, these resources could be shown as “reserved for school construction” (or a similar title) in a broad capital projects fund. However, if the government put these resources in a “school construction fund,” they would not be shown as reserved because they are not legally segregated from the other resources of *that fund*, all of which are there for school construction.

Table 2 Sample fund balance section from a governmental funds balance sheet

	General	Housing programs	Economic development	Other governmental funds	Total governmental funds
Fund balances:					
Reserved for:					
Inventories	111,111	—	—	8,888	119,999
Encumbrances	666,666	33,333	7,777	2,323,323	3,031,099
Debt service	—	—	99,999	—	99,999
Other uses	222,222	—	100,000	1,454,454	1,776,676
Unreserved, reported in:					
General fund	444,444				444,444
Special revenue funds		1,555,555		1,333,333	2,888,888
Capital projects funds			10,101,101	2,444,444	12,545,545
Total fund balances	\$1,444,443	\$1,588,888	\$10,308,877	\$7,564,442	\$20,906,650

Unreserved fund balance represents all resources that are not reserved and thus can be used for any purpose. Governments have the option of *designating* a portion of their unreserved fund balance in order to communicate what they *intend* to do with the resources. Actual practice varies widely. Roughly half of state and local governments report designations, primarily on the face of the balance sheet; a relatively small percentage of governments report their designations in the notes to the financial statements.^[3,4]

Information about designations is highly important to many financial statement users. Some users review individual designations of unreserved fund balance to determine which of a government's designations, based on experience, are likely to be used as identified in the balance sheet and which the users believe could be applied to any purpose. Thereby, these users construct an unreserved–undesignated fund balance amount that they use in their financial position ratios.

Designations, however, are highly problematic. Because they generally have no legal backing, but are merely statements of intent, there is no guarantee that a government will use those resources as it stated. The government is free to change its mind at any time. Second, because many governments do not report designations in their balance sheets, comparability problems arise; it may not be safe to assume that governments that do not report designations have not, in fact, identified intended purposes for their unreserved fund balances.

Third, many governments categorize their designations in a manner that communicates very little information to the financial statement user. For example, it is common for governments to designate unreserved fund balance for “future expenditures” or “subsequent years’ budgets.” These are meaningless titles, because all designations—in fact, all fund balance that is in an expendable form—are intended to be expended at some time in the future. Fund balance is, by definition, resources available for use in providing services or operating a government.

USING NET ASSETS AND FUND BALANCE INFORMATION TO ANALYZE FINANCIAL POSITION

Financial or operating position is likely the most common element of paradigms for analyzing the financial health of governments. Every major paradigm includes financial position or a similar term as a major component. For instance, it is one of the six basic financial factors affecting financial condition identified in the International City/County Management Association's financial analysis text.^[5] The “10-point test,” which is widely used by local governments to gauge their

finances, includes three ratios for measuring financial position.^[6] It is accorded similar prominence by Ives and Schanzenbach,^[7] Mead,^[8] Berne and Schramm,^[9] and Chaney, Schermann, and Mead.^[10] A review of the governmental financial analysis literature identified financial position as one of six major factors widely considered, along with surpluses and deficits, revenues and revenue base, expenditures/expenses and spending pressures, outstanding debt and debt service, pensions and other postemployment benefits, and liquidity.^[11]

Financial position is typically assessed with the use of ratios that divide net assets or fund balance by a measure of a government's annual financial activity or its aggregate financial size. There is considerable variation in what information analysts actually select for the numerator and denominator of that ratio, however, depending upon their particular interests and the types of governments they are examining. For example, some use total fund balance and others only unreserved or unreserved–undesignated fund balance. Similarly, total net assets or unrestricted may be employed. Furthermore, fund balance for the total governmental funds may be used, or only for the general fund or a combination of other funds. Likewise, some ratios include net assets for the primary government, and other just for the governmental activities or BTAs.

The denominator is either a flow indicator (revenues, expenses, expenditures) or, to a lesser extent, a stock indicator (such as assets or liabilities). Occasionally, a particular portion of revenues, expenditures, or expenses is utilized, such as operating revenues or operating expenses. In any case, the flow or stock information matches the scope of the fund balance or net assets information (total governmental funds, governmental activities, and so on).

A typical financial position ratio using net assets information from Table 1 might look as follows (the denominator amount would come from the government-wide statement of activities):

$$\left(\frac{\text{total primary government unrestricted net assets/}}{\text{total primary government revenues}} \right) \times 100$$

$$\text{e.g., } (264,702,452/6,000,000,000) \times 100 = 4.41\%$$

A typical financial position ratio using fund balance information from Table 2 might look as follows (the denominator amount would come from the governmental funds statement of revenues, expenditures, and changes in fund balances):

$$\left(\frac{\text{general fund unreserved fund balance/}}{\text{total general fund operating revenues}} \right) \times 100$$

$$\text{e.g., } (444,444/12,000,000) \times 100 = 3.70\%$$

Net asset information for the primary government is so new that rules of thumb or benchmarks have not yet been developed against which a net asset ratio can be judged. Regarding fund balance, the Government Finance Officers Association^[12] recommends that governments maintain a general fund unreserved balance of at least 5–15% of general fund operating revenues. By this standard, the fictional government in Table 2 would be deemed to have an undersized fund balance, *ceteris paribus*.

Analysts are aided in their assessment by schedules with 10-year trends in fund balance and net assets. Governments that publish their financial statements in a comprehensive annual financial report include such schedules for the most recent 10 fiscal years in the statistical section.^[13] The statistical section also includes 10-year schedules of changes in net assets and changes in fund balances, thus providing the information needed for the denominator. Analysts can analyze trends in financial position ratios, using these schedules, to determine not only a government's present financial position, but also whether its financial position has improved or deteriorated and to what degree.

CONCLUSIONS

Net assets and fund balance are two kinds of information state and local governments report in their financial statements that communicate the resources governments have that can be used to provide services. Net assets, a full accrual accounting construct, are the economic resources a government can employ in the provision of services and the operations of the government. Under modified accrual in the governmental funds, fund balance reports the current financial resources that a government has available to use. Net assets and fund balance each describe the resources that can be used for any purpose and those that are limited to specific uses.

When divided by indicators of annual financial activity, such as revenues, expenses, and expenditures, net assets and fund balances create ratios that inform the concept of financial position. Financial position—a government's financial standing at a given point in time—is one of the most commonly analyzed and most widely agreed upon elements of the financial health of

governments. Thereby, both net assets and fund balance are undoubtedly among the most crucial information needed for the financial analysis of governments.

REFERENCES

1. Governmental Accounting Standards Board (GASB). Statement No. 34. In *Basic Financial Statements—and Management's Discussion and Analysis—for State and Local Governments*; GASB: Norwalk, CT, 1999.
2. Governmental Accounting Standards Board (GASB). Exposure Draft. In *Net Assets Restricted by Enabling Legislation*; GASB: Norwalk, CT, 2004.
3. Governmental Accounting Standards Board (GASB). Board memorandum: results from preparer survey on net assets and fund balance, Dec 1, 2003.
4. Governmental Accounting Standards Board (GASB). Board memorandum: results of financial statement review, Dec 1, 2003.
5. Groves, S.M.; Valente, M.G.; Nollenberger, K. *Evaluating Financial Condition: A Handbook for Local Government*, 4th Ed.; International City/County Management Association: Washington, DC, 2003.
6. Brown, K.W. The 10-point test of financial condition: toward an easy-to-use assessment tool for smaller cities. *Gov. Financ. Rev.* **1993**, December, 21–26.
7. Ives, M.; Schanzenbach, K. *Financial Condition Analysis and Management*; Sheshunoff Information Services: Fort Worth, TX, 2001.
8. Mead, D.M. *An Analyst's Guide to Financial Statements*; Governmental Accounting Standards Board: Norwalk, CT, 2001.
9. Berne, R.; Schramm, R. *The Financial Analysis of Governments*; Prentice-Hall: Englewood Cliffs, NJ, 1986.
10. Chaney, B.; Schermann, K.R.; Mead, D.M. The new governmental financial reporting model: what it means for analyzing government financial condition. *J. Gov. Financ. Manag.* **2002**, 51 (1), 26–31.
11. Governmental Accounting Standards Board (GASB). Board memorandum: economic condition reporting literature review, Feb 4, 2002.
12. Government Finance Officers Association. Appropriate level of unreserved fund balance in the general fund, 2002. <http://www.gfoa.org/services/rp/caafr/caafr-appropriate-level.pdf> (accessed on September 29, 2004).
13. Governmental Accounting Standards Board (GASB). Statement No. 44. In *Economic Condition Reporting: The Statistical Section*; GASB: Norwalk, CT, 2004.

Networks, Organizational

Seok-Eun Kim

Department of Political Science, Kansas State University,
Manhattan, Kansas, U.S.A.

INTRODUCTION

The notion of network has become increasingly popular across disciplines. Political scientists conceptualize the complex interactions of multiple actors in the collective policy-making process as *issue networks*, organization theorists view patterned interconnectednesses of organizational systems as *networks*, and sociologists characterize complex social relations that exist among individual interactions as *networks or systems of networks*. In addition to the popularity in academics, the notion of the network is close to our daily life. For example, we often encounter a complex array of interconnected information-sharing systems such as local area network and wide area network. A group of television stations linked by wire and other relays provides another such example.

The proliferation of network usages rests on diverse theoretical foundations, which risk the term meaning so much to so many that it has little specific meaning to anyone. For example, “network” derives impetus from pluralism in political science, resource-dependence theory in sociology, and transaction costs in economics. Different scholars have treated networks differently, using such similar terms as issue network, policy network, policy subsystem, and policy community. In common, all these variants reject a simple process of formulating and implementing public policy and suggest new forms of governance that are neither formal hierarchy, perfect market, oligopoly, nor monopoly.

CONCEPTUALIZING NETWORKS

Despite the popularity of the notion, but in part because of that popularity, the meaning of network is still ill defined and often precludes advancement as a coherent theory. Nonetheless, hope exists, as in the work of O'Toole,^[1] who views networks as “structures of interdependence involving multiple organizations or parts thereof, where one unit is not merely the formal subordinate of the others in some larger hierarchical arrangement” (p. 45). In public administration, networks are considered as complex policy settings in which multiple actors interact to ensure that policy outcomes will suit their interests.^[2] According to these

definitions, networks have four features: 1) resource interdependence between actors; 2) involvement of multiple actors, each with their own interests; 3) lateral or horizontal patterns of exchange; and 4) unlimited territory.

These four features are usefully even if briefly, discussed. First, networks exist inasmuch as actors are interdependent in the policy process. The primary motive for interdependence is to obtain the necessary resources—such resources as people, information, and professional knowledge—to implement specific policies because individual actors are limited in terms of their resources and capacities to act. Hjern and Porter^[3] made it clear that “no single, comprehensive organization can command all the needed resources” (p. 214). Accordingly, network performance relies heavily on the ability of everyday organizations to create and to exploit essential resources with which organizations can manipulate environmental challenges for program success.

Second, networks consist of multiple actors, each with their own interests.^[2] In networks, government instrumentalities can become one or even several of the many actors. Complex public programs have increasingly been carried out in a mixture of public, private, and quasigovernment organizations, with the number of actors involved varying over time. Such combinations clearly require basic changes in the ways “implementation” is characterized. The single agency has served as the unit of analysis in much of traditional public administration. The unit of analysis increasingly has widened into three major classes of actors—executive bureaus, congressional committees, and interest group clienteles—in the classic “iron triangle.” However, Hecla^[4] advised against assuming a stable pattern of interaction. He warns that the closed system thought encouraged by the iron triangle tends to disguise fairly open networks where multiple actors increasingly impinge upon a protean “government.” From extensive search processes, for example, Laumann, Heinz, Nelson, and Salisbury^[5] identified at least 80 major corporate actors, each involved in the four U.S. national policy networks—agriculture, energy, health, and labor.

Third, networks are characterized by lateral or horizontal, not exclusively vertical, patterns of exchange.^[6]

Networks consequently reject both formal hierarchy and the perfect market, falling somewhere in or even beyond the alleged continuum between them.^[1] Actors in networks resist the hierarchical decision-making process in which federal government serves as the authoritative ruler, and in which lower levels of government, citizens, and private participants are considered more or less passive objects of networked interactions. The basic assumption behind the notion of networks is that actors, who possess relatively equal power, exchange their resources and information when they perceive mutual benefit from the interactions. However, influences of specific exchange may vary, depending on who occupies a better position in the network in terms of resources and information.^[2]

Fourth, the territories of networks seem unlimited. Networks exist worldwide, nationwide, regionally, and at local levels of decision events.^[6] Networks also take place among public, private, and quasigovernment organizations. For example, the U.S. government has long history in which private and non-profit providers actually take responsibility for provision of many public programs. Recently, in addition, international networks have been established in response to a growing dispersion of the resources and the limited capacities of a nation to deal with policy problems spanning across countries, such as environmental pollution and oil crises.

IMPORTANCE OF NETWORKS IN PUBLIC ADMINISTRATION

The growing importance of networks in public administration is no accident. Indeed, numerous reasons support the popularity and significance of the new form of governance. O'Toole^[1] declared that "complex networks are not only relatively common, they are also likely to increase in number and importance" (p. 46). In fact, O'Toole pointed out that the importance of networks is inescapable, not only because intergovernmental programs make up a sizable proportion of total government activities, but also because subtle arrays of collective interactions have considerably increased as evidenced, for example, by the expansion of quasigovernmental arrangements.

In addition, increasing reliance on forms of intergovernmental programs is a necessary by-product as policy problems become more "wicked" in such critical areas as health care, housing, and education.^[1] These policies are "wicked" because they are hardly manageable by single organizations; rather, they require network structures where complex issues must be coped with using a multitude of resources provided by many individual and collective actors. Indeed, scholars in policy implementation argue that basically

no program is fully implemented by a single organization. A complex pattern of multiple actors characterizes all implementation structures.^[3]

"Network" notions are also significant because they replace simplistic models. The development of network structures is a reaction to classical administrative systems in which politically neutral hierarchies are said to implement government will in an efficient manner. One major force for breaking the machine-like model came from American pluralism, a value respecting participation of various organized groups in the policy process. This pluralistic value has evolved in accordance with political pressures for maintaining "small government" and adding more organized groups in the policy process to protect representative or democratic systems of governance. For instance, the growing reliance on contracting out the provision of government service delivery proposes that government must remain as small as possible and should provide opportunities for others to maintain or improve the quality of public services.

The growing incidence of cross-cutting mandates provides another reason why networks have proliferated in government.^[1] The layers of mandates impose the same rules on many different programs in various agencies, thereby requiring multiple points of perspective to implement any policy in a consistent manner. For instance, environmental rules not only apply to all levels of governments, but also to various public policies, such as environmental protection, energy conservation, and national security. A necessary outcome is the establishment of networks in response to increasing demand for cooperation.

EFFECTIVENESS OF NETWORKS

The debates on network effectiveness are inconclusive. Earlier, networks typically were viewed negatively, especially in the iron triangle where actors in the policy process collaborate for their own benefit, which threatens democratic legitimacy of government performance (e.g., Ref.^[7]). Relatedly, scholars specializing in game theory view actors as rational utility maximizers who attempt to make strategic alliances with partners in networks to induce maximum benefits for their own.^[8] In policy settings where defecting is a dominant choice among actors, consensus requires high coordination costs to maintain network stability. For example, O'Toole^[9] noticed that the lack of coordination in the central government caused the implementation failure of an air pollution control network in Hungary. Scholars in public management also observe many limitations in networks of public-private collaboration. A web of public, private, and quasigovernment organizations diffuses responsibility, thereby

creating a lack of accountability in many of the important public deliveries of services.

Despite their limitations, there is a growing agreement among scholars that networks have become an important feature of modern public policy and administration. This evaluation gets impressive support in recent sources (e.g., Ref.^[10]). Overall, that is to say, networks work well in situations where policy resources are dispersed and where a given policy structure implies high interdependencies. Networks also lubricate the dissemination of information, thereby increasing learning opportunity and allowing ideas to transform into action quickly. Therefore, although networks may require higher costs in policy formulation, there seem to be significantly lower costs in policy implementation.^[11] If not “lower” in the latter case, the costs might be necessary or even unavoidable.

Nevertheless, the effectiveness of networks still remains uncertain, even though such work as that of Provan and Milward^[12] provided important insights about network effectiveness, particularly in comparison with hierarchical arrangements in service implementation. The authors conducted a comparative study of mental health service networks in four U.S. cities to measure effectiveness of service implementation. The researchers pointed out that effectiveness of service implementation must be assessed at the network level for a direct reason: accomplishing a goal—client well-being—depends on a web of complex agencies, each providing multiple services to improve quality of care. Interestingly, they found that the most effective of the four mental health service networks is centralized around a core agency under direct external control within stable system environment.^[13] This finding runs counter to the conventional theory, which suggests that decentralized and highly flexible structural systems can operate more efficiently than traditional hierarchical arrangements. Further, Provan and Milward argued that Simon's^[14] loosely coupled “nearly decomposable” systems may not be efficient in environments such as mental health settings, where networks must be tightly linked to provide effective service to clients.

IMPLICATIONS AND RESEARCH AGENDA

For better or worse, public administrators have increasingly been confronted with network arrangements for program operations. Consequently, Jacksonian simplicity in administration is no longer appropriate, and it could even be disruptive in understanding the modern public policy configurations. Nevertheless, public administrators are not well prepared for new forms of governance, and few have treated networks seriously.^[1] This lack of attention implies limited practical lessons for

public managers who have increasingly been facing network arrangements, recognize them or not.

The emerging networks have significant implications for effective governance, in which the end results of public policies seem to rest on the understanding of complex patterns of interactions among policy actors. For instance, scholars in public management have noticed the emergence of the “hollow state” as one awkward result of establishing service delivery networks where government becomes only one of the many actors in the provision of public service, and perhaps even one with only distant oversight capacities as in the new arrangements about security checkers in airports. It is important that public managers recognize the dynamics of networked interactions, and also that they find ways in which program results are not only mutually beneficial to policy actors, but also meaningfully accountable to the large set of publics.

Some contend that the presence of policy networks implies a weakening of neutral and scientific public administration because of the emphasis on broad political interactions (e.g., Ref.^[7]). Policy networks certainly deny the politics–administration dichotomy and require “issue skilled” public managers who are well informed about the important issues among actors in the network settings.^[4] Specifically, the role of public manager in the network context is to reduce uncertainty underlying network arrangements and to induce cooperative solutions among the actors.

The final implication for present purposes is that conventional theory may not be suitable to network settings and permits little heuristic guidance in practice. Scholars have noticed that the unit of analysis has gradually shifted from the single agency to networked arrangements. Also, there are increasing needs for further scholarly research on networks to explore highly flexible structural arrangements, cause and consequences of networked interactions, and preferences and strategies of the various actors. Given the growing demands for theoretical agenda, game theory may be a practical alternative. In particular, Stoker^[8] offered some important heuristic lessons that can apply to networked interactions of actors who may cooperate but often reluctantly because of lack of knowledge of other actors' true preferences. He used game theory as a way of predicting “interdependent strategic alliance” among actors who have conflicting interests.

Basically, anomalies exist in the course of exploring networked interactions because the multiple values involved tend to be incompatible. Managing networks requires necessary transaction costs to induce cooperation, regardless of the strategies applied—whether coercion, persuasion, or rewards. In this sense, some scholars suggest that trust may be a critical element in inducing cooperation, and reducing transaction costs engendered by bargaining, contracts, and monitoring

(e.g., Ref.^[15]). However, empirical research does not support the connection between trust and cooperation in network settings. This empirical question and many others require continued attention.

REFERENCES

1. O'Toole, L.J. Treating networks seriously: practical and research-based agendas. *Public Adm. Rev.* **1997**, 57 (1), 45–52.
2. Klijn, E.-H. Policy networks and network management: a state of the art. In *Managing Complex Networks: Strategies for the Public Sector*; Kickert, W.J.M., Klijn, E.H., Koppenjan, J.F.M., Eds.; SAGE Publications: London, 1997; 14–34.
3. Hjern, B.; Porter, D.O. Implementation structures: a new unit of administrative analysis. *Organ. Stud.* **1981**, 2 (3), 211–227.
4. Heclo, H. Issue networks and the executive establishment. In *The New American Political System*; American Enterprise Institute: Washington, 1978; 87–124.
5. Laumann, E.O.; Heinz, J.P.; Nelson, R.; Salisbury, R. Organizations in political action: representing interests in national policy making. In *Policy Networks*; Marin, B., Mayntz, R., Eds.; Westview: Boulder, CO, 1991; 63–95.
6. Marin, B.; Mayntz, R. Introduction: studying policy networks. In *Policy Networks*; Marin, B., Mayntz, R., Eds.; Westview: Boulder, CO, 1991; 11–23.
7. Fox, C.J.; Miller, H.T. Policy network. In *The International Encyclopedia of Public Policy and Administration*; Sharfritz, J.M., Ed.; Westview: Boulder, CO, 1998; Vol. 3, 1699–1702.
8. Stoker, R.P. *Reluctant Partner*; University of Pittsburgh Press: Pittsburgh, 1991.
9. O'Toole, L.J. Networking requirements, institutional capacity, and implementation gaps in transitional regimes: the case of acidification policy in Hungary. *J. Eur. Public Policy* **1997**, 4 (1), 1–17.
10. Chisholm, R.F. *Developing Network Organizations: Learning from Practice and Theory*; Addison-Wesley: Reading, MA, 1998.
11. Kenis, P.; Schneider, V. Policy networks and policy analysis: scrutinizing a new analytical toolbox. In *Policy Networks*; Marin, B., Mayntz, R., Eds.; Westview: Boulder, CO, 1991; 25–59.
12. Provan, K.G.; Milward, H.B. A preliminary theory of interorganizational network effectiveness: a comparative study of four community mental health systems. *Adm. Sci. Q.* **1995**, 40 (1), 1–33.
13. Rainey, H.G. *Understanding and Managing Public Organizations*; Jossey-Bass: San Francisco, 1997.
14. Simon, H.A. The organization of complex systems. In *Hierarchy Theory: The Challenge of Complex Systems*; Pattee, H.H., Ed.; George Braziller: New York, 1973; 3–27.
15. Golembiewski, R.T.; McConkie, M.L. The centrality of interpersonal trust in group processes. In *Theories of Group Processes*; Cooper, C.L., Ed.; Wiley: New York, 1975; 131–185.

New Institutional Economics

Brian E. Dollery

School of Economics, University of New England, Armidale, New South Wales, Australia

Joe L. Wallis

Department of Economics, University of Otago, Dunedin, New Zealand

INTRODUCTION

In common with the major earlier institutionalist tradition associated with Thorstein Veblen, Wesley Mitchell, John R. Commons, and Clarence Ayres, new institutional economics (NIE) represents a loose collection of ideas aimed at bringing institutional characteristics back to the core of economic analysis.^[1] Given its somewhat disparate nature, NIE is difficult to define with any degree of precision. Frant^[2] identified three “precursors” of NIE. First, Coase’s^[3] focus on the importance of transactions and the related problems of whether to employ markets or hierarchies to handle transactions, emphasized the costs attached to using the price mechanism. Second, there is the literature on the economics of property rights, deriving from Coase’s^[4] famous paper on how the assignation of property rights influences outcomes in the presence of externalities, which led to the so-called “Coase theorem.” And third, Alchian and Demsetz’s^[5] seminal attempt to apply the property rights paradigm to organizations engaged in productive activity, with the problems inherent in “team production,” like “shirking” and “monitoring,” is considered. Disagreement exists on the major dimensions of NIE. For instance, Rutherford^[1] adopts a fairly broad view of the main elements of NIE by including the economics of property rights,^[6] law and economics,^[7] rent-seeking and distributional coalitions,^[8] agency theory,^[9] transaction cost economics,^[10] game theory in institutional situations,^[11] and the new economic history of Douglas North.^[12] Matthews^[13] developed a fourfold taxonomy of NIE in which institutions are viewed as property rights, as kinds of contracts, as conventions, and as governance structures. By contrast, Boston et al.^[14] constrain their policy-orientated view of NIE to only two strands, namely, agency theory and transaction cost economics.

THE NIE CONCEPT

Unlike the older tradition, NIE scholars have few problems with a priori deductive theorizing. For example,

Furubotn and Richter^[15] have observed that:

The change in analytic approach adopted by the new institutionalists has not resulted from any deliberate attempt to set up a new and distinct type of doctrine in conflict with conventional theory. Rather, the tendency to introduce greater institutional detail into economic models has come about gradually over time because of the recognition that standard neoclassical analysis is overly abstract and incapable of dealing effectively with many current problems of interest to theorists and policymakers.

Although NIE does not contest the methodology of contemporary neoclassical economics, it makes several crucial changes to it. First, following Simon,^[16] NIE recognizes that in the real world, individuals possess an inherently limited capacity to process information and, accordingly, are “boundedly rational” in the sense that their calculations include only immediate and readily assimilated information. Bounded rationality necessarily implies that the complexities of actual economic exchange cannot be fully captured by hierarchical contracts or market mechanisms. This leads to a second difference between NIE and conventional economic analysis. Because bounded rationality prevents the construction of complete contracts between agents and principals, scope exists for opportunistic behavior by economic agents, who can conceal their preferences and actions from contractual partners.^[10] Indeed, it is precisely because of real-world phenomena, like bounded rationality and incomplete contracts, that economic activities have to be conducted in an environment characterized by asymmetric information and costly transactions, and it is these features that lend crucial importance to institutions.

Perhaps even more significant than these methodological issues are the differences between NIE and orthodox neoclassical analysis on the question of what should constitute the appropriate measure(s) to gauge economic efficiency. In common with the conventional approach, normative problems are usually framed in individualistic terms and focus on the efficiency of alternative conceivable institutional arrangements. But, many theorists working in the NIE tradition

have expressed strong reservations about using the allocative efficiency criterion to justify government intervention. For instance, Demsetz^[17] “nirvana fallacy” argument holds that by comparing real-world arrangements against the ideal of allocative efficiency rather than feasible institutional alternatives, policy makers have become far too inclined to prescribe government intervention. Similarly, De Alessi^[18] argued that many of the supposed inefficiencies identified in practical institutional situations by neoclassical economics are actually due to the existence of transaction costs.

Arguments along these lines have led many in the NIE tradition to favor a comparative institutions approach to the question of economic efficiency. The essence of this approach has been summarized by Furubotn and Richter,^[15] as follows:

Granting that the conventional efficiency standard is unsatisfactory, there is reason to seek a different criterion and move toward a concept of economic efficiency that takes fuller account of those real-life constraints that actually limit individual choices. This approach is precisely the one that many new institutionalist writers have adopted. These theorists point out that, in the real world, alternative systems of property rights exist and transaction costs are inevitably greater than zero. Hence, the “generalization” of neoclassical doctrine is advocated. The objective is, of course, to include in any general equilibrium model all of the constraints that decision makers must face in seeking to maximize utility.

The efficiency criterion required by this approach has been specified by De Alessi^[18] in the following terms:

Efficiency is being defined as constrained maximization. Efficiency conditions are seen as the properties of a determinate (equilibrium) solution implied by a given theoretical construct. On this view, a system’s solutions are always efficient if they meet the constraints that characterize it.

Often, notions of efficiency employed in the comparative institutions approach focus on productive or dynamic efficiency rather than on allocative efficiency. For example, North^[12] emphasizes the relative efficiency of institutions with reference to their impact on “social output” or “economic growth” and the ways in which they minimize transaction costs through time. This kind of argument does not indicate how “economic efficiency” defined in this way enhances individual utility or social welfare, and accordingly, is not completely satisfactory.

In general, it seems clear that theorists in the NIE tradition advocate a concept of efficiency that embodies organizational costs, like transaction costs, which

can be used to evaluate various feasible reorganizations of economic activity that could yield social gains net of the costs of reorganization. (See, for example, Refs.^[4,17,19,20].) This intellectual project is not without its critics^[21] and may be simply tautological. This argument has been summarized by Papandreou^[22] as follows:

The traditional notion of optimality in welfare economics has been criticized for comparing real-world outcomes with unattainable outcomes of a transaction-costless world. The alternative notion of “attainable” optimality implicit in models with endogenous transaction costs, suffers in that by offering a causal explanation of economic interaction it precludes other feasible alternatives with which to compare the outcome of the model. By disallowing comparisons of the “real-world” microeconomic model with microeconomic models that have different organizational costs, the set of feasible outcomes collapses to the outcome of the “real-world” model. Accordingly, the outcome that exists is the only possible one.

TRANSACTION COSTS

Despite the fact that Coase^[3] originally specified transaction costs as the “...cost of using the price mechanism” more than 60 years ago, a satisfactory definition of this concept remains problematic. As Allen^[23] has observed, “the literature on transactions costs is replete with papers which use the term and provide examples, but which never pause to define the phrase.”

In general terms, a transaction can be defined as an agreed exchange or transfer of goods and services across technologically separable boundaries. The costs involved in such an agreement and which facilitate such a transfer are collectively known as transaction costs.^[10] In other words, transaction costs are the costs of facilitating economic exchange or, in the language of Kenneth Arrow,^[24] the costs incurred in “running the economic system.” Transaction costs are contrasted with transformation costs (sometimes also called production costs), which are the costs involved in transforming inputs into outputs.

Theorists have adopted two generic ways of specifying transaction costs. First, it has been common to adopt somewhat narrow definitions of transaction costs closely bound with the notion of property rights. For example, Eggertsson^[25] delineated transaction costs as “the costs that arise when individuals exchange property rights to economic assets and enforce their exclusive rights.” Similarly, Barzel^[26] and McManus^[27] emphasized the costs of enforcement and negotiation of property rights. A second and more contemporary

view of the nature of transaction costs focuses on the costs of creating and maintaining the institutions characteristic of modern market economics. Furubotn and Richter^[15] described this approach as "...most easily understood as embracing all those costs that are connected with 1) the creation or change of an institution or organization, and 2) the use of the institution or organization." In the present context, we follow this latter approach.

The theory of transaction costs arose from Coase's^[3] question of why two institutions, the market and the firm, perform the same basic functions, yet continue to coexist. He argued that these two alternative methods of coordinating economic activities exist, because there are transaction costs associated with using the price mechanism. Accordingly, rational economic agents will seek to minimize transaction costs and will use markets or hierarchies, whichever is cheapest. Thus, for example, whether a firm decides to own or lease a particular machine will depend on the transaction costs involved.

The modern theory of transaction costs developed this line of thought and emphasizes the importance of the institutions underpinning exchange relationships, namely, markets, contracts, and firms in private sector settings. Transaction costs theory thus seeks to characterize the properties of transactions in order to determine which institution is optimal in any specific case. Williamson^[10] argued that the critical features of transactions are their uncertainty, the frequency with which they recur, and the extent to which parties to these transactions are obliged to make investments in transaction-specific assets. For example, once parties in an initial competitive market commit to a specialized transaction (involving, say, substantial investments in specific assets, or high "asset specificity"), then this leads an ongoing "small number" bargaining situation between the two parties, often termed bilateral bargaining. Williamson^[28] hypothesized that these kinds of arrangements encourage opportunistic behavior or "self-interest seeking with guile," and he called them "idiosyncratic" transactions, because the benefits of the transaction are dependent on the absence of opportunistic behavior in its execution. Obviously, specialized institutional arrangements will be required to govern idiosyncratic transactions.

Similarly, pervasive uncertainty, or the absence of perfect information, will also have a decisive effect on whether markets, contracts, or firms will be chosen to govern transactions. In general, Williamson^[10] argued that high levels of uncertainty will inhibit the use of long-term contracts in these cases. Analogous arguments also exist with respect to transactional frequency. For instance, in the absence of asset specificity, infrequent transactions will not usually require

contracts and will simply be undertaken through the market mechanism.

In common with other areas of NIE, transaction cost economics has been subjected to severe criticism. The first and most damaging line of attack is based on the proposition that it is not obvious that economic agents should wish to minimize total transaction costs rather than simply the transaction costs borne directly by themselves. Put differently, transaction costs minimization may take the form of a contest over "cost shifting" rather than a search for the optimal institutional arrangements. Second, although an attractive idea in principle, the desegregation of total costs into transformation costs and transaction costs may not always be feasible in practice.^[29] As Lane^[30] inquired: "could one measure the reduction in transaction costs by moving from a system of individual contracts to a hierarchy like a firm or bureau?" The transaction costs theory has also been criticized for the narrowness of its *homo economicus* behavioral assumption, which is asserted to disregard cultural, social, and other dimensions of human conduct.

For these and other reasons, transaction cost theory can at best offer only a partial explanation for real-world differences in economic organization. While it advances our understanding of the benefits and costs of alternative governance structures, it should, nevertheless, be applied with care and in tandem with other NIE models, like agency theory.

AGENCY THEORY

The literature on agency theory is centrally concerned with the ways in which one economic agent (the principal) can design and implement a contractual compensation system to motivate another economic agent (the agent) to act in the principal's pecuniary interests. Accordingly, agency theory is chiefly focused on the problem of economic incentives. By contrast, conventional neoclassical economics, with its reductionist construct of the organization as simply a production function converting inputs into outputs in the most productively efficient manner, assumed away principal-agent problems and the associated question of incentives. In the orthodox tradition, principals were perfectly informed on the nature of the tasks agents were hired to perform, and their subsequent actions could be monitored without cost. Agency theory can thus be viewed as the extension of the neoclassical theory of the firm to embrace "the problems posed by limited information and goal conflict within organizations."^[31] In other words, just as transaction costs theory adds realism to the theory of markets by including the costs attached to exchange relationships,

agency theory adds realism to the conventional “black box” conception of organizations.

A principal–agent relationship comes into being whenever a principal delegates authority to an agent whose behavior has an impact on the principal’s welfare. Numerous principal–agent relationships exist across virtually the entire spectrum of human endeavor. For example, principal–agent relationships include voters and their elected representatives, patients and physicians, employers and employees, shareholders and corporate managers, and so forth. The essence of all of these principal–agent relationships involves a trade-off for the principal. By delegating authority to an agent, the principal economizes on scarce resources by adopting an informed and able agent, but simultaneously takes on the risk that because the interests of the principal and agent will never be identical, the agent may fail to maximize the wealth of the principal. For instance, despite being paid a proportionate commission, real estate agents may wish to sell a property quickly at a lower price than could otherwise be obtained for the owner of the property. Because the interests of principals and agents are likely to diverge in most real-world relationships, agency theory focuses on the costs attached to such relationships and the efforts that principals and agents will take to economize on these costs. In their seminal contribution, Jensen and Meckling^[9] identified three categories of agency costs. First, monitoring costs arise from the resources principals invest in monitoring the behavior of agents and creating incentives for desirable behavior. Second, bonding costs will derive from resources invested by agents to guarantee successful outcomes to their principals. And third, residual loss costs consist of the losses to their well-being suffered by principals due to a divergence between their interests and those of their agents. According to Jensen and Meckling,^[9] total agency costs are thus comprised of the aggregate of monitoring costs, bonding costs, and residual loss costs.

The extent of total agency costs is postulated as dependent on the characteristics of particular principal–agent relationships. For example, in a typical agency relationship, agents almost always possess more information about the task assigned and the relative efficacy of their own performance. Agents often take advantage of this asymmetry of information by engaging in shirking or opportunistic behavior inimical to the interests of principals, and clearly, the greater the extent of this behavior, the higher will be the level of total agency costs.

Although in its initial stages, agency theory was applied exclusively to the firm, and especially, the modern corporation with its characteristic separation of ownership from control quintessentially representative of a principal–agent relationship, it was soon used

more widely once its explanatory powers were recognized. Jensen and Meckling^[9] recognized the generality of agency theory:

The problem of inducing an “agent” to behave as if he [or she] were maximizing the “principal’s” welfare is quite general. It exists in all organizations and in all co-operative efforts—at every level of management in firms, in universities, in mutual companies, in co-operatives, in governmental authorities and bureaus, in unions, and in relationships normally classified as agency relationships such as are common in the performing arts and the market for real estate.

Two specific aspects of the agency problem have received widespread attention in the literature. First, the phenomenon of adverse selection occurs due to informational asymmetries in principal–agent relationships. For example, because unhealthy people are more inclined to purchase health insurance, they will be overrepresented in the pool of insured people. In turn, this will force the price of insurance to rise, inducing many low-risk healthy individuals to forego health insurance. Second, the problem of moral hazard arises when a party to a contractual relationship can affect the outcome of this arrangement to their own advantage. For instance, if a person is fully insured and cannot be accurately monitored by an insurance company, then their behavior may change after the insurance has been purchased to induce larger payments. Adverse selection occurs due to information asymmetries in principal–agent relationships. For example, older people typically experience more difficulties in obtaining health insurance.

Given the problems inherent in principal–agent relationships, like adverse selection and moral hazard, a good deal of agency theory has focused on resolving these problems by specifying the optimal form of contracting (Refs.^[32,33]). Substantial progress has been made in a variety of economic relationships, including labor, land, credit, and product markets, and, more recently, in the public sector.

Agency theory is not without its critics (see, for instance, Ref.^[34]). In common with other areas of NIE, including transaction costs economics, agency theory has been criticized for the restrictive and descriptively unrealistic nature of its *homo economicus* behavioral postulate, which is postulated as neglecting the complexities of human relationships. More specifically, agency theory has been attacked for overlooking issues of authority and power in principal–agent relationships, and in particular, for emphasizing the role of opportunistic behavior by agents, while ignoring opportunism by principals. It has also been argued that agency theory places a heavy reliance on single principal–agent models that ignore questions of

competing principals or of uncertainty surrounding the identity of principals. Moreover, it has been observed that the policy implications that flow from agency theory are unclear, especially in public sector contexts.

REFERENCES

1. Rutherford, M. *Institutions in Economics: The Old and the New Institutionalism*; Cambridge University Press: Cambridge, 1996; 2–3.
2. Frant, H. The new institutional economics: implications for policy analysis. In *Policy Analysis and Economics*; Weimer, D.L., Ed.; Kluwer: London, 1991; 112–113.
3. Coase, R.H. The nature of the firm. *Economica* **1937**, *4*, 386–405, November.
4. Coase, R.H. The problem of social cost. *J. Law Econ.* **1960**, *3* (1), 1–44.
5. Alchian, A.; Demsetz, H. Production, information costs and economic organisation. *Am. Econ. Rev.* **1972**, *62* (4), 777–795.
6. Alchian, A.; Demsetz, H. The property rights paradigm. *J. Econ. Hist.* **1973**, *33* (1), 16–27.
7. Posner, R.A. *Economic Analysis of Law*; Little Brown: Boston, 1977.
8. Olson, M. *The Rise and Decline of Nations*; Yale University Press: New Haven, 1982.
9. Jensen, M.C.; Meckling, W.H. Theory of the firm: managerial behaviour, agency costs, and ownership structure. *J. Financ. Econ.* **1976**, *3* (3), 305–360.
10. Williamson, O.E. Transaction-cost economics: the governance of contractual relations. *J. Law Econ.* **1979**, *22*, 233–261, October.
11. Shubik, M. The general equilibrium model is not complete and not adequate for the reconciliation of micro and macroeconomic theory. *Kyklos* **1975**, *28* (3), 545–573.
12. North, D.C. *Structure and Change in Economic History*; Norton: New York, 1984.
13. Matthews, R.C.O. The economics of institutions and the sources of growth. *Econ. J.* **1986**, *96* (4), 903–918.
14. Boston, J.; Martin, J.; Pallot, J.; Walsh, P. *Public Management: The New Zealand Model*; Oxford University Press: Auckland, 1996.
15. Furubotn, E.G.; Richter, R. The new institutional economics: an assessment. In *New Institutional Economics*; Furubotn, E.G., Ed.; Edward Elgar: London, 1992; 1–32.
16. Simon, H. *Reason in Human Affairs*; Basil Blackwell: Oxford, 1983.
17. Demsetz, H. Information and efficiency: another viewpoint. *J. Law Econ.* **1969**, *82* (4), 713–719.
18. De Alessi, L. Property rights, transactions costs, and x-efficiency: an essay in economic theory. *Am. Econ. Rev.* **1983**, *73* (1), 64–81.
19. Dahlman, C.J. The problem of externality. *J. Law Econ.* **1979**, *22* (1), 141–162.
20. Bromley, D.W. *Economic Interests and Institutions*; Basil Blackwell: Oxford, 1989.
21. Samuels, W.J. Interactions between legal and economic policies. *J. Law Econ.* **1972**, *14* (2), 435–450.
22. Papandreou, A.A. *Externality and Institutions*; Clarendon Press: Oxford, 1994; 273–274.
23. Allen, D. What are transactions costs?. *Res. Law Econ.* **1991**, *14* (3), 1–18.
24. Arrow, K. The Organization of Economic Activity: Issues Pertinent to the Choice of Market Versus Non-Market Allocation. In *Public Expenditure and Policy Analysis*; Haverman, R.H., Margolis, J., Eds.; Markham: Chicago, 1970; 43–51.
25. Eggertsson, T. *Economic Behaviour and Institutions*; Cambridge University Press: Cambridge, 1990; 14 pp.
26. Barzel, Y. Measurement costs and the organisation of markets. *J. Law Econ.* **1982**, *25* (1), 27–48.
27. McManus, J.C. The theory of the multinational firm. In *The Multinational Firm and the National State*; Paquet, G., Ed.; Collier-Macmillan: London, 1972; 63–72.
28. Williamson, O.E. *The Economic Institutions of Capitalism*; Free Press: New York, 1985; 47 pp.
29. Dollery, B.E.; Leong, W.H. Measuring the transaction sector in the Australian economy, 1911–91. *Aust. Econ. Hist. Rev.* **1998**, *38* (3), 207–231.
30. Lane, J.E. *The Public Sector: Concepts, Models and Approaches*; Sage: London, 1995; 229 pp.
31. Levinthal, D. A survey of agency models of organisations. *J. Econ. Behav. Organ.* **1988**, *9* (1), 153–185.
32. Rees, R. The theory of principal and agent: Part I. *Bull. Econ. Res.* **1985**, *37* (2), 1–26.
33. Rees, R. The theory of principal and agent: Part II. *Bull. Econ. Res.* **1985**, *37* (2), 75–95.
34. Davis, G.; Gardner, M. Who signs the contract? Applying agency theory to politicians. In *The State Under Contract*; Boston, J., Ed.; Bridget Williams: Wellington, 1995; 140–159.

New Public Management

Eran Vigoda

School of Political Sciences, University of Haifa, Haifa, Israel

INTRODUCTION

New Public Management represents “*an approach in public administration that employs knowledge and experiences acquired in business management and other disciplines to improve efficiency, effectiveness, and general performance of public services in modern bureaucracies.*” Despite contradictory views about the meaning and implications of this doctrine, there is no doubt that it has become extremely influential in public administration theory and practice since the 1980s.

In the last decade, we have witnessed an ongoing debate between scholars and practitioners of public administration on the best way to revitalize and renew old style bureaucracies. Since the early 1980s, much work has been conducted in public administration theory and practice that claimed to go beyond the conservative approach in the field. Modern public administration has been urged to innovate and reform itself in a style better suited to the twenty-first century and the growing demands for higher efficiency and effectiveness. Drawing on the experience of the business/industrialized/private sector, scholars have suggested taking a more “demanding” attitude toward the dynamics, activity, and productivity of public organizations. Similarly, a significant conceptual change has transformed the “old” style of public administration into the “new” approach of public management. This “liberalization” of public administration is recognized today as the New Public Management (NPM) doctrine.

What is the meaning of NPM? What are its roots, and in what way is it actually a *new* arena in the study of the public sector? Has it achieved enough success in recent years to justify further inquiry into its principles and promises? These questions, as well as many others, have received considerable attention during recent years and, in fact, a large number of varying opinions have emerged. Proponents of this approach have provided detailed explanations and a variety of examples that demonstrate the enormous positive impact of the approach. Those who critique NPM counter with the argument that it has nothing new or notable to contribute to the field. They, thus, tend to treat it as a seasonal fashion that will eventually fade in face of other “new” doctrines. Therefore, advocates of NPM continuously face critics who are unconvinced

as to the merits of NPM as an administrative philosophy.

THEORETICAL FOUNDATIONS

To better resolve these conceptual and practical conflicts regarding the contribution of NPM to public administration, one must become familiar with several fundamental concepts. The first and probably the most basic premise of NPM emerges from the distinction between two proximate terms or fields of research—*administration* and *management*. Since the late 1880s, the monopoly on the term “administration” has been in the hands of political scientists. Scholars like Goodnow and Wilson perceived public administration as a separate and unique discipline that should consist of independent theory, practical skills, and methods. Yet, the term “management” referred to a more general arena, used by all social scientists and mainly by those who practice and advance theories in organizational psychology and business studies. Consequently, conservative administration science tends to analyze the operation of large bureaucratic systems, as well as other governmental processes aimed at policy implementation. Management, on the other hand, refers to the general practice of empowering people and groups in various social environments and the handling of manifold organizational resources to maximize efficiency and effectiveness in the process of producing goods or services.

A second premise of NPM derives from the nature and values of democratic nations. Citizens of modern democracies act as voters and delegate power to politicians and administrators to do what is best for people and societies. However, representative democracy leaves its fingerprints on the actions and operation of bureaucracies. For many years, bureaucracies worked in a manner far removed from citizens and with a lack of sensitivity to the growing needs and demands of heterogeneous populations. As suggested by Rainey,^[1] the 1960s and the 1970s were characterized by the initiation of unsuccessful public policies in Europe and in America. At least some of these unsuccessful experiences were due to lack of reliable analyses of the needs and demands of the public, while other failures were caused by incorrect assessments regarding

the power of bureaucracies. Over the years, efforts by governments to create extensive changes in education, welfare systems, health programs, internal security, and crime control were widely criticized for being ineffective and unproductive and for misusing public funds. Responsiveness to the real needs and demands of citizens was paltry. The crisis in practical public policy implementation, together with the increased cynicism of citizens toward government and public administration systems, generated rich scholarly activity aimed at creating useful alternatives for improved policy in various social fields, as well as in the administrative processes in general. Voters expressed their dissatisfaction with elected officials and, hand in hand with the academic community, called for extensive reforms in government. This call produced a large number of working papers, articles, and books that proposed extensive administrative changes. One of the most inspiring works, Osborne and Gaebler's *Reinventing Government*,^[2] is frequently mentioned as the unofficial starting point of such reforms, later known as NPM.

As time went on, a growing number of political scientists perceived public administration as an *old* and declining discipline. It was unable to provide the public with adequate practical answers to its demands and moreover, left the theoreticians with epidemic social dilemmas awaiting study. Evidence for this shift in attitude appears in the transformation of many schools of public administration into schools of public management that took place during the 1980s and 1990s. Looking for alternative ideas, management theory was proposed as a source for a new and refreshing perspective. It was suggested that public management instead of public administration could contribute to a new understanding of how to run the government more efficiently and thereby overcome some of its pandemic problems.

CORE DEFINITIONS

Based on the growing alienation of citizens, the ineffective performance of bureaucracy, and the growing demands for a real change in public policy and activity, NPM flourished. But what is the essence of NPM and the best way to define it? During the last two decades, many definitions have been suggested. In the early 1980s, Garson and Overman^[3] defined it as *an interdisciplinary study of the generic aspects of administration... a blend of the planning, organizing, and controlling functions of management with the management of human, financial, physical, information and political resources*.

Hood^[4] identified seven doctrinal components of NPM: 1) "hands on" professional management in the

public sector; 2) explicit standards and measures of performance, which were later defined as PIs (Performance Indicators); 3) greater emphasis on output control; 4) a shift to the disaggregation of unit; 5) a shift to greater competition; 6) a stress on private sector styles of management practice; and 7) a stress on greater discipline and parsimony in resource use. This definition implies that NPM relies heavily on the theory of the marketplace and on a business-like culture in public organizations. Other definitions were suggested in the 1990s and drew on the extensive writing in the field. For example, Hays and Kearney^[5] found that most of the studies on NPM had mentioned five core principles of NPM and thus concluded that they represent the most important philosophy of the discipline: 1) downsizing—reducing the size and scope of government; 2) managerialism—using business protocols in government; 3) decentralization—moving decision making closer to the service recipients; 4) debureaucratization—restructuring government to emphasize results rather than processes; and 5) privatization—directing the allocation of governmental goods and services to outside firms.^[6] All of these principles are mutually related, relying heavily on the theory of the private sector and on business philosophy but aimed at minimizing the size and scope of governmental activities. Integrated with ideas rooted in political economy, they have now been applied to public sector institutions.

Hence, governments that are far from being simple businesses have been encouraged to manage and run themselves like businesses. An integrative definition for NPM that relies on the previous works would thus argue that NPM represents "an approach in public administration that employs knowledge and experiences acquired in business management and other disciplines to improve efficiency, effectiveness, and general performance of public services in modern bureaucracies."

WHAT MAKES PUBLIC MANAGEMENT NEW?

New Public Management is growing in popularity in North America and across the world, and many governments have adopted ideas and recommendations that have proven beneficial, thereby arguing for the continued implementation of this strategy. A consensus exists today that NPM has become extremely popular in theory and in practice in public arenas. Relying on an extensive survey of public management research in America, Garson and Overman^[3] argue that this increasing popularity was due to the more aggressive connotation of the term "management" in comparison with "administration."

Yet, is NPM really a new doctrine in public administration or just another variation on old concepts and ideas? There is no doubt that the increased impact of new ideas and methods from the field of public management into that of administrative science is essential and natural. It reflects a special focus of modern public administration that must not be ignored. It may also be viewed as a major segment of the broader field of public administration since it focuses on the profession and on public managers as practitioners of that profession. Thus, public management had indeed some new elements of vital and innovative thinking. Furthermore, it emphasizes well-accepted managerial tools, techniques, knowledge, and skills that can be used to turn ideas and policy into successful programs of action.

Other scholars delineate (e.g., Ref.^[7]) six differences between public administration and public management that turn the latter into a new field of study and practice: 1) the inclusion of general management functions such as planning, organizing, control, and evaluation in lieu of simple discussion about social values and the conflicts of bureaucracy and democracy; 2) an instrumental orientation favoring the use of the criteria of economy and efficiency in lieu of equity, responsiveness, or political salience; 3) a pragmatic focus on mid-level managers in lieu of the perspective of political or policy elites; 4) a tendency to consider management as generic, aimed at minimizing the differences between the public and private sectors in lieu of accentuating them; 5) a singular focus on the organization where external relations are treated in the same rational manner as internal operations in lieu of a focus on laws, institutions, and political bureaucratic processes; and 6) a strong philosophical link with the scientific management tradition in lieu of close ties to political science or sociology.

While the emergence of NPM is frequently related to the increasing impact of positivist behavioral science on the study of politics and government (e.g., Ref.^[7]), the practical aspect of this process should also be considered. Practical public managers, as well as political scientists, refer to the difficulties in policy making and policy implementation that faced many Western societies in Europe, America, and elsewhere during the 1970s. These practical difficulties are viewed today as an important trigger for the development of NPM. Reviewing two recent books on NPM,^[8,9] Khademanian^[10] argues that American and Westminster advocates of the field find common ground in explaining why such reforms were necessary. The problem of an inflexible bureaucracy that often could not respond efficiently and promptly to the public needs contradicted basic democratic principles and values in these countries. Elegantly, Peter Aucoin summarizes a “trinity” of broadly based challenges with which Western

democracies have coped and with which they will probably continue to struggle in the future, partly through management reform: 1) growing demands for restraint in public sector spending; 2) increasing cynicism regarding government bureaucracies’ responsiveness to citizen concerns and political authority and dissatisfaction with program effectiveness; and 3) a growing international, market-driven economy that does not defer to domestic policy efforts. It seems that these challenges have led many Western governments in America, Britain, New Zealand, Canada, and elsewhere to the recognition that firm reforms and changes in public service should be made.

CRITIQUE OF NPM

Certainly, the NPM approach suggests a different type of interaction between citizens and rulers in democracies. However, the roots of such interactions can be found a century ago. For example, Weikert^[6] asserted that “the ideas behind NPM are not new” and that “NPM builds on a long history of using business practices in government and reflects a resurgence of old ideas about the form and functions of government.” During the first years of the twentieth century, reformers and business leaders demanded greater accountability in local government, and many politicians, as well as public officials, turned to business principles to improve governmental activities, invigorate performance, and reduce corruption. However, the vision of NPM is also far different from the old business-guided governance in that it looks to decrease government size and minimize its involvement in citizens’ lives. As is evident from the above principles, NPM advocates that governments and public administrative bodies view citizens as clients/customers of the public sector, while governments and the public sector are perceived as managers of large bureaucracies. According to this outlook,^[3,8,11] the state and its bureaucratic subsystems are equivalent to a large private organization operating in an economic environment of supply and demand. In this spirit, a major goal of government is to satisfy the needs or demands of citizens, namely to show higher responsiveness to the public as clients. In line with this, it is obvious that modern states must rely more on private and third-sector institutions and less on government to satisfy the societal needs of heterogeneous populations. The goal of satisfying the needs of citizens became central to the NPM philosophy.

Nevertheless, NPM may be criticized for not doing enough to encourage and incorporate the idea of collaboration or partnership between citizens and the public sector and for failing to apply these themes to modern managerial thinking.^[12] Unlike traditional public administration, the NPM movement focuses

on citizens as sophisticated clients in complex environments. The principles of NPM are compatible with theories of political economy such as regulative policy by governments or the policy of transferring responsibilities from the state sector to the private and third sectors. These ideas, and the governmental policies deriving from them, frequently challenge various social democratic principles, norms, and values in Britain, America, and many other Western democracies. Public authorities were urged to treat the public well, not only because of their presumed administrative responsibility for quality in action but also because of their obligation to marketplace rules and to economic demands and above all because of their fear of losing clients in an increasingly competitive businesslike arena. In fact, while NPM is an improvement over more classic views of public administration that saw citizens as subjects or voters, it does not go far enough in fostering the idea of vital collaboration between citizens, governments, and the public sector, which is in the essence of democratic civil society.^[13]

In line with this, “neo-managerialism”^[14] notes an additional obstacle to productive partnership that must also be recognized and surmounted. According to Terry, neo-managerialism supports the idea that administrative leaders should assume the role of public entrepreneurs. However, “public entrepreneurs of the neo-managerialist persuasion are oblivious to other values highly prized in the U.S. constitutional democracy. Values such as fairness, justice, representation, or participation are not on the radar screen (and) this is indeed, troublesome.” In many respects neo/new managerialism and NPM encourage passivity among the citizenry. They impart to citizens the power of *exit* (which indeed was virtually unavailable in the past), but at the same time, they discourage use of the original power of *voice* by citizens who may have much to contribute to their communities.^[12] Exit is an economic choice, while voice is more of a political selection by individuals in and around organizational systems. Exit is also classified as a generally destructive behavior, while voice is a productive one. According to this argument, NPM restricts and discourages the productive political voices of the people.

Hence, recent developments in the study of NPM have focused on the responsibilities of governments and public agencies in their interaction with citizens but similarly have paid far less attention to the active roles of citizens and to their obligations to the community. Most of the up-to-date NPM literature favors massive socialization of business management practices in the public sector to provide governments with better tools for policy implementation. But on the other hand, these orientations and practices have not yet been integrated with another core-construct of healthy democracies—genuine collaboration and

partnership with citizens based on equal opportunities for participation and massive involvement in running public life more effectively. This lack of emphasis on the idea of partnership and collaboration, in favor of good responsive management, may be deemed a flaw in contemporary NPM theory.

CONCLUSION: THE ONGOING PURSUIT OF “NEW” AND “NEWER” INITIATIVES IN PUBLIC ADMINISTRATION

Scholars agree today that at least some of the accumulated wisdom of the private sector in developed countries is transferable to the public sector. In an attempt to “liberate” the public sector from its old conservative image and moribund practice, NPM was advanced as a relevant and promising alternative. Thus, NPM literature has tried to recognize and define new criteria that may help in determining the extent to which public agencies succeed in meeting the growing needs of the public. New Public Management has continuously advocated the implementation of specific performance indicators used in private organizations to create a performance-based culture with matching compensatory strategies. It has recommended that these indicators be applied in the public sector since they can function as milestones by which to better gauge the efficiency and effectiveness of public agencies.

Moreover, citizens’ awareness of the performance of public services was suggested as a core element of NPM since it can increase the political pressure placed on elected and appointed public servants, thereby enhancing both managerial and financial efficiency in the public sector. Scholars who advocate NPM compare this process of public accountability to stakeholders/citizens to the role adopted by financial reporting in the private/corporate sector. As in the private sector, increasing external outcomes can have meaningful impact on internal control mechanisms, as managers and public servants become more sensitive to their duties and more committed to serving citizens as customers.

In view of the above and looking toward the future, Lynn^[15] suggested that NPM of the late 1990s has three constructive legacies for the field of public administration and for democratic theory and practice: 1) a stronger emphasis on performance-motivated administration and an inclusion in the administrative canon of performance-oriented institutional arrangements, structural forms, and managerial doctrines fitted to a particular context—in other words, advances in the state of the public management art; 2) an international dialog on and a stronger comparative dimension to the study of state design and administrative reform; and 3) the integrated use of economic, sociological, social-psychological, and other advanced conceptual models and heuristics

in the study of public institutions and management, with the potential to strengthen the field's scholarship and the possibilities for theory-grounded practice.

It seems that the ongoing debate about the net contribution and added value of NPM to the study and practice of public administration will continue. Whether it is a "new" approach, a partly new doctrine, or an old lady with a new hat, it is obvious that NPM is responsible for some of the meaningful transformations witnessed by modern societies in the last few decades. It is also likely the "newer" doctrines and methods will evolve along with bureaucratic, political, technological, and cultural developments. New Public Management has made its mark, but this is only one link in an endless chain of scientific advancement.

REFERENCES

1. Rainey, H. Public management: recent development and current prospects. In *Public Administration: The State of the Discipline*; Lynn, N.B., Wildavsky, A., Eds.; Chatham House: Jersey, 1990; 157–184.
2. Osborne, D.; Gaebler, T. *Reinventing Government*; Plume: New-York, 1992.
3. Garson, G.D.; Overman, E.S. *Public Management Research in the United States*; Prager: New York, 1983; 275, 278.
4. Hood, C. A public management for all seasons? *Public Adm.* **1991**, *69*, 3–19.
5. Hays, S.W.; Kearney, R.C. Riding the crest of a wave: the national performance review and public management reform. *Int. J. Public Adm.* **1997**, *20*, 11–40.
6. Weikart, L.A. The Giuliani administration and the new public management in New York City. *Urban Aff. Rev.* **2001**, *36*, 359–381.
7. Lynn, L.E. *Public Management*; Chatham House Publishers: New Jersey, 1996; 5–6, 38–39.
8. Aucoin, P. *The New Public Management: Canada in Comparative Perspective*; IRPP, Ashgate Publishing Company: Montreal, Quebec, 1995.
9. Boston; Martin, J.; Pallot; Walsh *Public Management: The New Zealand Model*; Oxford University Press, 1996.
10. Khademanian, A.M. What do we want public managers to be? Comparing reforms. *Public Adm. Rev.* **1998**, *58*, 269–273.
11. Pollitt, C.; Bouckaert, G. *Public Management Reform*; Oxford University Press: Oxford, 2000.
12. Vigoda, E.; Golembiewski, R.T. Citizenship behavior and the spirit of new managerialism: a theoretical framework and challenge for governance. *Am. Rev. Public Adm.* **2001**, *31* (3), 273–295.
13. Vigoda, E. From responsiveness to collaboration: governance, citizens, and the next generation of public administration. *Public Adm. Rev.* **2002**, *62* (5), 527–540.
14. Terry, L.D. Administrative leadership, neo-managerialism, and the public management movement. *Public Adm. Rev.* **1998**, *58*, 194–200.
15. Lynn, L.E. The new public management: how to transform a theme into a legacy. *Public Adm. Rev.* **1998**, *58*, 231–237.

New Public Service

Wouter Vandenabeele

Public Management Institute, Catholic University Leuven, Leuven, Belgium

Annie Hondeghem

Katholieke Universiteit Leuven, Leuven, Belgium

INTRODUCTION

The concept “new public service” has multiple meanings in the field of public administration. Patricia Ingraham and David Rosenbloom^[1] use it when referring to the new public administration movement. Paul Light^[2] uses the expression in a more empirical manner when describing the characteristics of a new generation of public servants at the beginning of the 21st century. However, none of those have received as much resounding responses in the international public administration community as the movement built on work in democratic citizenship, community and civil society, and organizational humanism and discourse theory. The remainder of this entry deals with this last, and by far most influential, meaning of new public service developed by Robert and Janet Denhardt.^[3] Conceived to counterbalance earlier movements in public administration, new public service in its initial description is contrasted to both “classical” public administration (hereafter classical PA) and New Public Management. Next to this description, its implication on administrative ethics is assessed. Finally, evidence of such a new public service ethic is discussed.

COUNTERBALANCING NEW PUBLIC MANAGEMENT

The classic approach to public administration has increasingly come under attack. The politically neutral and hierarchical bureaucracy associated with Woodrow Wilson^[4] and Max Weber,^[5] although it should get fair credits for its achievements, is often solely associated with bureaucratism and inefficiency. New Public Management tried to remedy these critiques not only by introducing private sector techniques into public administration, but also by adopting private sector values of cost effectiveness and entrepreneurial spirit as a normative framework. Against these new entrepreneurial principles, classical PA with its

negative image was bound to be judged bleakly. However, New Public Management was not entirely uncontested either, as numerous authors have criticized it.^[6] The Denhardts tried to bundle these critiques and use them as a starting point for an alternative normative framework to lay against the deficiencies of New Public Management principles: new public service. They define new public service as “a set of ideas about the role of public administration in the governance system that places public service, democratic governance and civic engagement at the center.”^[7]

SEVEN PRINCIPLES OF NEW PUBLIC SERVICE

By stating new public service in terms of normative principles, the Denhardts hope to achieve a more fair comparison of the traditional public administration values with New Public Management, a comparison they felt was not granted to classical PA. They state seven principles and contrast each of them with both classical PA and New Public Management.

1. *Serve, rather than steer.* Classical PA’s duty was to “row” society. Top-down leadership should provide the necessary services for society to survive. In New Public Management, leadership had to “steer” society. This meant setting policies, leaving implementation to agents (governmental or nongovernmental) and evaluate afterwards. Leadership was about creating incentives for agents to implement policies. Leadership in new public service is based on values and it is shared throughout the organization and with the community. Leaders must share power and lead with passion, commitment, and integrity, thus empowering citizenship.
2. *Seek the public interest.* In classical PA, no administrative responsibility to protect the public interest was necessary, as the public interest was determined by elected officials. In New Public Management, the public interest was

realized by the invisible hand of the free market, as it was conceived as the aggregation of individual or self-interest. New public service wants administrators to actively create a shared notion of public interest, consistent with democratic norms.

3. *Think strategically, act democratically.* The implementation of policies should not occur in a top-down manner, as in classical PA, or in a private or a utilitarian coproduction between individuals and government, as in New Public Management. Rather, the whole process should be a collective effort (for example, a network) with the community. It should be an interaction to create a lasting responsiveness of the government towards the community.
4. *Serve citizens, not customers.* In classical PA, people are considered clients or subjects. Therefore, little attention was given to the quality of the services rendered by government. New Public Management changed this drastically. According to public choice theory, people became customers. The quality of services delivered to these customers was consequently the most important aspect of governmental action. Although new public service recognizes the importance of quality of services, people are not mere customers. As citizens they have multiple roles, and they should express this in developing a concern for their community. Government should enable and support them in this community building, not just control or provide customer services.
5. *Value citizenship and public service above entrepreneurship.* The policy-role of the administrator in classical PA is kept deliberately limited. Whenever he or she enters this process, various “cloaks” or reasons are mentioned (for example, expertise or discretion) to cover the value-laden influence on the process. New Public Management expects administrators to act as entrepreneurs. Public servants are urged to develop policies and handle public funding as if it was their own money to achieve their targets. New public service thinks of administrators as “stewards” of public resources, rather than owners. They thus have a larger responsibility to their community. By assuming the role of “professional” citizen, public servants must maintain democratic values in administration and see to the responsiveness of government towards the community.
6. *Value people, not just productivity.* In classical PA, an organization was based on hierarchy and control, as this was the only way to make

people achieve the desired results. New Public Management based itself on public choice and principal agent theory to declare economic self-interest as the sole source of motivation. Therefore, incentives to satisfy this self-interest were used to motivate people. According to new public service, both approaches present only a very narrow view. Organizations are more likely to persist and deliver results if they operate through processes of collaboration and respect for people. The shared values of an organizational culture are very important in this respect.

7. *Accountability is not simple.* In classical PA, accountability was mainly hierarchical. This accountability scheme was grounded on the view of a monolithic bureaucracy, lead by a politically accountable government. New Public Management made entrepreneurial administrators accountable to values of efficiency and cost effectiveness, as it thought of the market as the main governance structure. According to new public service, the governance structure is much more complex. Public administrators are called upon to be responsive to various institutions and values within this structure. It is the role of public servants to explain and disseminate these governance complexities to the public and broker the different interests.

FOUNDATIONS OF NEW PUBLIC SERVICE

The Denhardtts ground the new public service framework on different foundations. The origins of the framework are certainly not new, as some of these date back to earlier movements in the science of public administration (e.g., new public administration), or even go back to the cradle of western democracies. Combining these different foundations into a set of norms provides an enormous intellectual heritage to structure governance.

At the foundation are theories of democratic citizenship. This citizenship is not defined from a legal point of view but as the individual's capacity to influence the political system. It is not only based on self-interest because an important element is public spirit, which enables people to look beyond their parochial self-interest. This public spirit can be nourished by participation, deliberation, and promotion of justice. The presence of a public spirit enables citizens to run government (that is what public administration is all about) in a democratic manner. Only then is government responsive to the people as citizens and not as mere clients, customers, or subjects.

A second element on which new public service is grounded are communities and civil society. Western (and especially American) democracy is based on a tradition of citizens engaged in all sorts of associations. This tradition should be reflected in public administration for two interconnected reasons. Public administration can build further on the available stock of social capital building even stronger networks in order to solve public problems. And it should also enable public administration to help to create and support communities and reinforce social capital.

Third, new public service is based on organizational humanism. Following the critique of hierarchy, control and positivism in classical PA, new approaches to management were adopted. According to organizational humanism, organizations should enhance human development rather than inhibiting it by hierarchy and control. The Denhardtts substantiate this by referring to the new public administration movement, which tried to offer a value-based alternative for classical PA.

Finally, new public service substantiates itself on postmodern discourse theory. This is a critique on separating the reality of organization from values, as propagated by the positivist approach to administration. According to postmodernism, this separation prevents a full understanding of the meaning of realities in public administration. This understanding can only be achieved through open communication. As public problems will be more easily solved through discourse and dialogue, government should be exemplary of open communication.

THE ETHICS OF NEW PUBLIC SERVICE

As a normative framework for a governance structure, new public service is inevitably linked with ethics. Moreover, these ethics are an important feature within the framework, as it explicitly refers to ethical behavior of administrators as a factor for success. The comparison of the new public service ethics with its predecessors, linked with classical PA and New Public Management, unveils some striking differences.

In classical PA, the prevailing system of ethics is the ethic of authoritative command. According to academic moral philosophy, the ethics of authoritative command are a consequentialist system.^[8] It has a certain definition of the good as a foundational principle and a theory of obligation by which an act is moral if it leads to the realization of that good. This good has to be politically determined outside the administration. The role of bureaucrats is limited to obedience to the representatives and the rules they make, following the tradition of the policy-administration dichotomy.

The ethics of New Public Management, although its manifestation could not be more different from classical PA, go along similar consequentialist lines. But instead of focusing on legal accountability of the administrators, attention is shifted towards a good that is determined by market forces instead of democratic procedures.^[9]

A major problem with the ethic of authoritative command is its lack of fit with reality. Governance has become such a complex matter that administrators cannot just follow rules anymore, they are bound to exercise some level of discretion. Therefore, the ethics of authoritative command refer to a situation that is to some extent fictional.

The ethics of New Public Management pose an even bigger problem. As it turns to market forces for determining the good, it threatens the very foundation of any democratic governance structure.^[10]

The ethics of new public service deviate from this approach when referring to organizational humanism, communitarianism, citizenship, and discourse theory.

Communitarianism reaches back to Aristotle's notion of man as a social/political animal and therefore man gets his identity from his community. Communities cultivate certain virtues to benefit the community and which man has the duty to realize. These virtues are at the center of communitarian ethics, but their interpretation is closely connected with the nature of a specific community.

The ethics of democratic citizenship lean very close to communitarianism. Citizens have to look beyond self-interest to a larger public interest of the community they belong to and whose fate is at stake. In fact, some consider the ethics of citizenship as an element of the ethics of communitarianism as it leads to citizens defining themselves in terms of their community.^[8]

Discourse theory states that not only man is a social construction. Reality as a whole finds its origins in social interaction. Only conversation can provide a clear understanding of reality and produce healthy individuals. "Keeping the conversation going" is therefore the ethical thing to do. As communities are obviously an important source of interactions, this also leans very close to communitarianism.

Although the ethics of new public service solves some problems, classical PA and New Public Management were struggling with, it has problems of its own. The reliance on benign communities from which virtues can be derived is an important weakness of the system. A community's shared heritage can be oriented towards oppression or destruction. After all, there is no empirical evidence that communities are intrinsically benign. Ancient Sparta, for example, was oriented towards war and oppression of the surrounding communities and this was reflected in the virtues it advocated. Also, communitarian tradition is less

outspoken in continental Europe. Historically, the religious pillars and the welfare state overcame the needs for other associations. With the decline of religious pillarization, it remains a question whether there is sufficient breeding ground to substantiate a system of communitarian ethics. Next to this, communitarianism carries the seed of totalitarianism. Enveloping every aspect of life in a community, excess application leaves no room for eccentric or creative behavior anywhere in society. At best, this can be unbearably boring. Worse would be that public administration reaches totalitarian dimensions and that it advocates a policy of elimination of all nonconformist individuals, like in George Orwell's "1984." Another critique is that some consider discourse theory ethics not viable in an era dominated by mass communication. Preponderance of this one-way method of communication prevents a discourse to fully develop. Finally, a general problem of new public service ethics is the lurking danger of relativism, because it does not have an overarching ethical principle. With moral values that are relative to community or discourse, anything can be considered ethical (or nonethical). In such an excessive application, discourse theory ethics therefore offers little guidance to public administration practitioners.^[11]

EVIDENCE OF NEW PUBLIC SERVICE ETHICS

As in seemingly every system of administrative ethics, new public service ethics has its inherent problems. These inherent problems can only be overcome by a belief in the rightness of the system. To avoid fundamentalism, however, research results can moderate excess application of an ethic. Put the other way around, these results can demonstrate the relevance of such ethics in particular fields of public administration.

In the case of a new public service ethics, a lot of value-based public administration research has been reported. Much of it is based on the public service motivation concept by Perry.^[12] This is a kind of value-based motivation, covering four dimensions: attraction to public policy making, compassion, commitment to the public interest and civic duty, and self-sacrifice. Behavior stemming from this motivation is an empirical expression of new public service ethical values. Brewer and Coleman^[13] link public service motivation to whistle blowing, resulting in high performance and high levels of job commitment. Naff and Crum^[14] find that public service motivated public servants are more satisfied and have higher performance ratings. Therefore, Rainey and Steinbauer^[15] suggest that public service motivation is critical in delivering public services. But also outside the narrow field of public administration, new public service ethics is visible. Brewer^[16] finds that because of their public

service motivation, public servants are more likely to contribute to social capital than others are.

CONCLUSIONS

The Denhardtts conceive new public service as an alternative normative framework on which to ground public administration and contrast it with both classical PA and New Public Management. The new public service rests on an eclectic set of foundations of democratic citizenship, communitarianism, and postmodern discourse theory, which has its implication in the ethical sphere. Although new public service ethics is able to solve some ethical problems of both classical PA and New Public Management, it has its inherent problems that limit its applicability. However, this need not to be insurmountable, as abundant empirical evidence of behavior in line with new public service ethics demonstrates. Whenever applied wisely and prudently, new public service and its associated ethical framework can be a valuable element in public administration.

REFERENCES

1. Ingraham, P.W.; Rosenbloom, D.H. The new public personnel and the new public service. *Public Adm. Rev.* **1989**, *49* (2), 639–656.
2. Light, P.C. *The New Public Service*; Brookings Institution Press: Washington, DC, 1999.
3. Denhardt, R.B.; Denhardt, J.V. The new public service: serving rather than steering. *Public Adm. Rev.* **2000**, *60* (6), 549–559.
4. Wilson, W. The study of administration. *Polit. Sci. Q.* **1887**, *2*, 197–222.
5. Weber, M. *Wirtschaft und Gesellschaft; Grundriss der Sozialökonomik*; Mohr Siebeck: Tübingen, 1922.
6. Denhardt, J.V.; Denhardt, R.B. *The New Public Service: Serving, Not Steering*; M. E. Sharpe: Armonk, New York, 2003.
7. Denhardt, J.V.; Denhardt, R.B. *The New Public Service: Serving, Not Steering*; M. E. Sharpe: Armonk, New York, 2003; 9.
8. Fox, C.J. The use of philosophy in administrative ethics. In *Handbook of Administrative Ethics; Public Administration and Public Policy*; Cooper, T.L., Ed.; Marcel Dekker: New York, 2001; 86, 105–130.
9. Frederickson, H.G. Comparing the reinventing government movement with new the public administration. *Public Adm. Rev.* **1996**, *56* (3), 263–270.
10. Box, R.C.; Marshall, G.S.; Reed, B.J.; Reed, C.M. New public management and substantive democracy. *Public Adm. Rev.* **2001**, *61* (5), 608–619.
11. Lilla, M.T. Ethos, ethics and public service. *Public Interest* **1981**, *63*, 3–17.
12. Perry, J.L. Measuring public service motivation: an assessment of construct reliability and validity. *J. Public Adm. Res. Theory* **1996**, *6* (1), 5–23.

13. Brewer, G.A.; Selden-Coleman, S. Whistle blowers in the federal service: new evidence of the public service ethic. *J. Public Adm. Res. Theory* **1998**, *8* (3), 413–439.
14. Naff, C.C.; Crum, J. Working for America: does public service motivation make a difference. *Rev. Public Person. Adm.* **1999**, *19* (4), 5–16.
15. Rainey, H.G.; Steinbauer, P. Galloping elephants: developing elements of a theory of effective government organizations. *J. Public Adm. Res. Theory* **1999**, *9* (1), 1–32.
16. Brewer, G.A. Building social capital: civic attitudes and behavior of public servants. *J. Public Adm. Res. Theory* **2003**, *13* (1), 5–26.

BIBLIOGRAPHY

1. Frederickson, H.G. *The Spirit of Public Administration*; Jossey-Bass: San Francisco, 1997.
2. Fox, C.J.; Miller, H.T. *Postmodern Public Administration: Toward Discourse*; Sage: Thousand Oaks, 1996.
3. Institute for Communitarian Policy studies, George Washington University. <http://www.gwu.edu/~icps/index.html> (accessed May 2004).
4. The Public Administration Theory Network. <http://www.pat-net.org> (accessed May 2004).

New Sciences Approach and Public Administration

Göktuğ Morçöl

School of Public Affairs, Pennsylvania State University at Harrisburg,
Middletown, Pennsylvania, U.S.A.

INTRODUCTION

“New Sciences” is a term used to refer to a group of natural science theories that have implications for the social sciences in general, and public policy and administration in particular. The philosophy and methods of the new sciences are evolving, and there is no definitive list of theories that are commonly considered as the new sciences, but *quantum mechanics*, *chaos and complexity theories*, *autopoiesis theory*, and *fuzzy logic* are the ones from which implications have been drawn for the social sciences. These are the theories that will be discussed in this article.

New science theories challenge the deterministic and objectivist assumptions of the classical (Newtonian) science, which is the primary source of the explicit and implicit ontological and epistemological assumptions of many natural and social scientific theories. Newtonian scientific thinking, which has been predominant since the seventeenth century, is based on the view that physical reality is fully determined and can be understood objectively and in terms of mechanistic principles (i.e., by reducing it into its parts and analyzing the relations between them). Newtonian thinking has been influential in the social sciences. For example, some theorists point out that our organizations are designed from Newtonian images of the universe: We assume that our methods can help us perceive the world objectively, separate organizations into parts for management purposes, attempt to influence individual behaviors by exerting direct force on them, and make plans to ensure predictability.^[1] Also, arguably, the policy analytic images of mechanical and linear relationships between policy inputs and their outcomes are Newtonian. Presumed mechanical and linear relationships between phenomena create a predilection for using quantitative methods in social inquiry. Some social scientists equate the use of quantitative methods with scientificity and assume that quantitative forms of knowledge are inherently superior to qualitative forms of knowledge. New science theories question and challenge these deterministic, objectivist, linear, and quantificationist assumptions and offer epistemological and methodological alternatives.

QUANTUM MECHANICS

Quantum mechanics is about subatomic phenomena, but it has broader ontological and epistemological implications. There are different, and sometimes conflicting, interpretations of quantum phenomena, but the term quantum mechanics is most often used to refer to the Copenhagen interpretation, whose main principles were formulated by Niels Bohr, Werner Heisenberg, and others in the early twentieth century.

The central insight of the Copenhagen interpretation is that all forms of radiation and all subatomic “particles” have, in fact, complementary characteristics: They are both wave and particles. This is called the *complementarity principle*. The *uncertainty principle*, which states that the position and momentum of a particle cannot be measured accurately at the same time, is a special case of the complementarity principle. (Position is a particle property, and momentum is a wave property; the more we know about the wave aspect of subatomic reality, the less we know about its particle aspect.) The complementarity and uncertainty principles contradict the Newtonian ontological assumption that the ultimate units of matter (particles) exist as discrete objects and in precisely definable locations in space and move around in precisely predictable trajectories.

The complementarity and uncertainty principles led to the ontological interpretation that reality—at least, at the subatomic level—is a probability function and to a questioning of the presumed distinction between the knowing subject and the object of his/her knowledge. Quantum scientists have shown repeatedly that subatomic phenomena “behave” as wave, when experiments are set up to detect waves, and as particles when experiments are set up to detect particles. Thus, it is concluded, methods of inquiry interfere with reality; in other words, we participate in reality as we observe it.

E. Samuel Overman derived conclusions from quantum mechanics for public policy and administration.^[2,3] His key insight is that quantum mechanics heralds a convergence between the natural and social sciences: It has shown that natural phenomena are indeterministic and unpredictable at the most

fundamental level of existence, very much like social phenomena as described by many social scientists. According to Overman, the complementarity principle of quantum mechanics suggests that the classical dichotomies of public administration and policy analysis—politics versus administration, facts versus values, theory X versus theory Y, rationality versus disjointed incrementalism, centralization versus decentralization—are in fact complementary aspects of phenomena. Conceptual dichotomies are epistemological illusions.

He suggests that Simon's concept of "bounded rationality" and the Hawthorne effect are examples of the uncertainty principle. Social reality cannot be observed in its entirety or unobtrusively. A social observation involves a complex relationship between the observer and the observed. According to Overman, quantum mechanics suggests a constructivist epistemology, that understanding of phenomena is a matter of interpretive construction. In public policy making, realities are constructed and constituted through symbol manipulation and image making. Performance appraisals in organizations are another example of participating in reality while measuring it. Although the appraisals are aimed at providing "objective" measures of employees performances, they end up influencing performances.

CHAOS AND COMPLEXITY THEORIES

What emerged as *chaos theory* in the late 1970s morphed into the *sciences of complexity*—alternatively called *complexity theory* or *complex systems theory*—in the 1990s. The sciences of complexity are composed of a constellation of theories that question the deterministic ontological assumptions of Newtonian physics. Some complexity theorists also offer interpretivist and phenomenological epistemological perspectives for science.

Complexity is customarily defined in terms of the number of elements in a system. Complexity theory expands this definition and says that it involves not only the number of elements, but also the non-linear (disproportionate) nature of relationships between them—as opposed to linear (proportionate) relationships. As such, complexity creates the conditions for unpredictability in the future behavior of a system. Complexity theorists also adopted the system-theoretic notions that the properties of complex systems emerge endogenously, and that once they do, they are irreducible to the properties of the system's components (i.e., system characteristics are holistic).

Complexity is also a function of the interactions between reality and its observer. As physicist Murray Gell-Mann puts it, a measure of complexity must

include the level of the detail of the description of a system, as well as the previous knowledge about the system and the language used in the measurement of complexity.^[4] Therefore, complexity is not only an ontological characteristic but also a product of the knowledge process. This view brings scientific understanding closer to interpretive and phenomenological epistemologies.

The most significant contribution of chaos theory is to show that "chaos" and "order" are not dichotomous or mutually exclusive categories. In fact, there is order within chaos and chaos within order. This notion defies the Newtonian ontological assumption that there is an ultimate orderliness and determinism in the universe, which can be known in its entirety by the human intellect. Chaos theory also shows that even when the relationships between a system's components are deterministic, the effects of the system may not be predictable in quantitative, precise terms. But there still, is an order within seemingly random system behaviors. This can be predicted and understood in qualitative terms.

Two leading theorists of complexity, Ilya Prigogine and Stuart Kauffman, describe the mechanisms of *self-organization*, which govern the emergence of order from disorder.^[5,6] What Prigogine calls "dissipative structures" are the systems that continually dissipate energy and matter to their environments to maintain their structure. These systems have the tendency to drift toward "far-from-equilibrium" conditions. Under those conditions, they reach certain "bifurcation points," where they fall into chaos or leap to a new, more differentiated order (i.e., they self-organize at higher levels of complexity).

According to Kauffman, biological systems are typically in non-equilibrium states, and they have the tendency to evolve to the "edge of chaos"—a concept similar to Prigogine's far-from-equilibrium conditions. At the edge of chaos, phase transitions take place into more complex forms. This is a self-organizational (endogenous) process, because it does not require an interference from external factors. Emergent complex forms are holistic. For example, once life emerges at the edge of a chaotic state of molecules, its complexity is not reducible to the functioning of the molecules of which it is made.

Prigogine and Kauffman's theories of self-organization have been applied in organizational theory, public administration, and economics. Goldstein derives the conclusion from self-organization theories that organizational change can be a self-generated and self-guided process.^[7] Organizations have an innate potential for transformation, due to the non-linear interactions within them. Organizational change occurs when the organization's components are out-of-balance (i.e., when they are in far-from-equilibrium conditions).

Then new forms of organization emerge. Goldstein and other organization theorists suggest that the theories of self-organization invite us to adopt less control-oriented management philosophies and accept transformational change. Such philosophies, they argue, can increase the probability of successful transformations in organizations and allow the emergence of better organizational forms. Kiel draws similar conclusions for public agencies: Public administrators must embrace the dynamic instabilities and uncertainties, which are the products of non-linear relations in organizations. These dynamics can be the sources of positive transformations in public agencies.^[8]

A group of libertarian, neoclassical economists, such as James Buchanan and Friedrich Hayek, draw parallels between the idea of dynamic self-organization and laissez-faire economics. They describe markets as self-organizing dissipative structures and argue that socialist planning has had devastating effects on economies. Similarly, Kauffman argues that decentralized and flatter organizations, democracies, and free markets work better than their alternatives. When organizations are broken into smaller units, each unit attempts to maximize its self-interest, which, in turn, maximizes the interests of the organization.^[6]

Kellert opposes the libertarian, neoclassical interpretations of complexity.^[9] He argues that chaos theory particularly refutes the notion that laissez-faire capitalism creates optimal solutions. To the contrary, it demonstrates that systems may end up chaotic. Chaos theory also shows that although it is impossible to make precise (quantitative) predictions in chaotic behavior regimes, it is still possible to determine the range and shape of fluctuations qualitatively (chaotic attractors, such as the well-known Lorenz attractor, illustrate this). The theory allows us to determine critical points in an economic system so that we can intervene and make macroeconomic adjustments.

Agent-Based Simulations

Some concepts of complexity theory are applied in *agent-based simulations*, which grew out of the studies on cellular automata, neural networks, and genetic algorithms in computer sciences. Agent-based simulations offer a methodological alternative to the deductive-nomothetic methods of Newtonian science. Newtonian methods strive to simplify reality with a small number of variables, while artificial life simulations generate complex and unique results.

In agent-based simulations, the behaviors of “agents” are simulated using specialized computer software. Agents are individual members of a group, each with its internal states and behavioral rules. They “live” in an environment that imposes some

constraints. In these simulations, the environment is a virtual lattice in which agents are initially placed at certain locations and are let move around and interact with other agents, according to initially set rules. Agents organize themselves into dynamical patterns, collective structures, on the lattice.

Epstein and Axtell’s study is the most comprehensive example of the applications of agent-based simulations in the social sciences.^[10] They simulated a society whose members (agents) were imbued with a set of economic, cultural, and reproductive rules. The agents conduct economic transactions, compete with each other to survive, pollute their environment, form cultural groups, go on wars against other groups, reproduce, and die. Epstein and Axtell’s simulations show that simple equations (behaviors of individual agents) can generate complex behaviors, and that internal dynamics of systems can generate complex, orderly, and chaotic patterns. Social structures emerge through a self-organizing process of individual interactions. Collective structures, once emerged, are not decomposable; they are holistic and are sustained beyond the life span of the agents that generated them.

According to Epstein and Axtell, the method of agent-based simulations can be the basis of a unified and evolutionary social science, which they call “generative social science.” In generative social science, evolutionary simulation models would replace nomothetic explanations and idiographic descriptions. Its focus would be on the behaviors of individual agents, and the processes in which collective structures emerge.

Epstein and Axtell also draw the conclusion from their simulations for economic theories that the aggregate, decomposable models of neoclassical economics are inappropriate. The neoclassical economics uses the rational actor as its unit and aggregates an economy from the behaviors of these actors. It assumes that static equilibria exist, or at least are achievable, and ignores the time dynamics. Epstein and Axtell illustrate that, left alone, markets do not generate equilibrium conditions; instead, they create far-from-equilibrium conditions over time. Society does not necessarily benefit from individualistic competition, because in the process, wealth accumulates through inheritance, and a skewed social hierarchy emerges.

Fractal Geometry

In agent-based simulations, agents are placed in discrete locations on a two-dimensional lattice. This is a limitation for simulating real life, in which things occupy places that cannot be measured in exact terms. *Fractal geometry*, which is considered as the geometry of chaos theory, offers better mathematical modeling tools for simulating real life. Fractal geometry allows

fractional dimensions—“fractal” dimensions between 1 and 2, or 2 and 3—to be used in descriptions. Fractal dimensions represent the degree of roughness, brokenness, or irregularity in an object. In Euclidian geometry, objects are described in terms of straight lines and curves, squares, cubes, and globes, that have finite lengths; and their dimensions can be calculated as whole numbers. There are no perfectly straight lines or perfect squares or circles in nature, however. Benoit Mandelbrot, founder of fractal geometry, studied objects that are irregular in dimensionality, but those irregularities remain constant over different scales, which he called “self-similarity across scales.” Mandelbrot found examples of such self-similar irregularities in nature. The branches of most plants and trees display fractal shapes, for example. In mathematics, fractal images are generated iteratively, by applying the same mathematical rule repeatedly on different scales.

Fractal geometry also demonstrates that the measurement of an object depends on the measurement instrument used. Different resolutions and different scales yield different results in a measurement. Therefore, measurements are not absolute; they are not “objective.” Mandelbrot emphasized that measuring dimensionality depends on the relation between an object and its observer.

AUTOPOIESIS

Autopoiesis is a theory of biological systems. It was developed by the Chilean biologists Humberto Maturana and Francisco Varela in the 1960s and early 1970s. Autopoiesis takes the idea of self-organization one step further and suggests that biological systems are *self-producing* and *self-referential*. It challenges the traditional view that systems are open to the influences from their environments and suggests that all living systems are organizationally closed and autonomous. These systems self-renew and self-create, and their most important products are themselves. A system’s interaction with its environment is a reflection of its organization. The environment is a part of the system, and it is coproduced by the system. In fact, there is no beginning or end to a system; it is a closed loop of interaction.

Autopoiesis theory suggests that the brain is an autopoietic system. It does not simply process information that comes from the environment; instead, it assigns patterns to incoming information. Thus, autopoiesis suggests a constructivist epistemology: The brain does not reflect reality, but it constructs a reality according to its own organization.

Autopoiesis theory has been applied to social organizations and political systems. In Gareth Morgan’s

interpretation, autopoiesis suggests that organizations attempt to achieve a self-referential closure in relation to their environments. They enact their environments as extensions of their own identity.^[11] Actually, most of the problems organizations encounter emanate from the identity they try to maintain. “Egocentric organizations,” which insist on maintaining their identities by imposing them on their environments, end up sustaining unrealistic identities. If an organization sees itself separate from its environment and acts accordingly, it loses sight of the fact that it exists in relationships and that survival is always with the environment, not against it.

Several theorists, including Niklas Luhman and Bob Jessop, drew conclusions from autopoiesis theory for social and political systems and the role of public administration in society. The first general conclusion is that government and public administration are self-producing and organizationally closed systems. The role of public administration is not simply to execute policies but to produce government through complex interactions with its political and social environment. The bureaucratic system aims to preserve its existence and reproduces itself. This poses challenges for democratic governance and administration.^[12]

The second general conclusion is that public administration, or government, cannot control social processes, because other social systems are autonomous too. Therefore, monocentric and hierarchical model of government should be replaced by models of governance in polycentric networks. Such models would be more reflective of realities in which governments do not control other organizations with regulations, but “steer” them through bargaining, negotiating, and participating in markets.^[13]

FUZZY LOGIC

The history of *fuzzy logic*—or, *fuzzy set theory*—goes back to the 1960s, when Lotfi Zadeh proposed it as an alternative to the Aristotelian *sylogistic logic*—or, the theory of *ordinary (crisp) sets*. Unlike ordinary sets, fuzzy sets include elements with partial membership (they are not categorical), and fuzzy logic uses operations that are not available to syllogistic logic. According to Bart Kosko, who is the leading theorist of fuzzy logic today, it challenges a series of “bivalent” assumptions—dichotomies such as black–white, existent–non-existent, either–or, zero–one—which are fundamental to logical positivist science.^[14] Bivalent science mistakenly assumes that its categories correspond to discrete realities and uses a binary mathematics in which numbers are rounded up to achieve discreteness. This is a fallacy, and there is a mismatch between bivalent science and fuzzy reality.

Bivalent science, according to Kosko, is a product of the Western worldview with roots that go back to Ancient Greek philosophy, particularly Aristotle. The Western culture sees binary precision as part of the scientific method. The Eastern philosophies, such as Buddhism, on the other hand, recognize fuzziness and make it a cardinal principle in their understanding of knowledge processes. He asserts that despite all the efforts by bivalent science to make our minds work in a binary fashion, our minds are still capable of knowing fuzzy realities. Eastern societies, because of their openness to fuzzy thinking, have quickly adopted the principles of fuzzy logic and used them successfully to develop more advanced and intelligent industrial products.

There are few applications of fuzzy logic in the social sciences. Among them, Charles Ragin's methodological applications are the most comprehensive.^[15] Ragin sees the applications of fuzzy sets as complementary of what he calls *diversity-oriented methodology*. The commonly used *variable-oriented (quantitative) methodologies* are based on the assumption that every unit of measurement is equivalent, and they ignore the complexity and diversity of social phenomena. Diversity-oriented approach, on the other hand, is concerned with exploring and constructing types and kinds of cases and recognizes causal heterogeneity. In this approach, individual cases are not viewed as collections of analytically distinct variables, but as configurations of sets of membership, and it is recognized that causes may have opposite effects on individual cases.

According to Ragin, fuzzy sets extend and deepen diversity-oriented research strategies. Fuzzy sets allow us to conceptualize diversity not only in the different configurations of set membership but also in the degree of their membership in such sets. As such, they can capture the complexity of realities better and thus help improve the link between theory and data analysis. Fuzzy sets combine qualitative and quantitative assessments in a single instrument. Fuzzy sets contain two qualitative states—full membership of an element in a set (“1”) and full non-membership (“0”)—as well as all the quantitative variations between the two (i.e., partial membership at different degrees; 0.35, 0.50, 0.76, etc.). Fuzzy sets use Boolean logical operations (such as *negation*, logical *and*, and logical *or*) and their own operations (such as *concentration* and *dilation*), which enhances their capabilities.

These enhanced capabilities enable and force researchers to be as faithful as possible to their

concepts and recalibrate them if necessary. Ragin stresses that fuzzy-set applications are not merely mechanical exercises, but they are fundamentally interpretive acts. They prod researchers to pay careful attention to the meaning of concepts; thus, they help find stronger links between theory and research.

REFERENCES

1. Wheatley, M.J. *Leadership and the New Science: Learning About Organization from an Orderly Universe*; Berrett-Koehler: San Francisco, 1992.
2. Overman, E.S. Policy physics: exploring the convergence of quantum physics and policy science. *Manage. Sci. Policy Anal. J.* **1987**, 4 (2), 30–39.
3. Overman, E.S. Policy physics. In *Quantum Politics: Applying Quantum Theory to Political Phenomena*; Becker, T.L., Ed.; Praeger: Westport, CT, 1991; 151–167.
4. Gell-Mann, M. What is complexity? *Complexity* **1995**, 1 (1), 16–19.
5. Prigogine, I.; Stengers, I. *Order Out of Chaos: Man's New Dialogue with Nature*; Bantam Books: New York, 1984.
6. Kauffman, S.A. *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*; Oxford University Press: New York, 1995.
7. Goldstein, J. *The Unshackled Organization: Facing the Challenge of Unpredictability Through Spontaneous Reorganization*; Productivity Press: Portland, OR, 1994.
8. Kiel, L.D. *Managing Chaos and Complexity in Government: A New Paradigm for Managing Change, Innovation, and Organizational Renewal*; Jossey-Bass: San Francisco, 1994.
9. Kellert, S.H. When is the economy not like the weather? The problem of extending chaos theory to the social sciences. In *Chaos and Society*; Albert, A., Ed.; IOS Press: Amsterdam, 1995; 35–48.
10. Epstein, J.M.; Axtell, R. *Growing Artificial Societies: Social Science from the Bottom Up*; Brookings Institution Press: Washington, DC, 1996.
11. Morgan, G. *Images of Organization*, 2nd Ed.; Sage: Thousand Oaks, CA, 1997.
12. Kickert, W.J. Autopoiesis and the science of (public) administration: essence, sense and non-sense. *Organ. Stud.* **1993**, 14 (2), 261–278.
13. Dunshire, A. Tipping the balance: autopoiesis and governance. *Adm. Soc.* **1996**, 28 (3), 299–334.
14. Kosko, B. *Fuzzy Thinking: The New Science of Fuzzy Logic*; Hyperion: New York, 1993.
15. Ragin, C.C. *Fuzzy-Set Social Science*; The University of Chicago Press: Chicago, 2000.

Newtonian Paradigm

Mary Ann McClure

*Department of Art, Music and Philosophy, John Jay College of Criminal Justice,
New York, New York, U.S.A.*

INTRODUCTION

Modern science began in the sixteenth and seventeenth centuries with the discoveries of Galileo, Copernicus, and Newton. Conceiving of nature as a machine that works with clock-like regularity, the Newtonian paradigm celebrated order and promised prediction and control. Central to this view of science was the presumption that we live in a universe governed by immutable laws. Human reason was applied to the natural world and a law-like mathematical order was assumed to exist in nature. With Newton's identification of gravity as the force that defined both the motion of the planets and motion on the earth, the same set of universal rules governed the terrestrial and celestial realms.

LOSS OF POWER TO PREDICT

Because all physical systems obey universal laws, accurate prediction “merely” depends on the accurate measurement of the initial conditions. Such confidence in the power of science to predict reached its ultimate expression in the intellectual bravado of Pierre Simon de Laplace, a leading mathematician of the eighteenth century. Laplace claimed that, with knowledge of the initial conditions and intelligence formidable enough to perform the calculations, we could predict the state of the universe at any future moment.

Laplace's claim of omniscience, however, was soon challenged by a compatriot. In 1880, a French mathematician Jules Henri Poincare entered a competition to solve one of the biggest mathematical problems of the day, the “three-body problem.” This problem sounds deceptively simple: “how do you determine the motion of three bodies in space if you know their starting positions?” To his surprise, Poincare discovered that one of the three bodies moves to all intents and purposes in a way that appears to be random.^[1] Yet, the motion is not really random because we know the mathematical equations for it. Poincare identified the phenomenon that would later be called chaos. Chaos occurs when any deterministic system—that is, a system that has laws that govern its behavior—has a solution that is so complex and so irregular

that it appears to be random unless we know a lot of hidden information about what it is doing.

The transforming insight of the Newtonian paradigm had been that the world obeys rules rather than being run by mystic powers or spirits unknowable to the human intellect. Furthermore, if the rules are simple, then you get predictable outcomes. Things that appear to be random, like the spinning of a ball in a roulette wheel, are what they are because they are governed by rules that are incredibly complicated. That is, complicated things are the result of hugely complicated rules. The concept of chaos undercuts this simple Newtonian view of the universe. The world of classical mechanics turns out to be much more mysterious than Newton, or even Einstein, imagined.

Chaos shows that some of the simplest rules you can imagine—simple, rigid, prescribed rules, with nothing random in them, like the rules that governed the motion of Poincare's third body—can generate behavior that is both random and complex. No longer do complicated things have to be the result of complex rules. Furthermore, simple rules can behave in such a complicated way that they do not have a predictable outcome.

SENSITIVE DEPENDENCE ON INITIAL CONDITIONS

Nevertheless Poincare's work was shelved and ignored by the major physicists and mathematicians of his time. Seventy years later, Edward Lorenz, a meteorologist working on mathematical models for the behavior of the earth's atmosphere, inadvertently made a discovery that would bring chaos theory to modern attention. In 1959, he left a room while leaving his computer running. When he was gone, the computer rounded off a few decimal points that resulted in an imperceptible difference in the initial conditions. When Lorenz got back, he discovered a radically different pattern had emerged from this insignificant change. He had thus discovered that insignificant differences in initial conditions, if given enough time, could produce different outcomes. Lorenz had identified a distinguishing characteristic of chaotic behavior: sensitive dependence on initial conditions.

Consider the motion of the earth around the sun. Its behavior, manifesting Newtonian regularity, is captured by a linear mathematical equation and is easy to predict. Linearity implies proportionality: small causes create small effects. If something pushed the earth slightly out of orbit, the result would be a new orbit close to the original. Chaotic systems are, however, non-linear: small causes can create big effects. In a chaotic system, the rate of change is expressed by a non-linear equation; in a non-linear world, exact prediction is both practically and theoretically impossible. For example, for insect populations, when the system is in a chaotic state, two initial populations that differ, say, by less than 1% will have different numbers in just a few years. Thus, if the initial configuration of a chaotic system is changed even slightly, its future behavior is altered radically.^[2]

Pointing out the consequences of his discovery, Lorenz observed, “It implies that two states differing by imperceptible amounts may eventually evolve into two considerably different states. If, then, there is any error whatever in observing the present state—in any real system such errors seem inevitable—an acceptable prediction of an instantaneous state in the distant future may well be impossible.”^[3] In practical terms, for Lorenz’s particular project, it also meant that weather is inherently unpredictable. Long-range forecasts are not possible. No matter how complex our models become, or no matter how accurate our measurement of the initial data, there is a limit beyond which we can possibly accurately predict.

In most cases, Newton, of course, had it right. We have been using him for the last 300 years to predict the motion of heavenly bodies, the flight of space craft, and also more mundane motions. However, if you fine-tune Newton, you find something a little different. Newton says that *if* you know the initial conditions perfectly, then you can predict anything in nature. The problem is the “if.” Variables within a system can be fixed only within some finite degree of accuracy. For example, it is impossible to pin down the position of a particle in space with perfect accuracy. In a chaotic system, these tiny variations get amplified. Lorenz called these amplification of the smallest differences, the “butterfly effect.” That is, the flapping of a butterfly’s wings in Chile leads to a tornado in Japan.

Admittedly, the predictive powers of modern science had already been challenged by quantum theory. But quantum uncertainty focuses on activity on the microlevel. Furthermore, the effects of quantum fluctuations are extremely small and tend to cancel each other out. But with chaos theory, phenomena at the level of ordinary everyday reality have become unpredictable.

ORDER WITHIN CHAOS: STRANGE ATTRACTORS

But within the unpredictability and apparent randomness of chaotic behavior, there is order, a hidden structure that is mapped out by strange attractors. Strange attractors model the behavior of chaotic systems in phase space. Phase space charts the changes in the system’s behavior over repeated cycles. An attractor is simply any point within an orbit that seems to attract the system to it. Attractors are labeled “strange” when their orbits wander in complex and unpredictable ways. For example, coordinates that are close in numerical value can have orbits far apart. But the orbits of the strange attractor always stay within a certain path so that the shape is quite distinct. Measuring a transverse cross-section of the volume of any strange attractor will always yield a fractal.^[4] A fractal is a form characterized by recursive symmetries that occur when the same pattern is repeated in different sizes and shapes. This characteristic of strange attractors links chaos theory to topology and the geometry of fractals developed by Benoit Mandelbrot.

Part of the power of chaos theory lies in the fact that an amazing range of diverse systems can be modeled as strange attractors, including the erratic eye movements of schizophrenics, variations in cotton prices, and outbreaks of infectious diseases. Chaos theory looks for changes in scale and follows the behavior of the system as iterative (the result of a calculation being feedback into the formula again and again) formulae change incrementally.

From this perspective, a proper explanation is one that is able to model large-scale changes through the incremental evolution of a few iterative equations. For example, in a study of the eye motion of schizophrenics, the emphasis of the model is not on the eye so much as the equations that, when iterated, would produce wandering orbits equivalent to the eye’s movement.^[5] From this point of view, it is coincidental that the system in question is an eye. It could equally well have been the fibrillation in the heart or the fluctuation in cotton prices. Models do not take into account a system’s specific mechanism. This represents a different concept of explanation than that given in traditional science where to explain means to uncover the system’s specific mechanism that causes behavior. Here you are producing a mathematical model that simulates the behavior. Chaos theory looks to the mechanism that will show how patterns arise alongside unpredictable behavior, providing an understanding of “how it happens” rather than “why it had to happen.”

To sum up, a chaotic system is characterized by deterministic randomness. Deterministic because the chaotic behavior itself arises from intrinsic causes that can frequently be expressed in simple equations and

whose order can be mapped by strange attractors. Randomness because the behavior remains irregular and unpredictable.

PRESUPPOSITIONS OF THE NEWTONIAN PARADIGM AND CHAOS THEORY

Chaos theory operates on a different set of assumptions than those of the Newtonian paradigm. First, in Newtonian paradigm, problems or events are linear. That is, small causes create small effects, and scientific theories have to be quantitative to even be considered theories. But chaotic systems are always non-linear systems in which small causes can create big effects. Because quantification is not possible with dynamic systems, chaos theory must turn to qualification. It asks—what does the system look like as it moves through changes?—and answers by watching the strange attractor being displayed on a screen. Thus, the study of chaos yields qualitative predictions for systems where detailed quantitative predictions are not possible. Qualitative theories become as important, if not more so, than the quantitative.

Second, from the Newtonian perspective, objects are considered to be independent of scale. Measuring a room with a yardstick or an inch ruler produces the same result. But chaos theory deals with the complex, irregular forms of turbulence and fractal geometry that are scale dependent. Measuring a rugged coastline with a mile scale or an inch scale produces different results because the inch scale is always going to capture more length.

Third, the Newton paradigm employs a foundationist approach that focuses on individual particles or units. To study a system is to follow the unit through time, adding the behavior of the units. This is a reductionistic method. One isolates the appropriate unit and understands the mechanism that binds the units to the larger whole. To understand how a clock works, you break it down into its simplest parts. This reductionist

spirit still pervades modern science. A salient example is the reductionist program of molecular biology, which has achieved its fullest expression in the human genome project.

In contrast, for chaos, the individual does not matter. Remember Poincaré's work on the three-body problem. Reductionism holds that the laws affecting the parts give the behavior of the whole. But the action of two bodies is linear and of three, chaotic. Thus, reductionism is contradicted. Chaotic behavior is governed by the strange attractor that affects the system as a whole. The emphasis shifts from individual units to recursive symmetries between scale levels.^[6]

In conclusion, the history of science dominated by the Newtonian paradigm has been the struggle to impose order on a world of turbulence and disorder, a world where the only straight line is one that man has introduced. Chaos science produces a new geometry that shows there is order within disorder. It destabilizes the order-disorder dichotomy, revealing a territory that cannot be assimilated into either order or disorder. Chaos theory represents the revisioning of the world as dynamic and non-linear, yet predictable in its very unpredictability.

REFERENCES

1. Kellert, S.H. *In the Wake of Chaos*; The University of Chicago Press: Chicago, 1993; 121–122.
2. Poole, R. Chaos theory: How big an advance? *Science* **1989**, *245*, 26–28.
3. Lorenz, E. Deterministic aperiodic flow. *J. Atmos. Sci.* **1963**, *20*, 133.
4. Stewart, I. *Does God Play Dice? The Mathematics of Chaos*; Basil Blackwell: Oxford, UK, 1989; 182–188.
5. Hayles, K. *Chaos Bound: Orderly Disorder in Contemporary Literature and Science*; Cornell University Press: Ithaca, NY, 1990; 170 pp.
6. West, B. Chaos and related things: a tutorial. *J. Mind Behav.* **1997**, *18*, 103–126.

No Child Left Behind Act of 2001

G.L.A. Harris

*Mark O. Hatfield School of Government, Portland State University,
Portland, Oregon, U.S.A.*

INTRODUCTION

The No Child Left Behind (NCLB) Act of 2001 became federal law with great fanfare and promise to overhaul the public school education system in the United States. According to the Act, public schools must demonstrate adherence to well-defined educational principles.^[1] The purpose of this essay is to show the evolution of the modern public school reform movement in the United States, the movement encapsulated in its present day form as the No Child Left Behind Act of 2001, assertions by both supporters and detractors of the Act, and the growing pains experienced by states as they attempt to navigate the implementation process.

OVERVIEW

The NCLB Act of 2001 is comprised of Titles I–X and consists of the following: Title I, Improving the Academic Achievement of the Disadvantaged; Title II, Preparing, Training and Recruiting High Quality Teachers and Principals; Title III, Language Instruction for Limited English Proficient and Immigrant Students; Title IV, 21st Century Schools; Title V, Promoting Informed Parental Choice and Innovative Programs; Title VI, Flexibility and Accountability; Title VII, Indian, Native Hawaiian and Alaska Native Education; Title VIII, Impact Aid Program; Title IX, General Provisions such as the Unsafe School Choice Option; and Title X, Repeals, Redesignations, and Amendments to Other Statutes such as the McKinney-Vento Homeless Education Assistance Improvements.

The NCLB Act purports to focus on what works by emphasizing proven educational programs and practices, reducing bureaucracy and increasing local flexibility, increasing accountability for student performance and empowering parents. Participating states in turn must align their own school districts' policies to the federal guidelines of the Act.

First, the NCLB Act requires annual report cards as performance reports. The report cards will reflect the degree to which a school is accountable by its performance on state designed annual assessments for reading and mathematics in grades 3–8. Reports will enable parents, educators, administrators, policy makers, and

the general public to track the performance of schools. Such assessments would reflect each state's ability to hold schools and districts accountable for student achievement standards the state mandated; student, school, and district progress annually to meet the standards; and the amount of federal aid in financing school reform.

Schools that meet or exceed states' standards would be eligible for academic achievement awards. However, those schools that fail to "meet the grade," would first be targeted for assistance, then remediation, and finally restructuring.

Second, NCLB promises flexibility and thus greater accountability in order to encourage more control of the decision making at the state and local levels where educators are more likely to be familiar with their students' needs. State and local jurisdictions in turn can spend at least half of the federal funds earmarked for other school programs for NCLB activities including the hiring of new teachers without federal approval.

Third, parents would gain greater control over which schools that their children attend. Parents could exercise an option to transfer their children from a school that failed to meet state standards for two consecutive years to a better performing public school within their districts. Such options would also extend to the children of low-income families who attend low performing schools that fail state standards for three years. Children who transfer may receive supplemental educational services such as tutoring, after-school services, and summer school. Further, students in poor performing schools that their parents deem to be unsafe can transfer to safer, better performing schools within their districts.

Finally, the Act would support educational programs or techniques that come from rigorous research with opportunities for educators to upgrade their skills through professional development. The Act's four-pronged emphasis on accountability, flexibility, parental choice, and research would provide the hallmarks for success.

THE BEGINNINGS

NCLB's roots are unclear at first reading. They appear to emanate from four earlier attempts at educational

reform. In a speech on April 22, 2004 at Harvard University's Kennedy School of Government, Education Secretary Rod Paige described the *Brown v. the Board of Education* May 17, 1954 decision as akin to the NCLB Act of 2001. He cited the vehement resistance of the time as an educational divide that is "cruel, vicious, demeaning, disrespectful and degrading." He went on to say that "I believe that when a child is left behind by an education system that is an injustice that affects us all. It is intolerable."^[2] Scott (see p. 179 of Ref.^[3]) also points to the U.S. Supreme Court's ruling in *Brown v. the Board of Education* (347 U.S. 483, 1954). Scott (see p. 179 of Ref.^[3]) establishes the *Brown* case as the starting point in modern school reform by the decision's holding that "racial separation adversely affected self-esteem, which reduced motivation and consequently retarded the educational and mental development of Negro children." Nowhere is such injustice prevalent and a desire for the "quality of education stronger" than in African American communities. (See p. 203 of Ref.^[3].) Scott cites Henninger (2003) in a 2003 poll by Public Agenda which found that 82% of African Americans parents, for instance, desire efforts to promote a quality education.

The Elementary and Secondary Education Act (ESEA) of 1965 marked a second attempt at closing the education achievement gap.^[4] In NCLB such efforts were not directed just towards bridging the gulf that existed between black and white children but also in achieving educational parity for the nation's poor and disadvantaged youths through federal funding of public schools K-12 educational programs.

However, following some 20 years of awkward and repeated trials at educational reform, including school busing during the 1970s, a 1983 report entitled "A Nation at Risk: The Imperative for Educational Reform" served as the third basis for NCLB. The report's authors accused the nation's public education system of being "a rising tide of mediocrity." (See p. 5 of Ref.^[5].) Mediocrity, according to the report, "threatens our very future as a Nation and a people." (See p. 5 of Ref.^[5].) Many reformers were legitimately concerned with the history in many school districts—especially urban schools—of students who graduate from high school unable to read and write at the minimal level; of teachers who lack the skills, concern or drive to reach students; of special education placements that seem to follow racial categories; and of rigid tracking. This national wake-up call, they hoped, would "generate reform of our educational system in fundamental ways." (See p. 6 of Ref.^[5].)

Finally, much of NCLB grew out of provisions written into various proposals that became laws after 1965. With the Improving America's Schools Act (IASA) of 1994,^[6] a revision of the Elementary and Secondary

Education Act of 1965,^[4] required that states develop assessments in keeping with local standards for reading, mathematics and science at all grade levels. The officials also required that state administrators use the data derived from assessments to track student performance improvement and to identify low performing schools.^[7]

NCLB crystallized national thinking about education reform as change that must give greater responsibility to local administrators under state government authority. Unlike its predecessors, the Act gives states and local districts flexibility in choosing how to spend federally allocated funding for education. States and local districts must set, adhere to and meet or exceed the state standards for student achievement. The standards and actual performance would lead to greater accountability. Parents, too, could exercise a wide range of options in hoping to secure the best education for their children. Teachers might pursue continuing professional development to ensure that they could employ proven techniques. In essence, the NCLB Act would attempt to "foster an environment in which every child can learn and succeed."^[1]

THE PROS

Proponents of NCLB point to many features of the Act's goal to maintain high standards. First, the Act recognizes that low achievement may frequently reflect education programs of compromising quality and instruction by inadequately trained or uncertified teachers.^[8] Testing and accountability through the Act could "illuminate problems that would otherwise go undetected, allowing teachers to tweak their lesson plans or add hands-on projects." (See p. 481 of Ref.^[8].) Testing not only would serve as a tool for teaching and learning but also good educators would not need to teach to the test or depart from individualized instruction. Educators would have the latitude to structure lesson plans in ways that motivate students.

Second, accountability would empower parents. Parents would get help in making decisions about where to help their children gain an education of higher quality. Annual report cards would separate the best performing schools from the lowest performing ones. State assessment programs would enable officials to raise achievement levels for all students. States would be able to close the achievement gap of low performing students, particularly those from low income and minority groups.

Further, proponents argue that the accountability systems NCLB called for already exist. Many schools "are beating the odds" according to Strahan (see p. 297 of Ref.^[9]), who cites case studies of four schools. Collectively, these case studies have highlighted the

success of three specific paths towards education reform.

First, academic engagement of students in meaningful classroom tasks fundamentally influences student achievement. Second, based on a 1998 Sanders and Horn report from the Tennessee Value-Added Assessment System,^[10] Strahan sees the quality of the teacher as another key contributing element in academic achievement. Strahan discounts such demographic variables as race, socioeconomic level and class size as “poor predictors of student academic growth.” (See p. 298 of Ref.^[9].) Third, collective effort by teachers appears to go a long way to improving academic achievement. Collective faculty efficacy or “the extent to which teachers have the belief that they can work together to produce effects” was a significant predictor of student achievement in that teachers “shared perception over time” helps to shape a child’s environment in school. (See pp. 298–299 of Ref.^[9].)

However, Strahan quickly notes that teachers may not necessarily teach to a particular program. While academic engagement in meaningful classroom tasks, the quality of teachers, and collective efficacy of the faculty have proven to be effective, the most successful schools are those that tend to use curricula in a way that accounts for their students’ needs.

THE CONS

Although NCLB received overwhelming bipartisan support, critics argue that the Act will serve only to label schools as failures while failing to provide those same schools with the resources or support necessary to increase achievement.^[8] Many opponents also question the validity and reliability of the Act’s assessment strategies in reforming the public school system.^[8]

Failure and success, as well as the assessment strategies called for in the Act, center on systematic student testing. Standardized testing has existed for some time and has gained popularity because Americans have come to accept that reward and punishment are effective tools to use in achieving any goal.^[11] While these testing-based assessment strategies may elicit temporary compliance and some evidence of immediate achievement, they could do so at significant costs, say critics. Most notably, research suggests that the more rewards or punishment are used as a way of inducing people to engage in an activity, in this case, to improve student’s performance on standardized tests, the more that these students will lose interest in what has to be accomplished in order to receive the reward or escape punishment.^[11] These research findings relate to the carrot and stick control that standardized tests provide. Testing becomes a high-risk condition for students. The Act provides that students may be held back in

grade levels when their test results in low scores, prevented from graduating based on low scores, and their teachers evaluated on students’ test scores with salaries increased or decreased accordingly.

Critics point out the fallacy of standardized testing. They show that such methods were originally designed and intended for the homogeneous population in France as a way of identifying children who were slow learners so that those slow learners could be given remedial work.^[12] The French government’s attempt at using the most objective measure avoided the prejudices of teachers who might be inclined to identify such students as mentally impaired. But it was the American version of this imported measure through Stanford University psychologist Lewis Terman who revised the Alfred Binet measure to establish norms for American children. However, the American tests were never designed to reflect differences among a heterogeneous population. In essence, the American version did not serve to bring slow learners up to par, becoming instead a gauge to categorize, isolate, and even ostracize people in school. Wade and Tavis citing the work of Gould (1996), claimed that the testers not only failed to consider the heterogeneity of the American population but that such diversity would also reflect differences in background and experience for which American standardized tests are not purported to measure.

Critics see the continued justification for the use of standardized tests as relatively inexpensive and easy to implement.^[13] These critics attack policy makers who favor simplistic and cheap solutions over complex and expensive ones. These critics argue that educators should emphasize more powerful and longer-term methods such as curriculum-focused professional development. Further, civil rights activists in both New York and Texas have challenged NCLB in court, arguing that this kind of high stakes accountability system has served to discriminate against poor and minority students. Researchers and educators believe that standardized test-guided systems narrow curricula and limit teachers’ flexibility and creativity.^[7] According to Price,^[14] standardized tests linked to student classroom performance threatens academic freedom. Moreover, a survey in 2003 of teachers in the Memphis and Shelby County Tennessee school districts on high stakes testing revealed that many teachers felt restricted in the classroom due to the undue demands that a heavy testing schedule places on classroom time.^[15] Other teachers have complained of the cheating in such testing or a systemic exclusion of weak students on test days in order to raise the average test scores as well as rendering assistance to students during which time test answers are changed.

Accordingly, accountability standards have created an environment where teachers and administrators

face discipline for violating security rules for standardized testing. A school principal has opted for retirement to escape the wrath of a disciplinary hearing involving charges of impropriety on her school's Tennessee Comprehensive Assessment Program (TCAP).^[15]

Critics charge that the NCLB Act only amounts to a "collective spine counting" by using the conclusion of scientifically based studies to identify patterns of student achievement. (See p. 304 of Ref.^[9].) Use of such scientific methods therefore cannot begin to capture the essence of meaningful reform. Accordingly, students, teachers, and administrators experience this process at a much deeper and personal level. What is unique about the schools profiled in the Strahan^[9] article is that they used given mandates for accountability but accomplished student achievement in their own ways, even exceeding expectations. In other words, they applied local solutions to solve local problems.

Opponents note the pattern of educators to use tests to administer a one-size-fits-all style of education while discounting individual attention and instruction.^[8] Opponents believe that educators will "teach to the test" while emphasizing test taking strategy at the expense of critical thinking. (See p. 482 of Ref.^[8].)

Opponents also have confidence that accountability need not always be as NCLB orders, a high-stakes testing report card. There are other ways of measuring accountability such as "using presentations and portfolios to gauge the students' ability to connect ideas, think critically and use information meaningfully." (See p. 483 of Ref.^[8].) The fear is that accountability standards based on high-risk standardized testing will encourage states to use such test scores to confirm their biases or findings.

IMPLEMENTATION AND REACTION

Many states have found the provisions of the NCLB intrusive. For instance, Virginia's House of Delegates recently approved a resolution requesting that Congress exempt the state from the Act's mandates.^[16] Jim Dillard, Republican head of the state's House Education Committee, considers the NCLB Act to be "utopian nonsense" and went on to say that "not only is it a massive federal intrusion, it is simply unworkable." (See p. 1 of Ref.^[16].) Utah also approved a similar measure as Virginia to opt out of the law's requirements although the state finally adopted a compromise to prevent the state from making expenditures on the law. Vermont's governor has also signed a similar ban.

Other states like Maine and New Hampshire have also proposed not using state funds for the law while Arizona, Hawaii, Minnesota, and New Hampshire have proposed measures to opt out of the law altogether.

State protests have already resulted in the federal government's reconsideration of testing requirements. The same tests will not be used for both native and non-English speaking students.^[16] Further, federal officials express willingness to negotiate teacher requirements in rural districts. And, governors like Jim McGreevey of New Jersey stated that, "Governors would look forward to working cooperatively in achieving the goals of NCLB. Yet, when you have some of the finest school districts in the nation being designated 'in need of improvement' or 'failing' because of an arbitrary benchmark, the legislation needs an agonizing reappraisal." (See p. 2 of Ref.^[16].)

Money became another contentious implementation issue. Some experts assert that plans tying federal funding to student performance make sense, but, cutting funds in poor performing schools does not.^[14] The carrot and stick strategy reinforced with funding, opponents argue, punishes those schools that need the funding the most and reinforces the notion that certain groups of people are beyond help. Many also view labels as "underperforming" or "failing" as synonymous with "undeserving." For example, the Act may aim to prevent schools from neglecting hard-to-teach groups. Many schools, such as Midwood High School in Brooklyn, New York, win science awards and have high graduation rates, but through the NCLB, state education departments label the schools "in need of improvement" because students fail to achieve the state's assessment standards. (See p. 2 of Ref.^[17].)

Many problems have surfaced from the actual implementation of the Act as it relates to compliance for testing. For example, many parents of children in New York's lowest performing schools, according to reports, were given only a few days to respond to the opportunity for moving their children to another, higher performing school.^[17] Available space for the students who wanted a transfer also became a problem. Students in Chicago encountered space problems with 270,000 students qualified for transfers, discovering that only 1097 spaces were available.

The most vociferous protests have come from teacher unions. Some threatened litigation against the U.S. Department of Education. The unions saw the Act's requirement, such as requiring that all teachers be qualified to teach in their respective subjects, as unreasonable. And, the teachers insist that test results could be attributed to a number of other factors including class size instead of teacher competency.

OPPOSITION DOES NOT MEAN COMPROMISING ACHIEVEMENT

Opponents contend that challenging NCLB, and thus standardized tests, does not mean they oppose high

academic standards.^[14] The challenge means that they oppose using such testing as the sole basis for reward or punishment. NCLB opponents believe that a more effective strategy is to support families and provide the resources public schools need to enable students with different needs to achieve at all academic levels.^[14]

Opponents point to the increasing pressure that states face to deliver improved test results or risk the loss of funding due to accountability standards and emphasis on standardized testing under the Act. Another approach might be evidence-based models of teaching to close the achievement gap, particularly among poor and minority students. But the literature suggests that Title I funds, estimated by the U.S. Department of Education in 2001 to be between \$400 and \$500 per student, are insufficient to close the gap alone.^[18]

Title I funds needed, in the form of preschool programs designed for early intervention, can cost up to 20 times the typical Title I allocation. Reduction in the size of classes could cost up to three times the allocation. Quality comprehensive school reform programs like Success for All, could require up to 20% more funds. Borman^[18] believes that much higher funding to these programs entails a greater and stronger commitment at the federal, state and local levels of government. This commitment must be long-term, coordinated, and systemic. Further, by investing in preschool, summer school and other school programs, the achievement gap can even be eliminated and produce long-term benefits for society.

In the meantime, according to Price^[14] with Schemo,^[19] “states are scrambling to rig tests results by lowering test standards so that they can claim academic improvements, though the federal Department of Education has vowed to end such practices and has declared such test cheaters enemies of equal justice and equal opportunity.” (See p. 7 of Ref.^[14].)

Realizing that states will have to expand their assessment programs, some resulting in the testing at additional grade levels, federal officials have promised to help cover the costs of developing these assessments. However, for the time being, states must absorb this additional financial burden without immediate relief.^[7]

UNINTENDED CONSEQUENCES

Many educators fear that the provisions of NCLB may have unintentional consequences. These include administering only certain mandated tests, manipulating data and teaching to the tests. An emphasis on testing could rob students of a well-rounded educational experience. Evidence of this tunnel vision approach to compliance has already appeared.^[7]

Some states have had to game the high risks in testing. State officials have installed a variation of compliance by administering only those tests mandated by the Act. These officials hope that stopping all but required tests will motivate students to focus on the required tests and succeed. The variation state officials have developed is to teach and test to the real test.

NCLB has provoked devious bureaucratic strategies often involving fraud as well. A Pittstown, Pennsylvania mother in January, 2004, grew increasingly concerned about her daughter’s receipt of unusually high test scores that she believed did not reflect her child’s mathematical ability.^[20] Suspicious, the mother asked that her child be retested. The child’s performance on the second test did not replicate the result of the first and the mother admitted that her daughter frequently “tries to get out of doing homework.” (See p. 5a of Ref.^[20].) The mother also discovered that the initial test showed answers that were originally marked wrong, erased, and changed to mark the correct answers.

The unduly stressful atmosphere created by the Act provoked another case in Connecticut. The Connecticut Education Association hired an independent economist to examine current data and to project how the Connecticut schools would fare under NCLB’s mandate of annual yearly progress (AYP).^[21] The intent was to help anticipate the effects of AYP in one of the highest achieving states in the nation. Even though the projected results were hypothetical, they were surprising. In the study, almost one-quarter of Connecticut schools would fail to meet the annual yearly progress standards. In essence, 155 elementary and middle schools and 88 high schools would be labeled “in need of improvement,” (see p. 47 of Ref.^[21]) and over the life of the Act or 12 years out, an astonishing 744 of the 802 elementary and middle schools or 93%, would fail to make their adequate yearly progress.

The Connecticut case reveals the questionable realism of the AYP formula. The formula fails to consider the differences among schools. The formula does not consider where a school starts and how quickly that school must reach its goal. Some schools, starting far behind others will need far more time to reach the state goals. Differences among schools put into question the AYP formula as the best gauge of student progress and thus school effectiveness.

Under NCLB, schools stand to lose significant funding if they fail to show mandated annual progress. However, this kind of pressure also supports those studies that indicate that such high stakes testing approach leads to rampant cheating and where poor performing students are reclassified as either disabled or special education students by excluding them from a school’s aggregated tests scores.^[22]

If this is the length to which some schools are willing to go in the quest for high achievement, then the NCLB Act of 2001 has forced an increasing number of cash-strapped states to do whatever is necessary to, in essence, “make it happen.” This blind eye at compliance then appears to further reinforce what other reformers have charged the public school system bureaucracy of doing. Teachers therefore implicitly understand that they must “teach to the test,” and in doing so, discourage critical thinking.

CONCLUSIONS

The NCLB Act of 2001 was created to redress the growing inequities and mediocrity that have culminated to define the nation’s public school system. The NCLB Act, it is hoped, would provide much needed resources to schools to achieve four clear cut goals: accountability for academic performance through annual reporting; built in flexibility for control at the local levels; increased parental involvement and thus school choice; and utilizing well-proven methodologies that are scientifically based. Continual assessment by using standardized tests is the cornerstone through which these goals can be achieved.

Proponents point to the Act’s use of high standards as a gauge to increase academic achievement as well as accountability. Small, but remarkable successes have also been reported for schools that are “beating the odds.”

Critics argue that the Act falls short by over relying on standardized testing as the tool to measure academic performance. Further, this one-size-fits-all style of education forces “teaching to the test” and discourages critical thinking in an attempt to preempt desired results.^[7]

An equal number of researchers also believe that standardized tests used to predict how likely students are to perform in subsequent educational settings and thus a school’s effectiveness, is deceptive.^[23] This can be attributed to what some researchers call “confounded causation” which they believe to consist of three main factors: what is taught in schools, a student’s native ability and out-of-school experience. However, it was the third factor that researchers found most troubling. Their argument: if children come from socioeconomically advantaged families and stimulus-rich environments, then they were more likely to succeed on items on standardized achievement tests than will other children whose environments do not mesh as well with what the tests measure.

These unintended consequences have not only resulted in some states forfeiting the right to funding that accompanies the Act’s implementation, but such action has forced the federal government to revise its

approach. One such revision includes negotiating teacher requirements in rural districts.^[16] However, some states are reshuffling priorities instead to avoid opting out of their share of the \$12.7 billion of federal funding promised by the Act.

The longitudinal study of 24,599 eight grade students by the U.S. Department of Education’s National Center for Education Statistics since 1988 shows that one consistent predictor of student achievement regardless of race and/or socioeconomic status is parental involvement.^[24,25] If this is the case, then the implementation of the NCLB Act of 2001 may be misguided and efforts should be expended on encouraging parents to invest time in their children’s education at home to help adequately prepare them for learning in school. According to Ramey and Ramey,^[26] for instance, such parental involvement is key to educating ones child.

Research also shows that early intervention at the preschool level is one proven strategy to help narrow the achievement gap and thus benefit society in the long run.^[18] However, under NCLB, and in the current era of austerity, educators and administrators might be tempted to go the path of political expedience at the expense of a quality education for children.

In theory, the NCLB Act of 2001 is perhaps the most revolutionary legislation to emerge in five decades. Its focus on accountability; flexibility and in turn local control given adherence to states’ assessments; enhanced parental control through annual report cards; and emphasis on scientifically based programs and techniques to improve students’ learning and performance; is to be commended, for without such established parameters, there would be an uneven handedness in its application and vast differences in the quality of education will exist from state to state and from school district to school district. However, the Act’s implementation becomes unwieldy as “the devil is in the details” and thus in its practice.

The federal guardians charged with the implementation of the NCLB Act of 2001 may be forced to reconcile some, if not all of the Act’s provisions, by taking a more practical approach as is already evidenced from the number of states, who despite their budgetary woes, have made the decision to opt out of the Act and thus lose needed funding. A re-examination of the extent to which states and in turn school districts may differ or have unique circumstances may be warranted. Further, and according to research, the most successful schools are those which, while meeting student achievement, are allowed to be flexible in their application. In other words, having a “one-size-fits-all” approach is impractical and in the long run serves to benefit no one especially the children for which the Act is intended.

REFERENCES

1. *No Child Left Behind Act of 2001. A Desktop Reference.* Prepared by the Office of the Undersecretary of Education. Public Law 107-110; U.S. Code 6301, 2002.
2. *Fifty Years After Brown v. Board of Education: What Has Been Accomplished and What Remains to be Done?* Remarks by Secretary Rod Paige at the Kennedy School of Government, April 22, 2004. Office of Public Affairs, U.S. Department of Education.
3. Scott, R. Five decades of federal initiatives concerning school desegregatory effects: what have we learned? *J. Soc. Political Econ. Stud.* **2003**, 28 (2), 177-215.
4. The Elementary and Secondary Education Act of 1965. Public Law 89-10; 20 U.S. Code 2701 et seq.
5. The National Commission on Excellence in Education. In *A Nation at Risk: The Imperative for Educational Reform*; U.S. Government Printing Office: Washington, DC, 1983; 20402.
6. Improving America's Schools Act of 1994. Public Law 103-382; 20 U.S. Code 6301 et seq.
7. Goertz, M.; Duffy, M. Mapping the landscape of high-stakes testing and accountability. *Theory Pract.* **2003**, 42 (1), 4-8.
8. Kucirik, E. No Child Left Behind Act of 2001: will it live up to its promise? *Georgetown J. Poverty Law Policy* **2002**, IX (2), 481-483.
9. Strahan, D. General patterns and particular pictures: lessons learned from reports from "beating the odds" schools. *J. Curriculum Supervision* **2003**, 18 (4), 296-305.
10. Sanders, W.; Horn, S. Research findings from the Tennessee value-added assessment system (TVAAS) database: implications for educational evaluation and research. *J. Person. Eval. Educ.* **1998**, 18 (4), 296-305.
11. Kohn, A. *The Schools Our Children Deserve: Moving Beyond Traditional Classrooms and Tougher Standards*; Houghton Mifflin Co., 1999.
12. Wade, C.; Tavis, C. *Psychology*, 7th Ed.; Prentice-Hall Publishers, 2003.
13. Elmore, R.F. A plea for strong practice. *Educ. Leadership* **2003**, 61 (3), 6-10.
14. Price, D.H. Outcome-based tyranny: teaching compliance while testing like a state. *Anthropol. Quart.* **2003**, 76 (4), 715-729.
15. High-stakes testing turns the screw. The commercial appeal (Memphis, TN). *Viewpoint* 2003, September 24, B4.
16. Toppo, G. States fight no child left behind, call it intrusive. *USA Today Section: Life*, February 12, 2004, 14d.
17. United States: Progress and Confusion; Education Reform. *The Economist* **2003**, 368 (8343), 50.
18. Borman, G.D. How can title I improve achievement? *Educational Leadership* (Association for Supervision and Curriculum Development). December 2002/January 2003.
19. Schemo, D.J. States get federal warning on school standards. *New York Times* **2002**, A25, 4.
20. Adams, B. Mom questions grading practices. *Times Leader* **2004**, May 3, 5a.
21. Weaver, R. Are the tests required by no child left behind making schools more accountable? No. NCLB's Excessive Reliance on Testing is Unrealistic and Frequently Unfair. *Insight on the News. Symposium*, May 24, 2004; 47 pp.
22. Schmidt, S. Parents trust teachers over testing: poll finds 65% prefer personal evaluations to standardized tests. *National Post (Canada). News* **2003**, July 11, A3.
23. Clinchy, E. *Transforming Public Education: A New Course for America's Future*; Teachers College Press: New York, 1997.
24. A Profile of the American Eight Grader: NELS: 88 Student Descriptive Summary of NCES Number: 90458, Release Date: June 27, 1990. National Educational Longitudinal Studies Data File User's Manual, Base Year: Student Component. National Center for Education Statistics, U.S. Department of Education.
25. *User's Manual: NELS: 88 Base Year Student Component Data File, NCES: 90464*, Release Date: June 22, 1990.
26. Ramey, S.L.; Ramey, C.T. *Going to School: How to Help Your Child Succeed: A Handbook for Parents*; Goddard Press, July 21, 1999.

Non-Monetary Incentives and Productivity

Hedy L. Isaacs

Department of Government, University of the West Indies, Kingston, Jamaica

INTRODUCTION

Non-monetary incentives refer to tangible and intangible means of motivating and rewarding employees. The humanistic philosophy that is the basis for using these incentives is that employees do not conform to the “men as machines” stereotype. Employees are social beings and non-monetary incentives may encourage them to achieve their own personal goals as well as their organization’s goals more readily, effectively, and efficiently. Agencies may use non-monetary incentives that are linked to employees’ needs and expectations about their job contents or contexts to accomplish their agencies’ goals. Non-monetary incentives may also be linked to employees’ expectations to balance their work lives with their personal lives. Their expectations may depend on the meaning and purpose of work. Non-monetary incentives are integral to the social and managerial aspects of an agency’s productivity improvement program. Agency managers have the task of identifying and developing appropriate non-monetary incentives that satisfy the totality of needs of the agency’s human capital. Humanistic management values, strategies, and ethos that couple employees’ needs, expectations, and attitudes with employees’ aptitudes, abilities, and knowledge in a particular job context may provide scope for improving employees’ productivity.

MOTIVATING EMPLOYEES

Individuals bring the totality of their needs to the workplace. How agency managers address these needs requires their awareness of, and sensitivity to, employees’ behaviors. Research conducted by behavioral scientists on the underlying motives for individual behavior, links the satisfaction of specific needs to individual motivation. The research results have gained prominence and may provide public managers with some signposts in their search for appropriate non-monetary incentives for public sector employees.

The work of pioneers Abraham Maslow, Elton Mayo, and Douglas McGregor, as well as more recent researchers—Frederick Herzberg, Clayton Alderfer, and David McClelland—has common themes: individual behavior is goal directed, and individuals’ needs

and expectations, if not satisfied, affect their performance on the job.

Maslow found that individuals have a hierarchy of needs. Individuals are continually developing and moving from one level of motivation to the next. When individuals satisfy their *basic* or physiological needs, the higher needs become evident. For example, hungry employees will not perform satisfactorily. Their goal will be to eat, and their motive will be to find food. *Safety* needs, commonly referred to as security needs, are at the next level. These needs are manifested in an employee’s desire for predictability and for protection from physical danger at the workplace. The *social* needs, at the next level, are manifested at the workplace by the employee’s need for secure social relationships. It is arguable that attempts to break up or to control groups at the workplace may lead to antagonistic and unproductive behavior by employees. *Esteem* or egotistic needs, at the next level, relate to our desire to be self-confident, to experience independence, and to be respected, appreciated, and recognized. All employees have this need, which, if properly used in the workplace, may result in higher levels of performance. The need for self-fulfillment or *self-actualization* is the highest level of need. At this level, employees are striving to achieve their full potential. According to Maslow’s approach, when managers recognize that these five levels of needs exist and use them effectively, worker productivity is likely to increase.

More research has disputed the view that an individual’s needs are satisfied in a systematic manner, as implied by Maslow. Clayton Alderfer accepted the view that individuals have similar needs to those Maslow identified. He found that individuals have three sets of needs: *existence*, *relatedness*, and *growth*. *Existence* needs correspond to Maslow’s physiological and security needs. *Relatedness* corresponds to social needs, and *growth*, to esteem and self-actualization needs. By placing these needs along a continuum rather than a hierarchy, employees can therefore deal with two sets of needs at the same time. Alderfer’s growth view holds that when some motive reaches some level of satisfaction, then a new set of motives is triggered.

McClelland’s contribution focused on the individual’s need to achieve. He identified the characteristics associated with individuals who have a high need to achieve. For example, individuals who believe that they

Individual's Goal To Balance Work Life and Personal Life

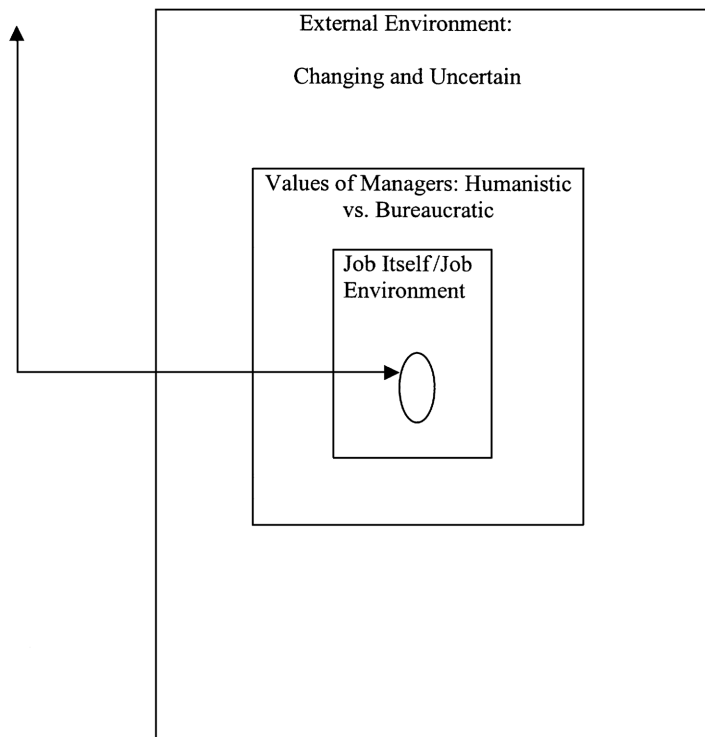


Fig. 1 Sources of employees' needs and expectations.

can make a difference to outcomes have a high need to achieve. These individuals are always thinking about ways to improve outcomes and face "head on" the challenges of solving work-related problems. If agencies stymie the efforts of high achievers, these employees will leave for other agencies that are more accommodating of their need to get things done.

Frederick Herzberg found that the job itself is a potentially motivating force. When individuals are satisfied with their jobs, they are also likely to be satisfied with their job environment.

Elton Mayo's pioneer research, however, found that employees were not simply motivated by the job itself, but by the human aspects of their work. Mayo's "Hawthorne Studies" found that pay and safe working conditions were not the only concerns of workers. Employees in Mayo's studies were also concerned about self-actualization, self-esteem, their ability to be in control of their jobs/to work without close supervision, to problem solve job-related issues when necessary, and to form and maintain informal relationships.

Managers' assumptions about human nature are also likely to affect individual behavior and productivity. Douglas McGregor found that if managers expect employees to be lazy and unproductive, this could be self-fulfilling, as employees are likely to conform to these expectations. Conversely, when employees are assumed to be self-directed, ambitious, and industrious, they are likely to be more productive.

According to Hersey, Blanchard, and Johnson,^a the assumptions that are made depend to a great extent on the values espoused by agency managers. "humanistic-democratic values" underpin managers' assumptions that employees are motivated by affiliation, esteem, and self-actualization needs. "Bureaucratic values" underpin managers' views that employees are motivated solely by their basic needs and by the conditions under which their jobs are performed. These values affect staff outputs and outcomes.

The totality of an employee's needs is complex and wide ranging. These needs may relate to the job environment, the job, or to management's attitudes. An employee needs a quality of work life that provides a comfortable, healthy, and safe work environment. An employee also needs interesting and satisfying work—to be an active participant in solving work-related problems and to be recognized, respected, and provided with opportunities to achieve and be competent.

Individual employees have different needs at different times. These needs, or most of them, have to be identified and satisfied. Managers, therefore, play a critical role in improving productivity. Their attitudes and approaches to managing their agencies' most valuable asset, the human capital, are determinants in productivity improvement. Their approach requires

^aFor detailed discussion see Ref.^[1].

Table 1 Employees' needs and expectations

Source	Needs and expectations
Individual	Balance work life and personal life
Job related	Work that is challenging, satisfying, and empowering
Job environment related	Quality of work life that encourages productivity
Management/supervision	Use talents, be creative/innovative to accomplish work goals
External environment	Agency awareness of cultural differences and of employees' personal needs in a changing environment

that they show concern for individuals' expectations and needs, as well as provide the incentives that satisfy these needs. The social and managerial dimensions of an agency's productivity improvement program may offer some guidance to managers.

PUBLIC PRODUCTIVITY PROGRAMS AND MOTIVATIONAL INCENTIVES

There are four facets of a public agency's productivity improvement program, according to Holzer and Gabrielan.^b Two of these facets are the social and managerial dimensions. In terms of linkages to motivational incentives, both of these aspects of the program focus on the human element and human potential for achieving productivity.

The social and managerial dimensions of public productivity programs present challenges and offer opportunities to public sector managers. Holzer and Gabrielan^[2] cite two social factors that challenge the managerial capacities of public managers: cultural differences and bureaucratic pathologies. Managers' awareness of employees' cultural differences is likely to affect employee productivity. When managers insist that employees, who are Seventh-Day Adventists, come to work on Saturdays, this is likely to have the opposite effect of improving such employees' productivity. Saturday is their Sabbath Day. Agencies that factor in religious holidays and days of prayer recognize and respect existing cultural differences.

Bureaucratic pathologies are evident when public agencies lack openness to change or accept employees' mediocre performance. Such pathologies stifle productivity. By eliciting, recognizing and using the needs of employees, productivity improvement may occur. When employees feel involved because their views are sought and respected, they are likely to be more committed to their jobs and to improving their

productivity. Motivational incentives that include work-force planning, the training and development of employees, and effective utilization of their skills and talents may also contribute to productivity improvement.

Providing employees with feedback through an effective performance appraisal system also results in productivity improvement. Positive feedback has tended to reinforce and improve performance. Exposure of employees to training and development programs serves to improve weak aspects of their performance, to enhance productivity, and is an integral part of an effective performance appraisal system.

The managerial dimension of the productivity improvement program focuses on the public manager's challenging task of implementing comprehensive public management strategies. Appropriate management strategies provide intrinsic rewards: employee recognition, partnerships, and opportunities for employees to achieve.

The managerial dimension also entails opportunities for cooperative partnerships, and for improving employees' competencies and problems of procedural rigidity. In terms of opportunities, Total quality management (TQM) is advocated. This typifies comprehensive public management that focuses on the potential of the human element to achieve productivity, provides non-monetary incentives, and leads to improved productivity. According to Holzer and Gabrielan,^c TQM involves using "theory-based improvement strategies that allow managers to improve capacity." Cohen and Brand^[4] (in Holzer and Gabrielan)^[3] further elaborated on each aspect of this management strategy:

Total = Search for quality in every aspect of work
 Quality = Meeting and exceeding customer expectations
 Management = Developing and maintaining organizational capacity

By embracing this strategy, agency managers demonstrate their awareness that individuals are encouraged and stimulated to perform when the agency

^bHolzer and Gabrielan have developed a framework for analyzing factors likely to affect public productivity improvement programs: The four facets of the framework are managerial, social, technological and informational factors.

^cSee Cohen and Brand Ref. [3].

Table 2 Non-monetary incentives

	Incentives
Individual	Flexible work schedule, elder care and child care facilities, physical fitness facilities
Job related	Recognition: awards, job enrichment, job rotation
Job environment related	Safe, healthy, attractive, and comfortable work environment
Management/supervision	Management values that are humanistic; management strategies that encourage employee involvement and empowerment, “teaming,” employee training, development and competence
External environment	Respect for cultural differences—public agency makes allowances for prayer days, religious holidays; addresses cultural differences with sensitivity and awareness

shows concern for the quality of the product and for the competencies of the employees who deliver the product.

Procedural rigidity may limit employees' abilities to be innovative. Rigidity *straightjackets* the skilled and talented employee who thrives on improving outcomes. Giving this employee room to be creative is likely to improve productivity.

The meaning and purpose of work are changing. Increasingly, employees are attempting to balance their work lives and their personal lives. Many employees believe that money is not all. Some employees are concerned with finding work that gives them the flexibility to manage their private and professional lives: work that both gives them satisfaction and allows them to spend time with their families. Connie Glaser and Barbara Smalley^d in their book, *When Money Isn't Enough: How Women are Finding the Soul of Success*, examined the phenomenon of individuals in the workplace “who are struggling to find meaning, balance and purpose” in their lives and the choices that they make. Individuals, and this apparently is not limited to women, are finding that they need more quality time with their family: their children, their parents, or elder relatives for whom they care. Employees are factoring these needs into their work expectations. Consequently, the provision of on-site child and elder care facilities are non-monetary incentives that are valued by these individuals. Employees' health concerns or expectations to have time for physical exercise as part of a daily regimen to maintain their health are sometimes factored into their totality of needs. Public agencies that provide health screenings and gyms are providing non-monetary incentives that are likely to improve productivity. Employees' wide-ranging needs and expectations, as well as possible sources of these needs, are summarized in Fig. 1 and in Tables 1 and 2.

REWARDING EMPLOYEES: IMPROVING PRODUCTIVITY SUCCESS STORIES

The humanistic philosophy focuses on the behavioral aspects of the work lives of employees. It focuses on the fact that employees have complex needs and are encouraged by incentives and stimuli other than money. Innovative public management strategies may enhance public managers' chances of improving productivity. According to Halachmi,^e strategic management is required for productivity improvement. Most progressive public organizations have adopted enlightened human resource management practices, rejecting authoritarian, bureaucratic style. Many public managers recognize that their management approaches have to be more flexible and accommodating because individuals have varying concerns. These managers are also aware that they need to be concerned with the social and psychological needs of their agency's workforce; enlightened human resource management practices are important for productivity.

Rose Ann Roberts^[7] documented numerous non-monetary approaches to productivity enhancement in the workplace and provided examples of the successful application of some of these approaches. She argued that, before introducing these approaches, agencies need to review their external environment and undertake systems or departmental analyses to determine forces that affect agency productivity. She advocated productivity enhancement approaches that include the introduction of organizational ombudsmen, quality circles, job descriptions updates and redesign, job teams, job rotation and team building, flexible job time, employee merit/recognition systems, employee education programs, and child care and elder care. Many of these approaches have been successfully used in the private sector. The public sector has also

^dSee Ref. [5].

^eHalachmi conceptualizes strategic management as a requirement for successful productivity improvement. Strategic management is not a panacea and offers no guarantees, according to him. For detailed discussion readers may refer to Ref. [6].

benefited from employing certain non-monetary incentives. According to Roberts, since 1989, some federal offices across the United States have opened on-site child care centers. In the private sector, these facilities also exist. Where the actual facilities do not exist, some organizations provide assistance with child care expenses, referral services, and counseling. Organizations have successfully used job rotation to improve productivity: At Tektronix Graphics Workstation Division, “jobs were rotated so that everyone became a coordinator for a week.” Roberts also described the case of the 1974 Union National Bank of Little Rock as an example of the value of employee merit/recognition systems: This case involved rewarding workers who increased their individual productivity by congratulating the highest producer on a regular basis. In the case of the bank, performance increased 200% in 6 weeks.

Other cases have also been documented elsewhere. An on-line government study, “HRM Policies and Practices in Title 5-Exempt Organizations (1998),” outlined the use of non-monetary incentives that include the following: transportation subsidies, health screenings, leave banks, leave transfers, access to fitness centers, and child and elder care information or facilities. Employees’ concerns to maintain their health, to be on time, and/or to be close to their young children or the elder family members for whom they care are “live” issues.

Motivating and rewarding public employees requires constant attention to employees’ needs and expectations from the time that the employees enter an agency until employees leave. Human resource

management practices that integrate responsive and innovative strategies may be critical for satisfying these needs and for improving productivity.

REFERENCES

1. Hersey, P.; Blanchard, K.H.; Johnson, D.E. Motivational Environment. In *Management of Organizational Behavior: Utilizing Human Resources*, 7th Ed.; Prentice Hall: Upper Saddle River, NJ, 1996; 88 pp.
2. Holzer, M.; Gabrielan, V. Five Great Ideas in American Public Administration, Idea 5: Program Effectiveness. In *Handbook of Public Administration*, 2nd Ed.; Rabin, J., Bartley Hildreth, W., Miller, G.J., Eds.; Marcel Dekker, Inc.: New York, 1998; 80–92.
3. Holzer, M.; Gabrielan, V. *Handbook of Public Administration*, 2nd Ed.; Rabin, J., Bartley Hildreth, W., Miller, G.J., Eds.; Marcel Dekker, Inc.: New York, 1998; 89 pp.
4. Cohen, S.; Brand, R. Total Quality Management in Government. In *Handbook of Public Administration*, 2nd Ed.; Rabin, J., Bartley Hildreth, W., Miller, G.J., Eds.; Marcel Dekker, Inc.: New York, 1998; 89 pp.
5. Glaser, C.; Smalley, B. *When Money Isn't Enough: How Women are Finding the Soul of Success*; 2000. <http://womencentral.msn.com/career/experts/expertbweek1.asp>. Accessed 6/20/01.
6. Halachmi, A. Strategic Management and Productivity. In *Public Productivity Handbook*; Holzer, M., Ed.; Marcel Dekker, Inc.: New York, 1998; 551–563.
7. Roberts, R. Non-monetary Incentives for Productivity Enhancement. In *Public Productivity Handbook*; Holzer, M., Ed.; Marcel Dekker, Inc.: New York, 1998; 463–486.

Non-Profit Organizations, External Relations in

Mordecai Lee

*Department of Governmental Affairs, University of Wisconsin-Milwaukee,
Milwaukee, Wisconsin, U.S.A.*

INTRODUCTION

A focus on external relations in non-profit organizations assists in understanding that an agency consists of more than the sum of its tangible components, such as employees, finances, and real property. Similarly, the skill sets needed by a non-profit executive extend beyond managing these “in house” components, such as human resources, budgeting, and asset management. An external relations perspective helps emphasize that the totality of a non-profit agency includes people and entities that are outside its formal and legal boundaries, and the duties of a non-profit executive include leading “beyond the walls,” such as through management of external relations.

INCREASING IMPORTANCE

Several recent trends have contributed to the increasing importance of external relations in non-profit organizations. First, the emergence of instant communications technologies permits and facilitates new forms of external communication that previously were cumbersome or expensive. For example, an Internet newsletter to contributors has many financial and temporal advantages over a hard-copy version that is mailed to a constantly changing mailing list.

Second, in an era of increasing privatization, subcontracting, third-party service delivery, alliances, public-private partnerships, and social entrepreneurship, more of the totality of a non-profit agency is outside its formal boundaries of, for example, activities that its salaried employees engage in.

Third, one manifestation of external relations, advocacy, is becoming an increasingly visible activity in the non-profit sector. Non-profit organizations with a primary mission of public advocacy are a fast-growing category within the non-profit sector. Salamon estimated that the number of non-profit civic organizations grew 68% between 1977 and 1992, increased their employment by 48%, and through 1995, increased their revenue by 216%.^[1] For other categories of non-profit agencies, advocacy is an important secondary activity. For example, a 1996 conference on the contemporary role of foundations identified their

increased “need for professional communication capabilities to educate their publics and influence policy debates.”^[2]

Fourth, the Guidelines for Graduate Professional Education in Non-profit Organizations, Management and Leadership issued by the National Association of Schools of Public Affairs and Administration (NAS-PAA) include external relations as one of 14 subjects that should be included in a curriculum.^[3] Similarly, Wish and Mirabella listed public relations—within the larger rubric of “fundraising, marketing and public relations”—as one of the seven major categories of a curricular model of non-profit management education.^[4] A survey of faculty, practitioners, and students regarding the skills needed by non-profit administrators listed public relations as one of eight “most important” core subjects.^[5]

Fifth, management theorists have increasingly emphasized the importance of external relations. It is a specialized staff function within large organizations as well as an operating concept shared by the agency’s management. The higher managers move up the hierarchy, the more they deal with external issues and constituencies. This necessitates a greater need for having and using public relations skills. In the private sector, corporate CEOs are now estimated to spend 25 to 75% of their time on external relations.^[6–9]

SCOPE

There is a lack of uniformity of nomenclature for this subject matter. External relations, as an overarching moniker, is sometimes also called public affairs, public relations, external communication, community relations, or governmental affairs.^[10] The critical decision point is what external relations encompasses. A literal definition of external relations would encompass all interactions between the salaried employees of a non-profit agency or its board members with any other individuals or organizations. This, however, would include activities that are the core mission of a non-profit agency, such as the marketing and delivery of its services to clients or its relationships with donors and volunteers. It would also cover business activities, such as purchasing and vendor relations. These latter

activities are generally ministerial in nature from the perspective of external communications strategies. When external relations is defined to cover the totality of the agency's out-of-house contacts, it then loses any analytic or applied utility. Focusing on everything is tantamount to highlighting nothing.

Similarly, a broad approach to external relations, as defined by the NASPAA guidelines, would include all advocacy organizations. However, for these non-profit agencies, external relations is their *raison d'être*.^a The study of their external relations would be tantamount to a review of the entire agency's operations. In this situation, too, the prism of external relations adds little to the observer or practitioner.

A pragmatic approach to external relations focuses on those discrete activities that are important to the life of a non-profit agency but are ancillary to the agency's central mission and purpose as well as its ministerial external business relationships. External relations covers secondary activities of an agency that are important to its overall mission but are neither its central nor primary purpose. It contributes to broad, amorphous goodwill toward the agency on behalf of the public at large and other more selected populations.

External relations is defined here as an agency's secondary out-of-house activities, whether planned or unplanned, that affect its standing with the public at large and with specific groups, exclusive of its ongoing external relationships with clients, volunteers, and funders. Specifically, external relations includes media relations, lobbying and advocacy, citizen relations, community relations, and crisis management. An effective external relations operation must be fully integrated into the primary mission of the agency, by working to further the agency's strategic plan, as well as its marketing, development, volunteer, and employee training programs. For more information on conducting all aspects of external relations, see Refs.^[12,13].

MEDIA RELATIONS

The most common purpose of non-profit external relations is to deal with reporters from the news media. This can be reactive or proactive. As a reactive activity, agency staff handle inquiries initiated by journalists covering a story that has some connection to the agency, such as its programs, financing, operating policies, or expertise. When answering media inquiries, two approaches are possible for larger agencies with specialized media relations staff. Some agencies prefer to rely on their expert staff to deal directly with the media.

This is a way to protect less expert employees of the agency from common mistakes when talking with a reporter. Using this approach, the spokesperson for the agency is the media relations professional. Other agencies use the media relations staff as facilitators. With this latter approach, staff focus on connecting the reporter to the optimal agency employee for that particular issue. In these situations, instead of having one person as the agency's permanent "face" and "voice," the agency has multiple spokespersons.

With proactive media relations, a non-profit agency initiates an interaction with the media, in an effort to persuade reporters or editors that a specific story idea is worth covering. In these cases, the reason for contacting the reporter is usually structured around an upcoming agency event, such as an unveiling of a new program, the announcement of a major contribution or the laying of a cornerstone of a new facility. These events are planned and intended to make the agency look good. Coverage is unpredictable and often dictated by the press or absence of other "breaking news" in the same news cycle as the agency event.

The scope of an agency's media relations activities should not be limited to the traditional mass media of daily newspapers, commercial television stations with news departments, and commercial radio stations with news departments. In addition, non-traditional mass media can be a further means for the agency to promote its message. The latter can include weekly newspapers, shoppers, public television and radio, student newspapers, radio stations, regional magazines, organizational newsletters, and specialized publications intended for such groups as businesses, women, residents of a particular neighborhood, or hobbyists.

Sometimes, a non-profit agency needs to bypass the media so that the public at large or a specific audience hears the agency's message despite a lack of coverage by the traditional mass media. In that case, communication activities fall under the rubric of citizen relations, rather than media relations (see below). For more information on conducting media relations, see Refs.^[14-16].

LOBBYING AND ADVOCACY

This area consists of two separate but related activities. First, lobbying (sometimes called government relations) describes the direct efforts by a non-profit agency to affect public sector decision making. This can occur at the local, state, or federal level and can be targeted toward the legislative branch, the chief elected official of the executive branch (mayor, county executive, governor, or president), or to a government agency. This form of lobbying focuses on issues, not

^aAn example of an advocacy organization is a Jewish community relations council. See Ref.^[11].

direct funding of the agency's programs. Hence, lobbying by the American Heart Association of Congress and the president for increased federal funding for heart research would be considered part of the agency's external relations activity. On the other hand, lobbying by the association of a specific federal agency for grants and contracts to fund the association's direct services would generally not be considered external relations.

The difference between the two examples is that the former reflects the disinterested views of the organization, while the latter concerns its self-interest. This distinction is the effect of the definition of external relations used here, namely, that it is not about the core mission and in-house operations of an agency, such as its ongoing relationships with clients, volunteers, and funders. This latter activity involved in government funding is sometimes called program funding, grantsmanship, or contract management. In addition, as mentioned earlier, this overview of external relations is not applicable to non-profit agencies whose main purpose is issue advocacy.

The second activity by a non-profit agency under this rubric is usually called grassroots advocacy. This is different from lobbying, because its focus is on efforts by the organization to affect public-sector decision making indirectly. A non-profit agency sometimes tries to mobilize its existing networks of volunteers, donors, and other supporters to express their opinions to a particular government official, agency, or institution. Grassroots advocacy activities can also be aimed at the public at large, trying to mobilize any citizen to voice the agency's views to a particular government agency or person.

Grassroots advocacy is sometimes contrasted to "astro-turf" advocacy. The latter describes efforts by for-profit businesses to mobilize public opinion through misleading appeals, front organizations that hide their true identity, efforts that are geographically pinpointed to influence the views of a small number of key elected officials, and the use of advanced technology to forward the calls of individuals from the front-organization's 800 telephone number directly to, for instance, the Washington office of their representative in Congress. For more information on conducting lobbying and advocacy, see Refs.^[17-19]

CITIZEN RELATIONS

A non-profit organization interacts with many selected publics, often identified by their relationship with the agency, such as clients, volunteers, donors, government agencies, and the media. Sometimes it interacts with groups identified by socioeconomic or ethnic status

(such as low income or Hmong), by geography (residents of southern Smith County), or by affinity (animal lovers or HIV positive). Yet, a non-profit agency also interacts with the broad and undifferentiated public at large. The phrase "public relations," in its literal meaning, conveys this activity. However, that specific phrase has evolved to represent a different and generally negative meaning. Nonetheless, in its original meaning, the phrase is helpful, because it identifies a distinct population group that a non-profit agency interacts with, namely, the citizenry as a whole. To avoid misinterpretation, another title for this activity is "citizen relations."

Akin to government agencies and commercial products, a non-profit agency has an image. This is a broadly held set of perceptions about the organization by public opinion as a whole, rather than by a small select group within the general population. For example, a non-profit might be viewed as "classy," "in trouble," "successful," "elitist," "caring," etc. These images evolve slowly (and change slowly) based on the invisible civic network that aggregates individual opinions held by the public at large. Non-profit agencies need to pay deliberate attention to the large and small details that contribute positively or negatively to its public image.

In that respect, citizen relations is brand management. An agency can design its casual interactions with the public along with its formal relations with its specialized publics (clients, volunteers, supporters) to create a consistent image. These can include such details as its name, logo, graphic look, signage, web page, public service announcements, media coverage, etc.

Citizen relations is not involved with the core mission of an agency, such as its customer services and fundraising. Rather, citizen relations is aimed to build goodwill toward the agency with the public at large. It enhances the image of the agency in civil society.

In public administration, citizen relations has a direct impact on the future of an agency. For example, the "favorability" ratings of federal agencies can be measured.^[20] A NASA astronaut, indoctrinated in the agency's sensitivity to public relations, stated bluntly that "public opinion helped determine funding."^[21] An official of a state agency observed that "name recognition is important in gaining support from the state Legislature."^[22] Elected officials want to be viewed as supporting popular agencies and criticizing those with low favorability ratings. This connection between image and funding, while more amorphous, also occurs in the non-profit sector. Agencies that have favorable images with public opinion are more likely to be successful, whether in fundraising, expansion, new programming, or political support. Through a well-planned and comprehensive approach

to citizen relations, agencies can gradually improve their public standing and, therefore, enhance their success.

COMMUNITY RELATIONS

A separate external relations activity is community relations. It is regrettable that the word "community" has become so overused in contemporary times that it has lost most of its previous substantive meaning. Nonetheless, as an organizing concept, it can be helpful in focusing attention on two distinct aspects of a non-profit agency's external relations.

First, community relations refers to populations living in close physical proximity to any of the agency's facilities. By their nature, the locations of a non-profit agency's buildings can create problems and even opposition to their existence. That is why a non-profit agency benefits by directing deliberate attention to the populations in the neighborhoods in which it is located. This might include ongoing neighborhood liaison, frequent contact, a hotline for reporting problems, open houses, advisory committees, and regular newsletters. Given the predominance of the NIMBY syndrome ("not in my back yard") in contemporary times, agencies are well advised to deal proactively with potential neighborhood problems. For example, before any public announcement about plans for a group home, an agency would do well to work intensively with the residents near the site to alleviate the automatic fears these proposals normally provoke.

Second, community relations refers to concentrated relationship building with a particular and identifiable segment of the population that would have a discernable interest in something the agency is doing or planning to do. These are not publics with which an agency routinely interacts for client service, volunteer recruitment, or donor development. Rather, there may be racial or ethnic groups that have a discernable but indirect sensitivity to the agency's activities.

For example, a housing redevelopment agency might have selected a specific neighborhood for its efforts. Certainly, current residents of the neighborhood might have concerns about gentrification pushing up property values to the point that they could not afford to continue living there. That would require the first category of community relations activities. However, there may be other population groups that are not directly affected, but are traditionally concerned about such issues, such as the city-wide African American population, the city-wide Hispanic population, etc. In this example, effective community relations would call for initiating contacts with these demographic groups, although they are not affected, as a whole, by the project in question.

CRISIS MANAGEMENT

All crises have external relations implications, because that is when an agency gets the most attention from the media, government, and the public. Sometimes an emergency is isolated to the agency, such as a major physical disaster at one of its facilities, for instance a fire, or the release of an extremely negative public report, such as an audit or newspaper expose. At other times, the emergency is community-wide and the agency is playing a major remediating role, for example in response to a public health epidemic, a major disaster such as a transportation-related accident, etc.

In all cases, the agency not only needs to respond substantively to the crisis but also have a mechanism to smoothly handle the sudden increased public attention. The guiding principle for these situations is the advertising tag line that "one never has a second chance to make a first impression." This requires significant advance planning, both managerial and logistical, as well as a close integration of senior external relations staff with the agency's decision-making group. For more information on conducting the external relations aspects of crisis management, see Refs.^[23-25].

REFERENCES

1. Salamon, L. *America's Nonprofit Sector: A Primer*, 2nd Ed.; The Foundation Center: New York, 1999; 140-141.
2. Spann, J. *Foundations: Exploring Their Unique Roles and Impacts on Society*; Nonprofit Sector Research Fund, Aspen Institute: Washington, DC, 1998; 19 pp.
3. National Association of Schools of Public Affairs and Administration (NASPAA); Commission on Peer Review and Accreditation. *Guidelines for Graduate Professional Education in Nonprofit Organizations, Management, and Leadership*; 1998; http://www.naspaa.org/guidelines/guide_npft.htm#3 (accessed June 2001).
4. Wish, N.B.; Mirabella, R.M. Curricular variations in nonprofit management graduate programs. *Nonprofit Manag. Leadersh.* **1998**, 9 (1), 99-109.
5. Tschirhart, M. Nonprofit management education: recommendations drawn from three stakeholder groups. In *Nonprofit Management Education: U.S. and World Perspectives*; O'Neill, M., Fletcher, K., Eds.; Praeger: Westport, CT, 1998; 61-80.
6. Ehling, W.P.; White, J.; Grunig, J.E. Public relations and marketing practices. In *Excellence in Public Relations and Communication Management*; Grunig, J.E., Ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, 1992; 357-393.
7. Simon, H.A. *Administrative Behavior: A Study in Decision-Making Processes in Administrative Organizations*, 4th Ed.; Free Press: New York, 1997; 294 pp.

8. Herman, R.D.; Heimovics, D. Executive leadership. In *The Jossey-Bass Handbook of Nonprofit Leadership and Management*; Herman, R.D., Ed.; Jossey-Bass: San Francisco, 1994; 137–153.
9. Grunig, J.E. What is excellence in management? In *Excellence in Public Relations and Communication Management*; Grunig, J.E., Ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, 1992; 219–249.
10. Lee, M. Public information in government organizations: a review and curriculum outline of external relations in public administration. *Public Adm. Manag.: An Interactive Journal* **2000**, *5* (4), 183–214. [www.pamij.com].
11. Lee, M. A Jewish “march of dimes”? Organization theory and the future of Jewish community relations councils. *Jewish Political Studies Review* **2000**, *12* (1–2), 3–19.
12. Booth, M. *Promoting Issues and Ideas: A Guide to Public Relations for Nonprofit Organizations*, 2nd Ed.; The Foundation Center: New York, 1995.
13. Levine, M. *Selling Goodness: The Guerilla P.R. Guide to Promoting Your Charity, Nonprofit Organization, or Fund-Raising Event*; Renaissance Books: Los Angeles, 1998.
14. Salzman, J. *Making the News: A Guide for Nonprofits and Activists*; Westview: Boulder, CO, 1998.
15. Martens, T. *Nonprofits and the Press: Newsmaking Strategies for California*; Nonprofit Sector Research Fund, Aspen Institute: Washington, DC, 1998; http://www.non-profitresearch.org/newsletter1531/newsletter_show.htm?doc_id=16180 (accessed June 2001).
16. Muehrcke, J. The 10 media trends that will drive your future and how to get ready now. *Nonprofit World* **1999**, *17* (2), 36–41.
17. Pidgeon, W.P. *The Legislative Labyrinth: A Map for Nonprofits*; Jossey-Bass: San Francisco, 2001.
18. Smucker, B. *The Nonprofit Lobbying Guide*, 2nd Ed.; Independent Sector: Washington, DC, 1999.
19. Rees, S. *Effective Nonprofit Advocacy*; Nonprofit Sector Research Fund, Aspen Institute: Washington, DC, 1998; http://www.nonprofitresearch.org/newsletter1531/newsletter_show.htm?doc_id=16986 (accessed June 2001).
20. Pew Research Center for the People and the Press. *Performance and Purpose: Constituents Rate Government Agencies*; Released April 12, 2000; <http://www.people-press.org/npr00rpt.htm> (accessed June 2001).
21. Lovell, J.; Kluger, J. *Lost Moon: The Perilous Voyage of Apollo 13*; Houghton Mifflin: Boston, 1994; 155 pp.
22. Associated Press. Historical Society Name Change Doesn't Sit Well With Everybody. *Milwaukee [WI] Journal Sentinel*; June 5, 2001; 3B. UMI Article Re. No.: UAAE-2258-133. <http://proquest.umi.com/pqdweb> (accessed June 2001).
23. Thompson, R.L. Contingency and emergency public affairs. In *The Nonprofit Handbook: Management*, 3rd Ed.; Connors, T.D., Ed.; Jossey-Bass: San Francisco, 2001.
24. Ogrizek, M.; Guillery, J.-M. *Communicating in Crisis: A Theoretical and Practical Guide to Crisis Management*; Aldine de Gruyter: Hawthorne, NY, 1999.
25. Berry, S. We have a problem . . . call the press! *Public Manage.* **1999**, *81* (4), 4–9.

Nuremberg Trials and Charter

Morton J. Tenzer

*Department of Political Science, University of Connecticut,
Storrs, Connecticut, U.S.A.*

INTRODUCTION

The Nuremberg Trials of major Nazi war criminals, properly the International Military Tribunal, were conducted from October 1945 to October 1946. Twenty-four of the leading figures of the Nazi regime were accused before a tribunal of judges from Great Britain, the United States, the USSR, and the provisional government of France under the authority of the Charter established by the four powers on August 8, 1945. One of the defendants was excused for age and health reasons, one committed suicide during the trial, three were found not guilty, 11 were sentenced to death, and the rest were sentenced to various terms from 10 years to life imprisonment.

The trial served many purposes. It publicized fully to the world for the first time the full range of monstrous activities of the Nazi regime, including the extermination of virtually the entire Jewish population of Central Europe and the slaughter of millions of other civilian victims deported for forced labor or deliberately killed for belonging to other categories of humanity, such as Gypsies, gays, political activists, and dissenters of various sorts. The trial established some important precedents: that such enormous crimes could be subjected to a fair and orderly judicial process; that no one, even a head of state, was immune from such judgement; and that conspiring to commit an aggressive war and crimes against humanity were subject equally to such judgement along with more traditional crimes against the laws of war. A fundamental principle was reinforced—that obedience to superiors' orders did not excuse or justify behavior of subordinates who comply with illegal orders. As a result of a subsequent trial at Nuremberg of Nazi doctors who performed hideous experiments on concentration camp prisoners, the Nuremberg Code was adopted to provide guidelines for research on human subjects, now part of the law in the United States. While the trial and punishment of the Nazi leaders did not deter wars from occurring in the subsequent half century, they had a profound effect on the regulations and behavior of public officials whose previous immunities had been stripped by the Nuremberg Charter^[1] and trials.

NUREMBERG CHARTER

At the end of World War I, Kaiser Wilhelm had been allowed to flee unharmed to settle in the Netherlands. No formal culpability for that war was ever established, and various controversies and myths developed about the origins of that war and its conclusion, which contributed to the rise of the Nazi party and its theory of a “stab in the back” that allegedly led to Germany's defeat. The victors over Napoleon in the last general European war a century before 1914 dealt with him summarily, exiling him first to Elba and then to St. Helena without any formal process. As the war in Europe drew to an end in 1945, it was apparent to the Allied leaders that Nazi Germany, in the course of the war, had committed crimes of such unparalleled gravity and extent that its leaders must be punished for their complicity in these crimes. President Roosevelt had warned of such accountability as early as 1942. Churchill wanted Adolph Hitler, the Nazi leader, and his associates shot summarily, but the United States and the Soviet Union favored holding trials, although they had vastly different concepts about how to conduct legal proceedings. There was some fear that Hitler might use his frenzied rhetorical abilities to disrupt a trial. When the war ended in May, 1945, Hitler and two of his closest associates, Joseph Goebbles, the chief Nazi propagandist, and Heinrich Himmler, the head of the SS and secret police, had committed suicide, and Martin Bormann, Hitler's top aide, was missing, although also presumed dead. Benito Mussolini, the Italian Fascist dictator, was also already dead. Of the top Nazis, only Hermann Goering, the regime's number two man, was captured alive. The Allies quickly decided to try Goering and other high-level Nazi officials, and on August 8, 1945, the four victorious powers (Great Britain, the United States, the USSR, and the provisional government of France) adopted the Charter of the International Military Tribunal empowered to conduct trials of Nazi war criminals.^[2]

The Tribunal would consist of four judges, each with an alternate representing the four signatories. Its jurisdiction included the power to try and punish major war criminals of the European Axis powers for crimes defined as follows: Crimes Against Peace, which

meant planning, organizing, or waging a war of aggression; War Crimes, the violation of the laws or customs of war, such as the murder of prisoners of war, deportation of civilians to slave labor, and devastation not justified by military necessity; and Crimes Against Humanity, the murder, extermination, enslavement, and deportation of civilian populations or persecutions on political, racial, or religious grounds. Additionally, the Charter stated that the official position of defendants, whether heads of state or other high rank, would not free them from responsibility for their actions, and the fact that defendants acted on orders of their government or a superior would not free them from responsibility.

Each signatory to the Charter was to appoint a Chief Prosecutor, and the prosecutors acting as a committee were to draw up an Indictment with full particulars specifying charges against the defendants and to collect all necessary evidence for presentation to the Tribunal. The American Chief Prosecutor, appointed by President Truman, was Justice Robert Jackson, on leave from the Supreme Court of the United States. He became chairman of the prosecutors' committee and bore the main burden of prosecutorial activities. The United States had by far the largest staff and played the major role in the collection of evidence, interrogation of prisoners, the running of the prison where the defendants were incarcerated, and the staffing of the courtroom.

The Charter also provided for a fair trial for the defendants by assuring an indictment with detailed particulars of the charges against each defendant, and by providing the defendants with the right to counsel, to present evidence supporting their cases, and to cross-examine witnesses for the prosecution. All documents were to be translated into the defendants' language. The Tribunal was given the power to summon witnesses and documents, and to impose punishment including death upon conviction of any defendant. It was required to explain the reasons for guilt or innocence of each defendant, and its proceedings were to be conducted in English, French, and Russian, as well as German. The permanent seat of the Tribunal was Berlin as the Russians wanted, but the trial and subsequent trials were held in Nuremberg, where the Americans had facilities and staff adequate for the purpose.

INTERROGATIONS

Even before the Charter, interrogations of the Nazi prisoners had begun.^[3] Practical difficulties included the shortage of able translators, the complexities of duplicating transcripts in many languages, the dispersion of the prisoners before they were brought to

Nuremberg, and the resistance of the prisoners to admitting guilt. The major difficulty, however, was the interrogators' lack of knowledge and understanding of the organization and mechanisms of the totalitarian Nazi regime. The prosecution was learning about the system as they were proceeding to draw up the indictments, and the early notions of a simple, tightly run dictatorial structure gave way to an understanding, perhaps never fully achieved, of the multiple layers of authority, duplicative agencies, and unclear lines of responsibility that characterized the Nazi system. It was difficult to coordinate the information received from hundreds of other prisoners, victim witnesses, and the voluminous documents that had fallen into Allied hands and required translation and incorporation in the prosecution's case.

Aside from Goering, Rudolph Hess (Hitler's longtime Nazi party associate) and Joachim Von Ribbentrop (the Reich Minister of Foreign Affairs) and a few others, it was not clear who should be tried as a major war criminal. Preparing the case was a logistic nightmare, made more difficult by uncertainty about the nature of the charges and the identity of the defendants. Work which should have taken months, if not years, had to be compressed into weeks, because the Allied Powers were determined that the trial be held as quickly as possible in the autumn of 1945.

INDICTMENT

After considerable debate among the Allied prosecutors concerning the identification of the defendants who were to be charged as major Nazi war criminals and the nature of the accusations against them, the prosecutors, working under severe time constraints, presented the indictment to the Tribunal on October 19, 1945, in Berlin. The trial was scheduled to begin on November 20 in Nuremberg, allowing the defendants a month to secure counsel and prepare their defense.^[4]

The indictment was innovative in at least two respects. First, following Anglo-American common law tradition, the defendants were accused in the first count of the indictment of being guilty of a *conspiracy* to commit the crimes against peace, war crimes, and crimes against humanity specified in the Charter. Second, several major organizations were indicted for their roles in the crimes of the individual defendants and were to be declared criminal organizations if found guilty. Membership in these organizations would create a presumption of guilt for defendants in future trials. These groups were the Reich Cabinet, the leadership corps of the Nazi Party, and police, security, and military staff units. While many of the defendants had multiple roles—Goering was a party leader, a cabinet

member, head of the four-year plan for economic development, a general in the SS, and Commander-in-Chief of the Air Force—they represented different elements of the Nazi power structure. Some were primarily party leaders, and others had been foreign policy leaders. Several had operated under Himmler in the persecutions and exploitation of slave labor and in the extermination program. Others had key roles in the German economic war effort and represented German business and finance involved in Nazi activities. Four were military officers of the high command, and others had held scattered positions in the government. There were 21 defendants who attended the whole trial and testified in their own cause. The indictment spelled out in great detail a gruesome account of the depredations the Nazi regime had committed in Europe. It referenced treaties broken, wars launched, numbers of transports and deportees from various countries, cities and towns destroyed, looted and expropriated property, and millions of Jews exterminated, as well as millions of others worked and starved to death.

THE TRIAL

Between November 20, 1945, and August 31, 1946, there were 403 open sessions of the Tribunal.^[5] Testifying for the prosecution were 33 witnesses and 61 for the defense, in addition to the defendants. Thousands of documents and affidavits were received, nearly overwhelming the ability of the staff to translate and collate all of the material.

The evidence against most of the defendants was overwhelming, as the prosecution presented through documentation and witnesses the details of the barbarities committed by the Nazi regime from its inception. Some unusual pieces of evidence were used. A specially produced four-hour film, *The Nazi Plan*, was shown on December 11. Tattooed human skin tanned at the Buchenwald concentration camp and a shrunken head of a Polish youth from Auschwitz were displayed along with photographs, charts, and tables illustrating elements of the prosecution's case.^[6] Perhaps the two most extraordinary witnesses were Dieter Wisliceny, one of Adolph Eichmann's assistants who testified about the transport of millions of victims to the camps, and Rudolph Hoess, the commandant of Auschwitz who admitted supervising the gassing of two million persons between the summer of 1941 and the end of 1943. Wisliceny said that Eichmann, who had operational responsibility for the extermination of the Jews, had shown him a written order from Himmler asserting this was the Fuhrer's (Hitler's) decision.^[7]

The visual and verbal evidence of the crimes of their regime shook the defendants. All present had pleaded not guilty. Goering, as in his interrogations, remained

the most arrogant and unapologetic. Most of the others tended to deny their own responsibility for the crimes and claimed that they did not know what the others had done. The military men were more willing to acknowledge responsibility, but they denied knowledge of atrocities, and haggled over details. Continued admiration, if not adulation, for Adolph Hitler was expressed by almost all the defendants who claimed only to be following his lead, inspiration, and orders. The denials of guilt were palpably false and the refusals to accept responsibility a strategy to avoid conviction and a possible death penalty.

The defense raised an important moral issue: what was the responsibility of men who served an evil government which punished dissent with violent ruthlessness? Can all the blame be shifted to the top? Does obedience to superiors excuse complicity? If one is resigned to events, detached emotionally from actions, is this acquiescence in a criminal sense? The generals and admirals were ready to assume responsibility for their actions but not for crimes committed under their command. At most, they would admit to having made mistakes. Von Ribbentrop agreed to political responsibility but not criminal responsibility, and others insisted on their innocence and claimed to have opposed many of the Nazi initiatives.

JUDGEMENT

The Tribunal rendered its judgement^[8] on September 30 and October 1, 1946. Of the organizations, the Reich Cabinet, the General Staff and High Command, and the SA were acquitted. Twelve defendants were sentenced to death by hanging. Three were sentenced to life imprisonment and four others to various lesser terms of imprisonment. Three defendants were acquitted. Arriving at these decisions had not been without controversy among members of the Tribunal, and the Soviet member formally dissented from the three acquittals and the life sentence of Hess, whom he believed deserved the death penalty. He also dissented from the decisions about the Reich Cabinet and the General Staff and High Command, which he thought should be declared criminal organizations along with the leadership corps of the Nazi party, the SS, the SD, and the Gestapo.

SIGNIFICANCE

The Nuremberg Trials had widespread ramifications. Although they have been criticized as "victor's justice," they established important precedents that bear on the behavior of states in international politics, and on individuals, especially government officials throughout the world. The Trial demonstrated that a judicial process

could be conducted to hold government officials, even heads of state, responsible for criminal actions taken leading to war and during war. Aggressive war, defined at Nuremberg, was construed as a crime violative of treaties going back to the Hague Conventions at the turn of the twentieth century. Activities undertaken even in the stress of war under direct orders of superiors can be construed as war crimes and crimes against humanity. Obeying orders is an excuse but in no way a justification for inhumane behavior.

Nuremberg Code

Subsequent proceedings held at Nuremberg tried over a hundred lower-level Nazi officials. Of these, the most significant was the Doctors Trial, in which a number of German doctors who had performed medical experiments on concentration camp inmates were put on trial. The almost unimaginably cruel nature of these sadistic experiments, whose subjects almost invariably died or were killed after their tortures, led the judges at the trial to include in their verdict a section called “Permissible Medical Experiments,” which became known as the “Nuremberg Code.”^[9] The Code became a landmark document in medical ethics and has become parts of the law and regulations of research on human subjects in the United States and elsewhere. That voluntary consent of the human subject is absolutely essential, the first point of the Code, has become the basis of all human subjects research, even in the social sciences. Other points in the Nuremberg Code bar research that causes unnecessary physical and mental pain or injury and forbids research when it is reasonable to expect death or disabling injury will result.

Nuremberg Principles

In 1946, the new General Assembly of the United Nations approved the Nuremberg Charter and appointed an International Law Commission to formulate principles derived from the Trial. It was not until 1950 that the General Assembly received the Commission’s report and adopted the seven so-called “Nuremberg Principles,” which basically restated much of the Charter and Indictment with regard to crimes against peace, war crimes, and crimes against humanity. The principles include the vulnerability of heads of state and government officials to the application of international law; that action under orders does not relieve responsibility for criminal acts; that accused have the right to a fair trial; and that complicity in crimes against humanity is punishable by law.^[10]

In 1948, the United Nations had adopted the Universal Declaration of Human Rights which

stemmed in part from the findings of the Nuremberg Trials. Article 3 states the right of everyone to “life, liberty and security of person.” Article 5 states “no one should be subject to torture or to cruel, inhuman or degrading treatment or punishment.”^[11] Also in 1948, the Convention on the Prevention and Punishment of the Crime of Genocide was signed in Geneva, providing a more definite legal basis for action against what was called the “odious scourge” of genocide, defined as “acts committed with the intent to destroy in whole or part, a national, ethnical, racial or religious group, as such.”^[12] The word genocide had not come into common use during the Trial in 1945, but clearly it means much the same as the Crimes against Humanity of the Nuremberg Charter.^[13]

For decades after the Nuremberg Trials, their direct influence on the behavior of states and public officials was minimized by the Cold War. The subsequent trials and de-Nazification program in Germany were prematurely ended. The Gulag system of punishment and labor camps in the Soviet Union, in which millions of people were incarcerated and died, continued unchecked until modified by Nikita Khrushchev in 1953 and was not ended until Mikhail Gorbachev’s rule in the 1980s. The indirect effects of Nuremberg, nevertheless, became widespread throughout western countries and the United States, evidenced by the increased sensitivity of public officials to their treatment of citizens and other inhabitants of their countries. Capital punishment was abolished in Europe over the decades.

In recognition of the responsibility of subordinates to disobey illegal orders of superiors within a government administration, federal- and state-level statutes protecting “whistle blowers” were adopted in the United States. Even the ombudsman movement, creating a position within agencies to mediate disputes between citizens and officials and among administrators, although originating in Sweden, was a reflection of the concerns raised in Nuremberg about the callous indifference of Nazi officials to the rights of the people with whom they dealt. The development of government guidelines for research on human subjects enforced by the United States Department of Health and Human Services, including the creation of Institutional Review Boards at hospitals, universities, and other research centers, was based on the Nuremberg Code.

The direct implementation of the Nuremberg Principles and the Genocide Convention languished for decades. In the 1970s, under the Carter Administration, a renewed emphasis was placed on human rights in American foreign policy. In 1979, the Office of Special Investigations in the Justice Department was created to search for Nazi war criminals who may have entered the United States with false credentials after the war. This office’s work led to 54 deportations, and continued as late as 2002.

When the focus was on the struggle with the Soviet Union, there was little reference or recourse to the Nuremberg Principles or the Genocide Convention, and atrocities occurred from Algeria to Cambodia without significant response by the international community. Some minor criminal prosecutions occurred such as the trial of the officer responsible for the My Lai massacre in Vietnam. The United States, a signatory to the Genocide Convention in 1948, did not ratify it for four decades until 1988.

A major shift in attitudes became evident in the 1990s with the end of the Cold War. In 1998, the former dictator of Chile was arrested in London for crimes committed in Chile under his rule, and the highest court in England held him eligible for trial despite the fact that he had been head of state. The genocidal activities in the former Yugoslavia led to the trial of Serbs and Croats in 2000 for war crimes at The Hague, and the chief defendant was Slobodan Milosovic, the former president of Yugoslavia. War crimes trials were held in Rwanda to punish the perpetrators of the genocide committed by the Hutu inhabitants of that country against the Tutsi minority, and in 2002, Indonesian military leaders went on trial for atrocities committed by their forces in East Timor. At Rome in 1998, a treaty was signed by 120 countries, including the United States, to establish an International Criminal Court that would prosecute individuals for war crimes, genocide, and other gross human rights violations. On April 11, 2002, the United Nations announced that 66 nations had ratified the treaty, bringing the court into existence and replacing the ad hoc system of tribunals that had begun at Nuremberg.^[14] More than 50 years after they were conducted, the Nuremberg Trials had become a vital precedent for the prosecution of government leaders and officials for crimes against peace, war crimes, and crimes against humanity.

CONCLUSION

The Nuremberg Trials of 1945–1946 punished many of the leading figures in Nazi Germany's attempt to

conquer Europe and exterminate entire populations. The trials served as an inspiration for the adoption of international treaties and conventions, UN resolutions, and national laws regulating the functions of public administration and the behavior of public officials. Many decades passed before the power of the Nuremberg Trials as a precedent became evident, but at the turn of the twenty-first century, war crimes trials dealing with events in Europe (Yugoslavia), Africa (Rwanda), and Asia (Indonesia) demonstrated the viability and relevance of Nuremberg again.

REFERENCES

1. <http://www.Yale.edu/lawweb/avalon/imt/proc/intconst.htm> (accessed March 6, 2002). Also see Marrus, Michael. *The Nuremberg Trial, 1945–46: A Documentary History*; Bedford Books: Boston, 1977.
2. Conot, R.E. *Justice at Nuremberg*; Harper & Row: New York, 1983; 9–24, 197–199.
3. Overy, R. *Interrogations: The Nazi Elite in Allied Hands, 1945*; Viking: New York, 2001.
4. <http://www.yale.edu/lawweb/avalon/imt/proc/count.htm> (accessed March 17, 2002).
5. Taylor, T. *The Anatomy of the Nuremberg Trials: A Personal Memoir*; Knopf: New York, 1992.
6. Conot, op. cit., 197–199.
7. Arendt, H. *Eichmann in Jerusalem*; Viking: New York, 1963 and Hoess, R. *Commandant of Auschwitz*; World Publishing: New York, 1960.
8. <http://www.yale.edu/lawweb/avalon/imt/proc/judent.htm> (accessed March 18, 2002).
9. http://www.ushmm.org/research/doctors/code_expl.htm (accessed March 6, 2002).
10. Knoll, E., McFadden, J.N., Eds.; *War Crimes and the American Conscience*; Holt: New York, 1970; 178–180.
11. <http://www.un.org/overview/rights.htm> (accessed March 19, 2002).
12. http://www.unhchr.ch/html/mem3/b/p_genoci.htm (accessed March 19, 2002).
13. Kuper, L. *The Prevention of Genocide*; Yale University Press: New Haven, 1985.
14. War Crimes Tribunal Becomes Reality, Without U.S. Role, NY Times April 12, 2002, Vol. CLII, A3.

Offset Requirements in International Public Procurement

Travis K. Taylor

Department of Accounting, Economics and Finance, Christopher Newport University,
Newport News, Virginia, U.S.A.

INTRODUCTION

Consider the following international transaction between a multinational corporation and the government procurement office of Australia. In an effort to win the bid to sell new aircraft to Australia, an American company made arrangements to provide extra (legal) benefits to the Australian economy. Among the benefits this included the marketing and sale of Australian lollipops from Sweets Ltd to the U.S. navy. This arrangement, which actually took place in the early 1990s, is formally known as an offset contract. Common in both developing and industrialized countries, the offset arrangement is one of the most important innovations in international public procurement policy. This article describes the legal framework of offset contracting and employs the tools of the *new institutional economics* to explain why governments are electing to leave the price margin for some types of purchases.

WHAT IS AN OFFSET?

Offsets are contracts that require the seller to transfer extra economic benefits to the buyer as a condition for the sale of goods and services. In many instances, governments prefer to realize these benefits in the form of in-kind transfer, instead of bargaining for discounts of the sale price.

Although few outside those actually engaged in the practice of offsets are aware of this contractual arrangement, it has become an increasingly popular tool of government procurement. Common in exchanges involving aerospace goods and services, offsets permeate the market for defense systems and weapons. These arrangements also arise when governments purchase telecommunications equipment, computers, and a myriad of other goods that need not embody high technology.

Offsets often appear under the guise of *compensation packages, industrial benefits programs, cooperative agreements, and countertrade policy*. The arrangement, at its most fundamental level, is straightforward: a purchasing government obliges a foreign seller to include extra benefits with the sale of the base good. The foreign firm may then sign individual offset contracts

with local firms in the purchasing government's local economy.^a Aside from this buyer-seller relationship, offsets vary considerably in form. Offset transactions may take, but are not limited to, any of the following forms: subcontracting, technology transfer, countertrade, foreign investment, marketing assistance, training, co-production, and licensed production.^b

One can make a further distinction between direct and indirect offsets. Direct offsets call for work that relates in some fashion to the good purchased by the government. Indirect offsets require the seller to transfer some form of economic benefit to the purchasing country that is unrelated to the purchased good.^c The hierarchical structure of offset design is shown in Fig. 1.

Consider two different government procurement examples. In 2003, the American firm Lockheed Martin announced the sale of 48 F-16 aircraft to Poland for \$6.023 billion. As part of the contract, Lockheed Martin agreed to provide Poland with one of the more lucrative offset packages ever recorded. The company would create—over the course of 10 years—economic benefits (jobs, technology transfer, and so forth) in Poland worth \$9 billion.^[1] The basic offset arrangement is depicted in Fig. 2.

An example of a civil offset arrangement involves AT&T's successful expansion into the Middle East. To win the bid to supply a telecommunications infrastructure for Saudi Arabia, AT&T included offsets in 1994. AT&T agreed to provide billions of dollars in new work and technology transfer for the Saudi economy.^[2] By 1998, AT&T had fulfilled most of its obligations by teaming with local Saudi businesses to form equal equity joint ventures. The joint ventures are now successfully supplying production inputs to AT&T (and spin-off Lucent Technologies) and other export markets.

We call offsets non-standard contracts because they require *conditionality* and *additionality*.

^aThe contracts may or may not designate specific local firms with whom the foreign seller will work, depending on the particular offset agreement. Contractual design varies considerably because offsets can accomplish multiple objectives.

^bCountertrade includes barter, buyback, and counterpurchase transactions.

^cIn offset language, researchers and practitioners usually refer to the purchased good as the "base good" or "base exchange."

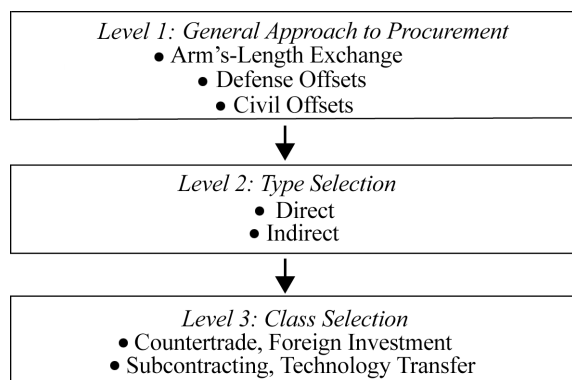


Fig. 1 Hierarchical design of offset policy by a purchasing government.

Conditionality means that the buyer will only purchase from the seller if the seller provides economic benefits to the buyer. This is, effectively, a reciprocal exchange condition. Offsets also require the seller to demonstrate that the new business transferred to the buyer's country would otherwise not occur. This is the additionality requirement. Competition with offsets involves price-quality-content trade-offs, thereby adding another dimension to the price-quality margin of most markets.

Although Organization for Economic Cooperation and Development (OECD) international trade statistics do not officially track offsets and other compensation agreements, studies show that the market is sizeable. For example, estimates of countertrade as a percentage of world trade range from a conservative 8% to a high of 20%.^[3] Countertrade that is a result of government procurement requirements qualifies as offset, whereas an agreement between two private firms does not. Most aerospace and defense goods sold to foreign governments include offsets ranging from approximately 10% to 150% of the sale price. In the U.S. defense industry alone, 42 U.S. companies signed 513 new offset arrangements between 1993 and 2004 valued at \$55.1 billion.^[4] These data indicate that offsets accompanied the sale of 173 U.S. defense products during the period (Table 1). The actual number of offset contracts is far greater than 173 because the seller typically agrees to provide offset work to numerous firms in the purchasing government's economy.

Moreover, offset contracts are increasing in use. The number of non-Soviet bloc governments officially using industrial cooperation and compensation policy increased from 17 to 78 between 1964 and 1986.^{[5]d}

^dIt has long been recognized that Soviet bloc countries engaged in various forms of offset and countertrade policy during the post-World War II years. Their experience is reviewed in the historical development of offsets chapter in Taylor (2001). (Taylor, T.K. 2001. Ph.D. Dissertation. University of Connecticut.)

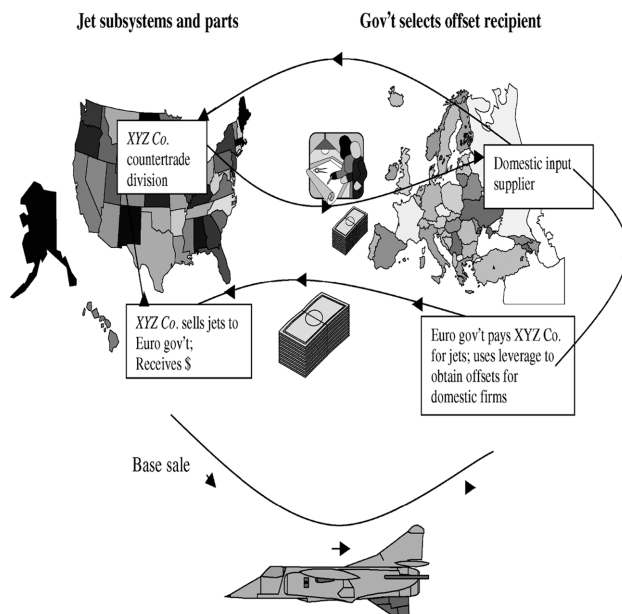


Fig. 2 Simplified offset arrangement.

Wood reported that approximately 130 countries maintain offset offices or agencies with their respective governments.^[6] Yet, for a policy instrument of such apparent importance, there is a dearth of comprehensive examinations.

LEGAL FRAMEWORK

The offset is an extra sweetener—some form of economic activity beyond the base exchange—that governments require or encourage when accepting bids for procurement orders. *Defense* offsets arise when a government purchases military goods from a foreign seller. If the government requires offsets for the procurement of non-defense goods, it is known as a *civil* offset program. Although not the norm, civil offsets are gaining popularity despite their illegality under World Trade Organization (WTO) laws.^c Article XVI of the WTO's Fair Trade Practices prohibits offsets not associated with military purchases. Governments use several strategies to circumvent this article.

Article XVI is virtually impossible to uphold. Offset is one of the more nebulous policy instruments in government. In fact, many procurement requirements never mention the word "offset," preferring the more innocuous and vague "compensation packages," "cooperative agreements," and "industrial development." Another technique is for the purchasing

^cThe data show that approximately 10–20% of the offsets originate from governmental purchases of civil goods and services.

Table 1 Summary of U.S. defense offset arrangements, 1993–2004 (\$ millions)

Year	U.S. companies	Agreements	Base export value ^a (\$)	Offset value (\$)	Offset ratio ^b %	Countries
1993	17	28	13,935	4,784.4	34.3	16
1994	18	49	4,792.4	2,048.7	42.7	20
1995	20	47	7,529.9	6,102.6	81.0	18
1996	16	53	3,119.7	2,431.6	77.9	19
1997	15	60	5,925.5	3,825.5	64.6	20
1998	12	41	3,029.2	1,768.2	58.4	17
1999	10	45	5,656.6	3,456.9	61.1	11
2000	10	43	6,576.2	5,704.8	86.7	16
2001	11	34	7,017.3	5,460.9	77.8	13
2002	12	41	7,406.2	6,094.8	82.3	17
2003	11	32	7,293.1	9,110.4	124.9	13
2004	14	40	4,927.5	4,329.7	87.9	18
12 year total	42	513	77,208.6	55,118.5	71.4	41

^aBase export value is the summation of U.S. sales to countries who require offsets

^bOffset ratio % is the compensation ratio defined as the (offset value/base export value) (From page 33 of Ref.^[4].)

government to consider, although not require, offsets with bids. In other words, the government does not mandate offsets, but it is highly recommended. In some cases, procurement officials will not consider bids without offsets.

The Czech government's procurement policy is illustrative. When it purchased a fleet of jet aircraft in 1999, offsets played a surprisingly large role. In reviewing bids from aerospace suppliers, the Czechs reportedly placed most importance on the offset package. Offset considerations made up 50% of their decision whether to buy from a seller, followed by technical specifications (30%) and financial factors (20%).^[7] This procurement policy departs from the price margin: the content of offset packages is supplanting price and quality competition of the base good.

Because procurement officials hope to achieve multiple and diverse objectives, it is not surprising that offsets vary considerably in design. A general model of offsets would, ideally, take into account the economic and political factors that codetermine policy. The following equation reflects these political economy variables:

$$\text{Policy} = f(\text{development strategy, transaction costs, market structure, political institutions, national security, bureaucrat maximization})$$

WELFARE EFFECTS OF OFFSETS

Governments know little about the welfare effects of offsets. Hooper noted that offsets are “a complex subject about which very little is known—they have

been described as smoke and mirrors . . . nobody knows who benefits.”^[8] Many governments view offsets as a beneficial tool for economic development. Development strategy today focuses on acquiring economic capabilities necessary to compete in the global marketplace. Offsets arise from the government's bargaining power in procurement and the belief that, for various reasons, it is more feasible to acquire these capabilities and benefits through non-standard contracting than in markets. However, although the government's purchasing leverage may enable it to extract rents from suppliers in an imperfectly competitive environment, there is no free lunch.

Economists are understandably skeptical when they observe any government intervention that subverts the invisible hand in markets. Orthodox economic theory has long vouched for the superior efficiency of markets and money as a medium of exchange (see, e.g., Refs.^[9,10]). In traditional arms-length exchange, competition in terms of price and quality drives buying and selling decisions. Goods and services exchange for money, which allows market participants to disassociate current production from consumption.

In offset arrangements, the buyer and seller may conduct a considerable portion of the procurement contract off the price margin. Offsets are thus susceptible to the double coincidence of wants inefficiency of barter. In addition, these arrangements are subject to extra administrative and bureaucratic costs that are less acute in markets. That said, the obvious question is “why do offset contracts persist today”? Studies demonstrate that offsets continue to flourish, in part, because they help build economic capabilities and are politically attractive.^[1,11,12]

OFFSETS AS A CAPABILITIES ACQUISITION STRATEGY

Economic capabilities are the skills and competences of a firm. The capabilities view considers the firm to be a pool of core and ancillary competences that become routinized over time to perform production activities.^f The ability of firms to grow capabilities internally and acquire capabilities externally explains many successes and failures in international markets.^[17]

The boundary of the firm—while always changing—depends on the transaction and governance costs of purchasing competences in markets, and the relative production cost differences between the firm and market competitors.^[18] The critical comparisons are therefore between the cost of cultivating capabilities internally, purchasing capabilities on the market, and purchasing final goods and services (off the shelf) in markets.

The government's aim in many countries is to promote the acquisition of the capabilities required to compete in high value-added markets. Most countries must purchase a majority of its information technology, aerospace, and defense systems in international markets because they have neither the budget nor the political support to arrange for indigenous production. Purchasing items off the shelf in markets is unlikely to transfer the capabilities required to produce these products. Ultimately, governments must decide whether it is more desirable to purchase capabilities (technology, skills, and routines) directly in markets, and/or develop capabilities over time by encouraging strategic alliances.

Many countries employ offsets as part of a broader economic development strategy. The offsets usually target high value-added industries and hasten the speed of technology dissemination from leaders to laggards. The main objective is to raise the level of economic capabilities in the targeted industry and, more generally, in the country.

We can view offsets as a government tool to enhance dynamic capabilities. Many offset arrangements transfer technological know-how, teach idiosyncratic routines, and forge long-term strategic alliances. Numerous studies have shown that these alliances “can be vehicles for organizational learning, helping

firms to recognize dysfunctional routines, and preventing strategic bottlenecks.”^{[16]g}

An implication of this reasoning is that *building* capabilities with offsets may be superior to anonymous, one-time *purchases* of technology in markets.^[11] For example, buying an advance-warning-aircraft-system off the shelf.^h may transfer some basic technologies and knowledge of maintenance operations; however, it is unlikely that the transaction will aid the purchasing country in the cultivation of new core competences. Long-term joint ventures cultivated through offsets may bear more fruit as multinational corporations are induced to teach the recipient cost-minimizing techniques.

BUNDLED GOODS

Another strand of research treats offsets as bundled goods.^[12] In this approach, the welfare effects of offsets may be either positive or negative, depending on the presence of economies (diseconomies) of scope and scale. Just as bundling can be shown to offer efficiencies under certain assumptions, the same can be said about particular offset arrangements. It is critical to note, however, that offsets, in the technical definitional sense, differ from bundled goods and services. To see why, recognize that the purchase of a bundle will only occur when the individual components are not sold separately or quantity discounts exist for the bundle, vis-à-vis buying the components separately. It is not unusual that neither of the above hold in offset contracting.

This approach is helpful but of limited use because the importance of capabilities, transaction cost, and public choice issues dwarf production economies. We need only observe recent government procurements to confirm that countries are often willing to accept production diseconomies to secure technology transfer and augment domestic capabilities.

For example, in the 1990s, Japan opted for a coproduction agreement instead of purchasing 200 F-15 aircraft off the shelf from a U.S. supplier. The relatively short production run in Japan resulted in severe diseconomies of scale. The total cost of coproduction to the Japanese was 2.5 times greater than the price for the aircraft in arms-length exchange with the American firm.^[4] Certainly, institutions and political economy variables played a crucial role in this procurement decision. International procurement is replete with similar contractual arrangements.

^fSeminal contributions to this literature include Penrose,^[13] Nelson and Winter,^[14] Chandler,^[15] and Teece and Pisano.^[16] The capabilities literature burgeoned after reaching a critical mass in the early 1980s. The work of Penrose on the theory of the firm is noteworthy.^[13] She was the first to argue that the boundary of the firm is a function of skills and know-how, not simply physical stocks of resources and technology. The capabilities approach acknowledges the contributions of transaction cost economics, but maintains that the boundary of the firm depends heavily on the number and type of skills and routines cultivated.

^gSee, for example, Prahalad and Hamel^[17] and Teece and Pisano.^[16]

^hA final good or service sold in markets. The production process is already complete.

CONCLUSION

Like other government policies, there are benefits and costs associated with offsets. Short-term gains for the targeted industries benefiting from an offset obligation are often visible and politically attractive. Under certain high-technology procurement settings, the offset may actually improve the efficiency of the exchange. In addition, a well-designed offset policy has been shown to contribute to the economic vitality of developing industries.

In a general equilibrium framework, however, the governmental action may result in a misallocation of resources and deadweight loss. Furthermore, because many offset obligations are not usually fulfilled for 5–10 yr, the long-run welfare effects on the purchasing economy will likely exceed the initial contract value.ⁱ Corruption, faulty contract design, and non-compliance in offset fulfillment must be vigorously monitored and addressed to reap the fruits of a well-conceived offset. A rich empirical investigation of this topic is not only warranted—it is needed in this time when governments are actively using these contracts as part of a dynamic economic development strategy. Many governments are currently experimenting with various kinds of offsets and other non-standard contracts in an effort to find efficient procurement instruments that can exploit government bargaining power. This heuristic process is slow and expensive; procurement officials would be wise to perform extensive analysis and auditing of offsets as recommended in Taylor and Brauer and Dunne.^[11,19]

REFERENCES

1. (accessed November 14, 2006); <http://www.lockheedmartin.com/wms/findPage.do?dsp=fec&ci=11701&rsbci=0&fti=0&ti=0&sc=40>.
2. Matthews, R. Saudi Arabia's defense offsets programmes: policy and performance. *Defence Peace Econ.* **1996**, 7, 233–251.
3. Hammond, G. *Countertrade, Offsets and Barter in International Political Economy*; St. Martin's Press: New York, 1990.
4. United States Commerce Department. *Offsets in Defense Trade: Annual Report to Congress*. U.S. Government Printing Office: Washington, DC, 2006.
5. Liesch, P. *Government Mandated Countertrade*; Avebury: Aldershot, England, 1991; 4 pp.
6. Wood, D. *Australian Defence Offsets Program*; Proceedings of the Defence Offsets Seminar; Department of Defence: Canberra, Australia, 1992.
7. Countertrade & Offset. 1997. "Thomson Fulfills Polish Offsets with Local Radio Maker." 15 (January): 1–8.
8. Hooper, N. *The Economics of UK Defence Procurement*; Center for Defence Economics York: U.K., 1990; Vol. 1, series 8: Discussion Series Paper.
9. Williamson, S.; Wright, R. Barter and monetary exchange under private information. *Am. Econ. Rev.* **1994**, 84 (1), 104–123.
10. Hayashi, F.; Matsui, A. A model of fiat and money barter. *J. Econ. Theory* **1996**, 68, 111–132.
11. Taylor, T.K. Using offsets as a government procurement strategy. *J. Public Procurement* **2002**, 1.
12. Hall, P.; Markowski, S. On the normality and abnormality of offset obligations. *Defence. Peace Econ.* **1994**, 5, 173–188.
13. Penrose, E. *The Theory of the Growth of the Firm*; Basil Blackwell: Oxford, U.K., 1959.
14. Nelson, R.; Winter, S. *An Evolutionary Theory of Economic Change*; Harvard University Press: Cambridge, MA, 1982.
15. Chandler, A. Organizational capabilities and the economic history of the industrial enterprise. *J. Econ. Perspect.* **1992**, 6 (3), 79–100.
16. Teece, D.; Pisano, G. The dynamic capabilities of firms: an introduction. *Ind. Corp. Change* **1994**, 3 (3), 537–556.
17. Prahalad, C.K.; Hamel, G. The core competence of the corporation. *Harvard Bus. Rev.* May–June, **1990**, 79–91.
18. Langlois, R.; Robertson, P. *Firms, Markets and Economic Change: A Dynamic Theory of Business Institutions*; Routledge: New York, 1995.
19. Brauer, J.; Dunne, J. *Arms Trade and Economic Development: Theory, Policy and Cases in Arms Trade Offsets*; Routledge: London, 2004.

ⁱPath dependency from the use of offsets is a potential by-product for some industries building complementary assets that are specific to a particular system.

Ombuds and Ombuds Programs

Tina Nabatchi

Lisa B. Bingham

*School of Public and Environmental Affairs, Indiana University,
Bloomington, Indiana, U.S.A.*

INTRODUCTION

This chapter discusses the role of an ombuds, a person within an organization who assists employees, stakeholders, or customers in resolving disputes by offering a wide variety of alternative dispute resolution (ADR) processes (for more information on ADR processes, refer to Nabatchi and Bingham in this volume; for more information on ADR in the public sector, refer to [Nesbit and Bingham](#) in this volume).

OMBUDS

Contemporary conflict resolution theory and practice suggest that organizations should design and build integrated conflict management systems, as opposed to a singular dispute resolution program, to better manage organizational conflict.^[1] In turn, many scholars and practitioners have asserted that an ombuds office is both a desirable and cost-effective element in an efficient dispute resolution system.^[2,3]

In Scandinavian countries, where the concept originated, an ombudsman, or an ombuds, is a public official appointed to hear citizen complaints and to conduct independent fact-finding investigations to correct abuses of public administration. The original word, *ombudsman*, is not gender-specific in Scandinavia; however, to avoid gender connotations, authors and scholars in the United States variously refer to an ombudsman as ombuds, ombudsperson, ombuds practitioner, etc. In the United States, the concept and function of an ombuds have developed differently, and have grown in both significance and prominence as they have blossomed in a number of public and private settings.

In the United States, an ombuds operates like a third-party neutral inside an organization by assisting employees, stakeholders, and/or customers of the organization in resolving workplace disputes through confidential, informal means. An ombuds has a diverse set of responsibilities and must serve many functions, such as supplying information and other resources, channeling communications, handling complaints, and resolving disputes.^[4,5] For these many functions, the

ombuds also plays a wide variety of roles, such as counselor, consultant, informal go-between, facilitator, formal mediator, and/or informal fact finder.^[6] Moreover, the ombuds uses various conflict resolution skills, such as listening, conflict coaching, facilitation, mediation, and shuttle diplomacy,^[6] but also has the ability to help individuals gain access to other informal or formal dispute resolution processes within the organization, such as traditional grievance procedures and equal employment opportunity (EEO) processes (for more information on grievance procedures, refer to Nabatchi and Bingham in this volume).

The ombuds serves an integrated function for the organization, and has the flexibility and capacity to handle numerous and varied disputes. In addition to providing assistance to people with complaints, concerns, and conflicts, the ombuds often also acts as an upward feedback mechanism, problem prevention device, and change agent in the organization.^[6] In this respect, the ombuds serves as a troubleshooter for the organization. An ombuds generally works closely with all personnel in an organization, from top-level management to front-line staff; therefore, an ombuds is often able to identify processes and practices that create employee (and other) conflicts, or that otherwise have negative effects on the organization. In this respect, the ombuds serves as a bellwether of organizational problems that could develop into lawsuits and as an early warning signal for new and emerging problems. Moreover, the ombuds may notice a pattern of problems, which suggests the need for policy or procedural change within the organization. The ombuds will try to eliminate these sources of organizational conflict through problem solving or brainstorming meetings, retreats, trainings, and other conflict management sessions. These features of an ombuds not only result in a more integrative and systemic approach to managing conflict in an organization, but also allow the ombuds to serve both a corrective and preventative role as a resolver of conflicts.

The effectiveness of an ombuds office depends on its ability to demonstrate confidentiality and neutrality—key principles in the ombuds profession. In terms of confidentiality, ombuds generally do not keep case

records, and the majority of their work is unknown to anyone except those who use their services.^[7] However, in the federal sector, there are certain exceptions to confidentiality for ombuds; in some circumstances, federal employees must report information when it relates to a felony, waste, fraud, abuse, corruption, or crimes by government officers or employees.^[8] These reporting requirements interact with federal statutes providing for confidentiality in dispute resolution.^[8] Similarly, state and local governments vary in the degree of confidentiality they afford dispute resolution processes involving administrative agencies.

In terms of neutrality, the ombuds must: 1) resist taking sides during a conflict; 2) possess no stake in the outcome of a dispute; and 3) demonstrate fairness, objectivity, and evenhandedness in relationships.^[4] To promote these qualities, the ombuds is outside of the organizational chain of command or reporting structure, and although all ombuds are made available and paid for by the host institution, not all ombuds are on the host organization's payroll. Some may work through a contract with another organization to provide an additional layer of neutrality.

OMBUDS PROGRAMS

Ombuds programs have been developed in both the public and private sectors. One reason for the growth of ombuds programs in the United States is that they are able to personalize the dispute resolution processes of large organizations such as corporations, municipalities, and government agencies. Thus, they provide people with individualized and flexible attention such that they are able to overcome the difficulties associated with the formalism and bureaucracy of large organizations. Today, ombuds programs are found in large corporations such as Haliburton (formerly Brown and Root), educational institutions such as the Massachusetts Institute of Technology (MIT), and government agencies such as the National Institutes of Health (NIH) and the Internal Revenue Service (IRS).

Ombuds programs can focus on different organizational constituencies. Traditionally, they provide services to those both within and outside the organization, including members of the public. In the federal government, most ombuds offices focus on organizational issues within the agency.^[8] In some cases, Congress established ombuds offices through legislation (e.g., Federal Deposit Insurance Corporation, Office of the Comptroller of the Currency, Small Business Administration, and State Department).^[8] In other cases, the agencies themselves created the office and staffed it with career employees insulated from politics (e.g., Customs Service, Federal Bureau of Investigation, Food and Drug Administration, and Department of Justice).^[8]

Research on federal workplace ombuds offices suggests that they can be effective when the EEO office has too many non-EEO complaints, or the employee assistance program (EAP) is receiving workplace complaints outside of its mandate. In addition, ombuds programs can be effective when personnel-related offices are not working together, employee morale is low, and employee-management communication is poor. Finally, such programs can be effective when there are poor labor-management relations and when there are frequent employee claims of retaliation.^[9]

BENEFITS AND RISKS OF OMBUDS PROGRAMS

Ombuds programs appeal to organizations for many reasons. First, many dispute resolution programs are accessible to a limited set of employees (e.g., those covered under collective bargaining agreements). In contrast, ombuds programs are accessible to everyone in the organization. Likewise, many dispute resolution programs are rigidly designed to handle specific types of disputes (e.g., those involving workplace contract or policy issues); however, ombuds programs are flexible enough to address typical workplace disputes related to contracts, policy, or law, but can also reach beyond these conflicts to handle individual and group disputes involving interpersonal issues, environmental problems, or racial and ethnic tensions. In short, ombuds programs are more inclusive of organizational personnel, can address a broader range of issues, and have greater ability to resolve conflict constructively and provide stability to the organization than single-process dispute resolution programs.

Second, most dispute resolution programs are reactive in that they only address conflicts that have been made "public." In other words, the programs respond only to disputes that have been formally expressed. Conversely, ombuds programs are both reactive and preventative: they can address formally expressed disputes, but can also address disputes where one party is unwilling or unable to publicly confront the other party.

Third, traditional dispute resolution programs focus on the symptoms rather than the source of problems. They engage in problem solving between the disputing parties, but not problem solving for larger organizational issues. Ombuds programs help identify and address the root causes of problems through systemic change. In theory, this allows an organization to decrease the highly visible costs of conflict and to increase organizational morale, productivity, and communication.

Finally, and perhaps most importantly, ombuds programs assist employees in finding fair solutions to

problems, and may facilitate problem solving at lower levels than would be necessary through traditional dispute resolution processes. This not only saves valuable resources for the organization, but also promotes a sense of justice, fairness, and participation among organizational constituents. Research shows that higher perceptions of justice in an organization lead to higher employee productivity and morale. Moreover, evidence suggests that people who learn how to effectively resolve disputes are more likely to resolve future disputes independently and more likely to teach these methods to other colleagues.^[10]

Despite the numerous potential benefits of ombuds programs, there are also some risks, especially if the ombudsperson violates guarantees of neutrality and confidentiality. One potential risk is that employers may distort the ombuds title in unilaterally adopted non-union arbitration programs. For example, one employer had its ombuds represent employees as their advocate in arbitration and select the arbitrator on behalf of both parties. After repeated selection of the same arbitrator who always ruled for management, it was clear that this particular ombuds was not neutral. Similar structures, where the ombuds departs from a neutral role, give the appearance of a conflict of interest. This reduces stakeholder trust in the ombuds office, which is a critical element for program success.^[11]

CONCLUSION

The use of ombudspersons continues to grow in a variety of settings, including educational institutions, hospitals, nursing homes, corporations, and government. Ombudspersons now have a professional association and a code of ethics.^[12] (For additional information and resources about ombuds and ombuds programs, refer to The Ombudsman Association (TOA) at www.ombuds-toa.org). As the benefits of ombuds programs are increasingly measured and understood, and

as conflict management theory and practice increasingly move toward dispute system design efforts and integrated conflict management systems, the prominence of such programs is also likely to increase.

REFERENCES

1. Lipsky, D.B.; Seeber, R.L.; Fincher, R.D. *Emerging Systems for Managing Workplace Conflict: Lessons from American Corporations for Managers and Dispute Resolution Professionals*; Jossey-Bass: San Francisco, 2003.
2. Rowe, M.P.; Perneski, T. Cost-effectiveness of ombudsman offices. In *Corporate Ombudsman Newsletter*; May, 1990.
3. Slaikeu, K.A.; Hasson, R.H. *Controlling the Costs of Conflict: How to Design a System for Your Organization*; Jossey-Bass: San Francisco, 1998.
4. Gadlin, H.; Pino, E.W. Neutrality: a guide for the organizational ombudsperson. *Negot. J.* **1997**, *13*, 17–37.
5. Robbins, L.P.; Deane, W. The corporate ombuds: a new approach to conflict management. *Negot. J.* **1986**, *2*, 195–205.
6. Rowe, M.P. The ombudsman's role in a dispute resolution system. *Negot. J.* **1991**, *7*, 353–362.
7. Isenhardt, M.W.; Spangle, M.L. Ombudsperson. In *Collaborative Approaches to Resolving Conflict*; Sage Publications, Inc.: Thousand Oaks, 2000; 169–175.
8. Senger, J.M. *Federal Dispute Resolution: Using ADR with the United States Government*; Jossey-Bass: San Francisco, 2003.
9. Meltzer, D.L. The federal workplace ombuds. *Ohio State J. Dispute Resolut.* **1998**, *13*, 549–609.
10. Rowe, M.P. Helping people help themselves: an ADR option for interpersonal conflict. *Negot. J.* **1990**, *6*, 239–248.
11. Bingham, L.B. Emerging due process concerns in employment arbitration: a look at actual cases. *Labor Law J.* **1996**, *47*, 108–126.
12. The Ombudsman association. *Code of Ethics*; 1997. (available at www.naples.cc.sunysb.edu/Pres/ombuds.nsf/pages/ethics).

Open Systems Theory

Rupert F. Chisholm

*Department of Management, Pennsylvania State University at Harrisburg,
Middletown, Pennsylvania, U.S.A.*

INTRODUCTION

Roots of open system theory date at least from the 1920s. At that time, scientists from several disciplines, including biology, philosophy, physics, and sociology, began to question prevailing mechanistic theories in their fields. For example, Ludwig von Bertalanffy, the “father” of general systems theory, observed that existing biological theories failed to explain the essentials of the phenomena of life. Instead, he advocated considering organisms as wholes, complex systems comprised of organization at various levels.^[1]

Thinking continued to develop in various disciplines over the next 25 years or so. In 1954, as a result of being located together at the Center for Advanced Studies at Stanford University, von Bertalanffy and several colleagues from diverse fields established the Society for General Systems Research. This organization aimed to identify commonalities of concepts, principles, and models in various fields to help transfer knowledge, increase understanding, and foster theory development among disciplines. Organizing the Society signaled a basic change in thinking about organizations and how best to understand and manage them.

THE NATURE OF OPEN SYSTEMS

Basically, a system is “a set of interacting units with relationships among them.”^[2] Organizations are a type of concrete (versus abstract) living system. Because all living systems are open (versus closed systems that have no transactions with an external environment), organizations are, by definition, open systems.

Fig. 1 gives a graphic representation of an organization as an open system. According to open systems theory (OST), all organizations (public, private, and non-profit) depend on the external environment to provide certain critical inputs to exist and carry out work over time. As the figure indicates, these required inputs include money, personnel, various types of materials and supplies, and information. The exact makeup of these critical inputs depends upon the industry, primary technology, strategy of the organization, and other contextual factors.

Transforming the critical inputs is the second phase of the overall process. Organizations transform required inputs into services or products that are exported back to the external environment. During this phase, an organization must “add value” to inputs brought into the system. Transforming inputs involves using various combinations of technologies. Technology includes all the knowledge, information, material resources, techniques, and procedures required to convert system inputs into outputs. For example, a road paving crew uses machinery and equipment, general knowledge, techniques, and information about the specific site to surface a segment of state highway. And, mental health staff members use various therapeutic techniques, knowledge of their clients’ individual, family and social situations, physical environments, and possibly medication to improve psychological functioning. As an open system, an organization must carry out transformation processes that produce services or products that meet the needs of other individuals, groups, or organizations in the outside environment. Only by consistently delivering services required by external constituents can an organization remain viable over time.

Exporting outputs to the environment comprises the third phase of the total process. Organizations produce several types of outputs: services and products, information, waste, and others. Some outputs are planned for and intended (e.g., the basic service delivered), and others are unintentional (e.g., the employees’ attitudes developed while delivering the service). As a result, an organization must attempt to clearly identify all key outputs and how they are perceived by stakeholders. To do this requires effective feedback from critical stakeholders (e.g., clients, employees, suppliers, sponsors) and other potentially important individuals, groups, and organizations. Receiving effective feedback is necessary to assure that the organization is producing and delivering valued services of the right kinds, at the right times, and in ways that are acceptable to recipients. And, as indicated below, receiving feedback does not occur automatically. Instead, it requires designing effective processes for constantly obtaining valid information about reactions to current services and service delivery and about basic long-term changes in the environment that may alter future service demands.

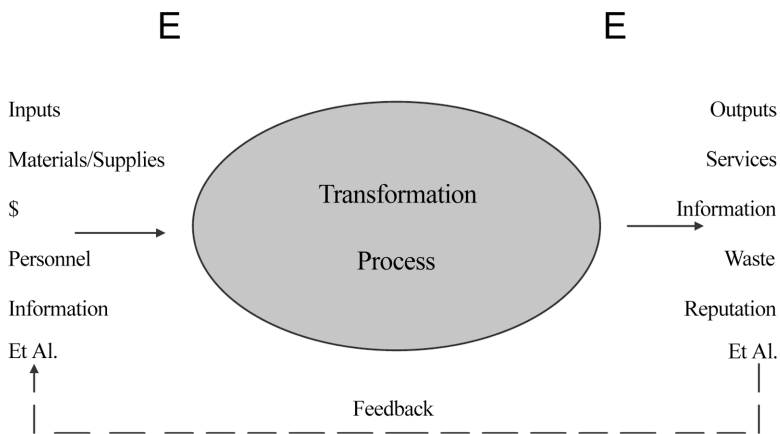


Fig. 1 Open systems view of organizations.

To summarize, organizations are complex systems that continuously interact with the external environment by importing critical inputs, transforming inputs into valued services and products, and exporting these services and products to the external environment. Because of the continuous input-transformation-output process, organizations depend upon the outside environment for their existence. And, to remain viable over time, they must deliver services (products) that meet the requirements of clients and other key decision makers outside the organization. Maintaining viability also requires designing and managing effective feedback processes.

KEY FEATURES OF OPEN SYSTEMS

Open systems theory also calls attention to several additional features that are important for understanding and managing organizations. These features include negative feedback, boundary, negative entropy, dynamic equilibrium, system levels, equifinality, and holism.

Negative Feedback

Feedback is crucial to the existence and functioning of organizations. According to OST, feedback is the flow of information back from the output phase to the input phase of the input-transformation-output process. The broken line in Fig. 1 represents this flow. Feedback can be positive or negative. Negative feedback, which indicates that the system is off course or is not meeting its goals, is most important for understanding and managing organizations. The house thermostat is an example of a simple negative feedback system. Here, the desired temperature is set, and when the actual temperature deviates above or below it beyond a prescribed limit,

this information automatically triggers an “on” or “off” switch. For many aspects of organization functioning, feedback does not occur automatically, and the process is considerably more complex than that of a thermostat. First, crucial processes that require feedback systems must be identified. Then these systems must be consciously designed, monitored, and managed. In addition, organizations must remain alert to make changes in the goals or standards against which to compare actual conditions as changes in the external environment and within the system occur.

Boundary

Following OST, a boundary is the imaginary line that separates an organization from its external environment. In Fig. 1, the oval surrounding the transformation process depicts the organizational boundary. Space outside the system boundary represents the external environment, E. A boundary carries out two essential functions for organizations: 1) it filters inputs and outputs; and 2) it binds internal system elements together and maintains system integrity. No system can exist and function if inputs of any kind or in unrestricted amounts are allowed to enter: it will soon disintegrate due to input overload. Therefore, each organization must make decisions and establish criteria that determine which inputs are permitted to enter, in what forms, in what amounts, when, and how. The boundary also operates during the output phase. In effect, the filtering qualities of a boundary result from the many decisions made, guidelines established, and constraints that emerge from formal and informal system functioning over time. By differentiating the organization from its environment, a boundary also helps hold elements of the system together. Members of the organization come to think of themselves as “insiders” and develop a shared view of themselves,

the organization, and the outside world. Over time, a unique organizational culture emerges that helps form and reinforce the boundary and provide coherence to the system. Although there are certain physical features (e.g., buildings, perimeter fences) that partially reflect boundaries, other aspects are primarily psychosociological (e.g., “I’m a Welfare Department employee”; “I’m a doctor at the XYZ Medical Center”). Establishing and maintaining appropriate dimensions and degrees of permeability of boundaries are crucial to maintain the system and achieve organizational effectiveness.

Technically, the environment [(E) in Fig. 1] includes everything outside the system boundary. However, complexity has led to conceptualizing the total environment as task and general environment.^[3] The task environment includes all the organizations, groups, agencies, and individuals that affect short-term organizational functioning. These include suppliers, information sources, regulators, customers, lenders, citizen groups, and others—any entity that can affect goal setting and goal attainment activities of the system. The general environment is the environment that exists beyond the task environment: legal, social, political, economic, technological systems that surround and provide the context for the organization and its task environment. Future opportunities and new constraints on organization actions emanate from the general environment. Consequently, organizations must be proactive to identify and determine the implications of emerging forces in the general environment and reactive to feedback from the existing task environment. “Opportunistic surveillance” is required to discover and create opportunities for an organization and to maintain long-term system health.^[3]

Negative Entropy

Negative entropy (negentropy) results from the open system nature of organizations. The entropic process is a universal law of nature in which all forms of organization move toward disorganization and death.^[4] Following the second law of thermodynamics, over time, physical systems tend toward states of maximum disorder and toward a leveling of differences. In contrast, open systems tend to become more elaborate through the differentiation of internal structures and processes. Greater complexity can emerge because of the basic design and dynamics of the system. It is made possible via continuous transactions with the environment. Conceptually, the process of differentiation can continue indefinitely provided the system continues to generate sufficient negative entropy (negentropy). Organizations as open systems can generate negentropy in many ways: by attaining

favorable ratios of inputs to outputs, by improving the internal alignment of internal components and processes, and by improving location in and linkages with the external environment. Acquiring information about basic changes in the general environment, how these provide future opportunities, and realigning the organization with emerging external conditions can provide substantial negentropy. Theoretically, an infinite number of possibilities exists, and each organization must strive continuously to discover and invent ways of sustaining the negentropic processes.

Dynamic Equilibrium

Organizations as open systems operate in states of dynamic equilibrium. Such states mean that, despite constant movement which results from continuously importing, transforming, and exporting inputs and outputs and internal system dynamics, an organization tends to remain in a steady (but not static) state. In a steady state, forces that act to influence a system in one direction tend to be offset by countervailing forces in the opposite direction. Using Lewin’s^[5] force field diagram is one way to conceptualize this phenomenon (see Fig. 2). Here, factors affecting phenomena in organizations (e.g., quantity of output; quality of work) are represented as driving and restraining forces. At a particular time, driving forces to change the level of quality or quantity of output tend to be offset by restraining forces in the opposite direction. These offsetting forces are not necessarily precisely equal, but they tend to keep the organization functioning within certain upper and lower limits and maintain system integrity. This feature is called “dynamic equilibrium.”

Dynamic equilibrium results from two types of basic systems processes. First is primary regulation, which results from interactions that naturally occur among components of the system or between the organization and the external environment. For example, behavioral norms emerge spontaneously from interactions that occur among organization members as they carry out work activities. Similarly, informal roles also emerge over time from spontaneous interactions inside the organization and with individuals in the external environment. These naturally occurring phenomena are aspects of primary regulation processes of all social systems that help structure expectations, attitudes, and behavior and make employee collaboration and cooperation possible.

Secondary regulation results from consciously designing feedback systems to regulate key system processes. Budgets and management information systems are examples of such regulation. Katz and Kahn^[4] identify the necessity of having appropriate feedback systems for organizations. However, because secondary

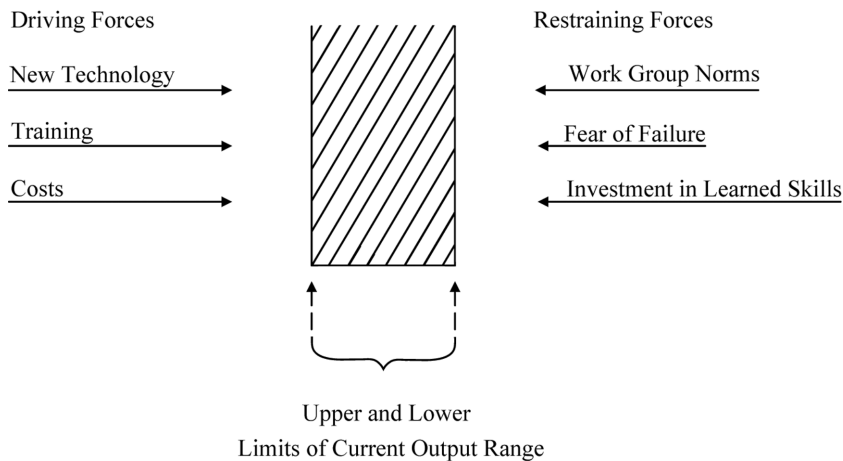


Fig. 2 Force field representation of factors determining current level of work group output.

regulation tends to limit the potentiality (number of possible future states of a system) of the organization,^[1] proliferating secondary regulation processes is not necessarily desirable. Such proliferation may lead to inflexibility, inability to respond quickly, and lack of capacity to perceive new opportunities in the environment. Bureaupathology^[6] can result from overemphasis on such formal control mechanisms. In short, organizations must carefully consider using both primary and secondary regulatory processes. For example, team-based organizations rely heavily on primary regulation to improve flexibility and response time. Organizations also must monitor and redesign these systems to meet new requirements.

System Levels

Another important OST principle is that systems exist at several levels. Miller,^[2] for example, identifies the levels of cells, organs, organisms, groups, organizations, societies, and supranational systems for living systems. For organizations, primary levels include the individual, group, organization, society, and trans-societal. Two things are important about system levels. First, each system level has certain emergent features not found at lower levels of complexity that affect system functioning in highly relevant ways. To illustrate, group norms, roles, and values affect group and member functioning to a considerable degree. And, these factors are not present at the individual level that is immediately below the group level. Second, systems exist within systems. The group is a subsystem of the organization (the suprasystem), that, in turn, is a subsystem of the society, its suprasystem. Hence, within an organization, group functioning is strongly affected by features of the organization. And, as an open system, societal norms and values influence actions of the total organization, as well as those of groups and individuals within it. Linkages among system levels also

lead to the influence of subsystems on higher levels (e.g., groups on organizations or organizations on society). In general, the effects of higher-level systems on those of lower levels is greater than the reverse. Because systems exist within systems of greater complexity and a given system depends on the higher-level systems for critical inputs (and for accepting outputs), understanding an organization best starts at the next higher level or from the outside in. Following this OST approach, organizational analysis best starts with the environment, where the organization is located in it, its basic orientation to it (vision, mission, strategy), and how it carries out basic transactions with the environment.

Equifinality

Traditional organization and management concepts were based on a mechanistic model of human beings and of organizations. Taylor's^[7] scientific management, in particular, attempted to determine the "one best way" of conducting work operations. The theory of bureaucracy and classic management principles also rested on closed system, deterministic logic—"doing 'x' will result in 'y.'" Much current organization and management theory remains rooted in such thinking. In contrast, OST states that reaching a desired state or a particular set of goals can be achieved in many different ways (equifinality). Thus, organizations in the same industry can succeed following quite different strategies and using diverse designs. Relatedly, a given set of starting conditions does not necessarily lead to an expected set of outcomes (multifinality). For example, having the best educated, most highly qualified employees does not guarantee having the most effective organization. In short, as open systems, organizations have considerable flexibility to determine how to reach a particular set of goals (end state). And, establishing certain starting conditions does not guarantee reaching a predefined set of goals. These

two principles mean that managers have considerable latitude to create and design structures and processes that meet the particular requirements of an organization.

Holism

Holism involves the essential quality of organizations as open systems. Ideally, this OST principle means that “the whole is more than the sum of its parts.” Put more precisely, the statement should read, “. . . *can be* more than . . .”. For, with open systems, this is a possibility, not a guarantee. The synergy that occurs under certain conditions to make decisions of effective groups or organizations superior to those of individual members illustrates holism as a positive phenomenon. At the same time, as indicated in the previous section, because open systems are far from deterministic mechanical systems, group decisions also can be of lesser quality than those reached by individual group members (e.g., Bay of Pigs invasion). Thinking about organizations should include the potential of the holistic principle and also an awareness that this phenomenon does not automatically take place.

IMPLICATIONS OF OST

Contrasted with earlier mechanistic closed system views of organizations, OST provides a revolutionary perspective for thinking about and managing organizations. The following statements highlight several important implications of this perspective.

1. Open systems thinking emphasizes the criticality of the external environment in providing required inputs, in determining the acceptability of outputs, and in affecting appropriate design of internal structures and processes. This basic dependency on the environment suggests that understanding an organization best starts from the outside and works inward. Attempting basic organization change should proceed in a similar way.
2. Understanding environmental demands and constraints on an organization is essential to understanding organizational functioning. However, the environment does not dictate organization design. Rather, equifinality indicates that there is more than one route to organizational effectiveness. And, because each organization is unique in certain important

ways (e.g., development history, culture), an organization must attempt to develop its own design and ways of operating to reach optimal effectiveness.

3. OST emphasizes the dynamism inherent in organizations. This dynamic quality stems from never-ending interaction with the environment, dynamic equilibrium that maintains the organization and key processes within certain limits, the tendency to generate negative entropy, and negative feedback. Focusing on the dynamic nature of organizations as open systems helps reveal the severe limitations of traditional static views of organizations as closed mechanical systems. OST challenges theorists to continue to develop improved models that increasingly capture more of the dynamic, ever-changing nature of organization.
4. Open systems thinking also focuses attention on maintaining the input-generating capacity of the external environment. Because organizations rely on the environment to provide critical inputs, concern about the ability of the system to generate inputs indefinitely into the future must exist. Otherwise, individual organizations create conditions that make future sustainability impossible. In short, from a long-term perspective, organizations cannot determine how well they are doing for themselves without determining how well they are doing for the suprasystem (environment).^[8]

REFERENCES

1. von Bertalanffy, L. *General System Theory*; George Braziller: New York, 1968.
2. Miller, J.G. *Living Systems*; McGraw-Hill: New York, 1978.
3. Thompson, J.D. *Organizations in Action*; McGraw-Hill: New York, 1967.
4. Katz, D.; Kahn, R.L. *The Social Psychology of Organizations*, 2nd Ed.; Wiley: New York, 1978.
5. Lewin, K. Frontiers in group dynamics. *Hum. Relat.* **1947**, *1*, 5–41.
6. Merton, R.K. *Social Theory and Social Structure*; Free Press: New York, 1957.
7. Taylor, F.W. *Scientific Management*; Harper and Row: New York, 1911.
8. Berrien, F.K. A general systems approach to organization. In *Handbook of Industrial and Organizational Psychology*; Dunnette, M.D., Ed.; Rand McNally: Chicago, 1976.

Operations Research

William L. Waugh Jr.

*Department of Public Administration and Urban Studies, Georgia State University,
Atlanta, Georgia, U.S.A.*

INTRODUCTION

Operations research or management science involves the development of mathematical models to simulate real-life problems. The goal is to design, analyze, and improve the performance of a system or process. By examining the structure of a problem, analysts can search out optimal solutions. Models may be created to examine the operation of a wide range of systems, including financial, engineering, industrial, and administrative systems.^[1]

The roots of operations research can be traced to the pre-World War II period. British research on military deployments used radar to track aircraft, and scientists termed part of the project “operational research.”^[1] During the war, operational research techniques were applied to submarine warfare, strategic bombing, and other military problems. War-related research involving chemists, physicists, statisticians, mathematicians, and planners began using the same quantitative tools.^[2] Following the war, the techniques were applied to financial, industrial, and governmental problems. As the field developed in the 1950s, professional societies were formed, first in Great Britain and the United States and then elsewhere. Since the postwar period, the applications have increased and become more specialized. Now, operations research is the term commonly used in industry and engineering, and management science is the term commonly used in dealing with administrative systems.^[1]

THE PRACTICE OF OPERATIONS RESEARCH

Operations research begins with the construction of mathematical models to represent real-life phenomena, such as systems and processes. The modeling begins with the identification of a problem or issue to be studied. The controllable and uncontrollable aspects of the system or process, the goals or purposes, and the constraints or limits are identified. Data are collected to approximate as best as feasible the attributes of the model. When the model is operational, analysts must then determine the extent to which the model

represents reality and how applicable their findings will be to the real-life problem.

Complex models are possible, but care has to be taken to make sure that the complexity does not interfere with the development of practical solutions. Simplistic models may be sufficient and, in many cases, are preferred, because they are easier to explain to decision makers. The development of algorithms, the “sequence of operations that can be carried out in a finite amount of time,”^[1] to model problems may be easy or may even be impossible. Not all problems are solvable. The size of the problem, the number of steps in the process, and the availability of data may confound analysis. Optimality may have to give way to practicality. The nature and size of the error may make optimization difficult if not impossible. Suboptimal solutions may be the best that can be expected.

The practice and methodologies of operations research have grown with the development of computer software and hardware. New PC-based software provides tools for dealing with specific classes of problems, and the expanded capacities of hardware are making it possible to include many more variables and much more data in the models. It is also making it possible for individual managers to use operations research/management science techniques for the growing number of decision-making situations that they face in the workplace. Specialized software is increasingly available through Internet “libraries” sponsored by professional organizations as well as private vendors.

Operations research has been used to address a wide variety of problems in the public and non-profit sectors. It has been widely used by the Department of Defense in much the same ways it was used during World War II. It has been used for bus scheduling, postal delivery and mail sorting, waste management, land use regulation, highway safety, education, and agriculture.^[3] Some common operations research/management science techniques follow.

The Linear Programming Model

The linear programming model is one of the most common operations research/management science

technique. This is a class of models in which the objective function, a principal criterion representing effectiveness (e.g., profit maximization or cost-efficiency) and the constraints on the system or process (e.g., time, human resources, and financial resources) can be expressed in linear terms. There may be more than one or even no optimal solution. There may be no feasible solution, as well.

Network Analysis

Network analysis deals with the linkages among elements in a system. Networks are common in a variety of systems, including highway systems, aviation flight patterns, utility grids, and communications systems. Path analysis, the analysis of paths between and among centers or nodes, can provide solutions to flow problems, such as traffic routing and cost of problems. The flow of work can be analyzed and decisions made concerning how to speed up or improve the quality of the work. PERT-CPM (Program, Evaluation, and Review Technique and Critical Path Method) is a common tool to analyze projects and to determine how they may be made faster and more cost effective.^[4]

Integer Programming

Integer programming is useful in dealing with systems in which one or more values cannot be a fractional. For example, industrial processes typically do not manufacture incomplete products, such as half of an automobile or a fourth of a laser disk, and administrative processes typically do not serve fractions of individual clients or employees.

Non-Linear Optimization

Non-linear optimization provides tools for dealing with problems that are non-linear in nature. The tools can accommodate processes that are non-sequential or in which variables change at different rates. For example, doubling the speed of computer processors does not necessarily double the speed of computations, because other variables affect the speed of computations.

Markov Process

Markov processes are useful in dealing with processes that are probabilistic, where decisions have to be made based upon data that may be uncertain or, in some cases, missing. Prior event probabilities are used to predict the probabilities in the subject system.

Queuing Theory

Queuing models are commonly used in dealing with systems in which people are queued or in line, situations in which customers or clients may be waiting for service and decisions need to be made concerning how many servers are necessary to satisfy the demand. The number of customers and system capacity have to be weighed against the acceptable length of waits and costs.

Simulation

Simulation or observing the behavior of a model to see how it operates so that improvements can be made is also a relatively common operations research technique. Simulations permit the examination of interactions among variables in a system when researchers vary the conditions. One or a few variables may be changed to analyze the impact on the model.

There are also a wide variety of decision analysis techniques, including game theory, decision trees, and utility theory, to explain decision making. Knowledge about the psychology of decision making, including non-rational approaches, is also applied. Also, there are heuristic techniques that can be applied to help identify optimal decisions.

CONCLUSION

Operations research/management science techniques have been invaluable tools in many decision-making and problem-solving contexts. The British and American efforts in World War II are illustrative of that utility. The application of quantitative techniques has greatly aided public, non-profit, and private decision makers in finding better solutions to engineering, industrial, financial, and other administrative problems. Financial resources can be allocated more efficiently; school buses can be routed more effectively to reduce wait times for students and costs to school systems; construction projects can be managed better to reduce time and other resources; management systems can be streamlined; communications systems can be designed to link organizations and individuals more effectively; garbage pickup can be managed to reduce costs; and so on.

The use of such techniques has not been without critics, however. While few argue that the techniques have not been of benefit in many instances, many argue that sometimes the techniques have been misapplied. Rational decision processes, of which operations research and management science are examples, may best be applied at the level of operational decisions.^[4,5] Policy-level problems are usually less clearly defined,

and the goals are less likely to be solely economic. Not all decisions are simply technical, without social and political implications. In short, efficiency and speed are not always the preferred values, and optimal solutions based on those values may be less important than consensually derived solutions. Policy decisions, in particular, may be characterized by intense disagreement over means and goals. For the products of operations research, i.e., the “optimal” solutions, to be accepted and adopted, there need be some sensitivity to the political context of the solution and the need for compromise, little strong opposition to and at least some support for the solution, some attention to jurisdictional prerogatives, sensitivity to the goals of society and the real goals of the decision makers, congruence with societal, group, and individual goals, no offsetting side effects, and achievement of prior results cheaper and faster.^[6]

There are biases inherent in operations research/management science. Operations research has been described as “value free,” because the techniques have been used in a variety of contexts and often with a single-minded focus on efficiency, without attention to the normative issues raised. It is difficult to quantify some values and, in quantitative decision-making processes, non-quantified values tend to be undervalued. Indeed, it is argued that some values, some of the most important social and political values, cannot be quantified or to do so distorts their importance.

Time and data constraints, as well as the nature of the problems being addressed, dictate flexibility in

analysis. The application of operations research techniques is less important than the identification of practical solutions to problems at hand. Operations research techniques offer sets of tools that can be used along with other sets of tools. They are particularly effective at the operational level, where efficiency is the goal. They are much less effective at the policy level, where efficiency is less important than other goals, and optimal solutions are less important than politically acceptable solutions.

REFERENCES

1. Carter, M.W.; Price, C.C. *Operations Research: A Practical Introduction*; CRC Press: Boca Raton, FL, 2001; Sections 1, 2, and 4.
2. White, D.J. *Operational Research*; John Wiley & Sons: New York, 1985.
3. Henry, N. *Public Administration and Public Affairs*, 7th Ed.; Prentice-Hall: Englewood Cliffs, NJ, 1999; 166 pp.
4. Gortner, H.F.; Mahler, J.; Nicholson, J.B. *Organization Theory: A Public Perspective*, 2nd Ed.; Harcourt Brace: Fort Worth, TX, 2000; 373 pp.
5. Olshfski, D.; Collins, M. Management science. In *Defining Public Administration*; Shafritz, J.M., Ed.; Westview Press: Boulder, CO, 2000.
6. Nagel, S.S. Conflicting evaluations of policy studies. In *Public Administration: The State of the Discipline*; Lynn, N.B., Wildavsky, A., Eds.; Chatham House Publishers, 1990; 452 pp.

Orange County Bankruptcy

Susan MacDonald

California State University, Stanislaus, Turlock, California, U.S.A.

INTRODUCTION

The Orange County Board of Supervisors filed for [Chapter 9](#) protection from bankruptcy on December 6, 1994, declaring losses in the county's investment pool of approximately \$1.7 billion or roughly 20% of the entire pool. At the time, there were 187 government units, in addition to the county, participating in the Orange County Investment Pool (OCIP). Many of the participants were cities or towns located in the county; a few were towns outside the county's geographic limits that had asked to participate because of the county's record of high yields. All of the county's school districts were invested in OCIP because state law required them to be. Finally, regional water, transportation, and other special districts participated in the pool, although their involvement was voluntary.

Groups had been drawn to the investment pool, in part, because it was large and represented a wealthy county. Median family income was around \$48,500, fifth highest in California, and total property values were estimated between \$100 and \$150 billion. These and other wealth indicators gave investment pool managers access to a greater variety of financial instruments, lower fees by brokers and other financial advisors, and access to more expert and timely information than many other local government investment pools.

The county's treasurer, Robert Citron, had been in office since 1970, having secured voters' confidence in seven elections, the most recent being that in June. He had also been recognized for his financial accomplishments by the County's Board of Supervisors and by groups like *City & State* and the American Society of Public Accountants, which had pronounced him an "outstanding public official" during the treasurer's re-election campaign. The pool itself had repeatedly outperformed the state of California's investment pool. Between the recession years of 1991 and 1993, OCIP produced returns between 8% and 9%, while the state generated returns in the range of 5–6%. The differential meant additional revenues to Orange County of approximately \$250 million per year. Further, in October 1994, Merrill Lynch had underwritten a \$110 million bond issue to which both Standard and Poor's Corporation and Moody's Investors Services had accorded the county the second highest

rating, AA. Residents had re-elected County Treasurer Robert Citron that June—only six months before the bankruptcy. John Moorlach, Citron's opponent in the campaign, had made the Orange County Investment Pool, and its level of risk, the central issue in the campaign. Nevertheless, citizens re-elected Citron by more than 60%.

UNDERLYING CAUSES

Accounts of the bankruptcy clearly blame Citron for the debacle, but most trace the origins of the county's difficulties to the passage of Proposition 13 in 1978. Double-digit inflation in the 1970s had affected both citizens and the state. Those with assets, like the State of California, wanted to preserve their assets' value, which was being eroded by inflation. As early as 1975, California's Auditor General recommended that state and local agencies be allowed to invest in both security loans and reverse repurchase agreements, then illegal. Security loans would make it possible for a California agency to lend a stockbroker or financial institution one-year Treasury notes worth \$100,000; in exchange, they would receive cash or a promissory note *in excess of* the Treasury note's current market value as well as regular interest payments or dividends. Reverse repurchase agreements resembled security loans except that the contract was verbal instead of written and a date of repurchase was specified. In its report, the Auditor General estimated that the state and local agencies held more than \$7.8 billion in assets, and that in any single day as much as \$175 million could be temporarily transferred to financial institutions, providing an additional net income to California agencies of \$3 million per year.

While state officials introduced measures to preserve their assets, California *citizens* were taking steps to preserve theirs. Californians, especially retirees, were experiencing the effects of inflation with *both rising prices* and the *taxes* that they paid. Retirees on fixed incomes were not affected by bracket creep in their income tax as much as in their *property taxes*. For some time, counties had been reassessing property values and raising the tax rates applied to housing; by 1977–1978, some retirees were paying more in annual property taxes than they had initially paid for

their homes. This fueled the movement for property tax reform and Proposition 13.

Proposition 13 accomplished several things. First, it stipulated that all properties that had not been sold in two years (i.e., 1976–1978) would have their assessed values returned to 1975/1976 levels. Properties that had sold during this time would be valued at their market rate. Second, it set a ceiling for property tax rates at 1% of the property's value, and limited increases in the assessed value of property to 2% annually. Third, a provision, little noticed before the proposition's passage, made allocation of property tax revenues the prerogative of the state. To cope with this unsolicited responsibility, legislators passed AB 8 in 1979, establishing a formula for reimbursing property taxes to counties and other local governments. The formula was based largely on the city or county's *history of taxation and predisposition to self-tax in 1978*, as well as the unit's range of responsibilities and access to other sources of revenue. Finally, AB 8 required that 2/3 of the state legislature or voters, whichever was appropriate for the tax measure at issue, approve any future tax increases.

Historically, Orange County's taxes had been low, at least by California standards. This meant that the Legislature's formula for redistributing property taxes did not favor the county. County officials had tried to make legislators aware of their predicament. A study by the State's Legislative Analyst's Office, conducted after the bankruptcy, reported that in 1996/1997 Orange County received only \$40 in property tax revenues per person from Sacramento,^[1] placing it second from the bottom of all California counties. That year property tax returns to counties ranged from \$38 to \$1167 per capita, with the statewide average at \$106.

In the years immediately following Proposition 13, California was able to offset losses in county revenues with a special bailout. This stopgap measure sufficed as long as the economy remained strong and state tax receipts held up. These conditions held in California until the late 1980s. By 1987, however, strains were beginning to show. In August that year, Orange County officials testified before a joint hearing of the Senate and Assembly local government committees that they would be forced to lay off as many as 140 staff to meet the budget shortfall that year. Unable to accommodate the county's request for additional financial support, the Legislature adjusted the county's responsibilities for state mandated services.

Counties in California serve *three* functions—as administrative arms of the state, service providers for all county residents, and municipal service providers for those who reside in the county but outside the limits of any city. In their state administrative role, counties provided health care, welfare, and court services, and managed elections. Property taxes covered many

of their county and municipal services, as well as matching support for their state administrative responsibilities. Unfortunately, funding from the AB 8 formula was not adequate for all counties. Recognizing this, the Legislature passed at *least 9 bills* liberalizing county investment policies and as many as 19 bills overall between 1978 and 1994 relaxing safeguards on local government investments.^[2] The most important of these for Orange County were AB 346 in 1979, which authorized the use of reverse repurchase agreements, and AB 3576 in 1992 that enabled treasurers to invest as much as 20% of their portfolios in a variety of collateralized mortgage instruments. The Legislature accompanied this move toward greater autonomy with a relaxation of reporting requirements. In 1983, it passed SB 389, which eased the quarterly reporting requirements of local treasurers in effect since 1933. Although replaced the following year with stricter reporting requirements, these were allowed to expire in 1991. In 1993, the Legislature made treasurers' reports entirely voluntary.

In sum, Proposition 13 reduced the property taxes collected by counties and placed control for the taxes that were collected in the hands of the California Legislature. These measures hurt all counties, but affected Orange County more than most because the formula disadvantaged those with lower tax rates in 1978. This, in turn, forced local officials to look to non-traditional sources for revenue. One of these was greater development, because new housing was priced at market rates and could be assessed special fees that lay outside the parameters of Proposition 13. It also led officials to seek alternate sources of revenue, including new investment opportunities.

The impact was evident in the distribution of county revenue sources; property taxes, once accounting for more than 30% of all county revenues, fell to just over 15% in 16 years. In the meantime, *other revenues*, that had accounted for less than 1% of all county revenues in 1977–1978, rose to almost 3% by 1994. In Orange County, *other revenues* represented 12% of all revenues in 1993–1994, while property taxes contributed only 13.3%.

FINANCIAL INSTRUMENTS

Two investment miscalculations are generally credited with causing the bankruptcy: excessive use of derivatives and the leveraging of Orange County's financial portfolio. Derivatives are financial instruments, whose returns *are derived from* an underlying stock, commodity or other asset *other than* interest. Frequently, the return will be calculated by a formula, such as the difference between short- and long-term interest rates.

In 1994, more than 70% of the OCIP was invested in federal agency notes. These financial instruments are issued by U.S. government agencies and corporations, rather than by the U.S. Treasury. As such, they provide somewhat higher yields and carry a little more risk than Treasury bonds or notes. In OCIP, these agency notes were split 60–40 between those with *fixed-rate yields* and those with *floating rates*. The floating-rate notes thus accounted for almost 30% of the overall pool. Floating-rate notes differ from other agency notes and are often referred to as *structured notes*; they greatly resemble derivatives, both in their management of uncertainty and their difficulty to price. Structured notes, like other securities, have two components: principal and interest; however, their yields vary considerably.

The simplest structured note is one in which the yield *varies with* interest. These notes have great appeal because they adapt to market conditions, allowing the value of the principal to remain unaffected by the market, but providing higher yields when interest rates rise. A more complicated note is the *inverse floater*. This structured note provides yields that are *inversely related* to interest rates, usually calculated using the London inter-bank offer rate or LIBOR. For example, if the six-month yield (represented as 180/360 days) on a \$100 million note were set at 9% minus LIBOR and LIBOR were at 8%, the yield would be calculated as: $\$100 \text{ million} \times (9 - 8\%) \times 180 / 360 = \$500,000$. Inverse floaters may carry upper and lower limits within the formula to prevent negative yields. While the yields on inverse floaters may be attractive, the notes themselves are almost impossible to price. In the 1990s, large financial institutions purchased supercomputers for the sole purpose of calculating the prices of these and other complex financial instruments.

Federal agencies like the Federal National Mortgage Association (FNMA or Fannie Mae), the Federal Home Loan Bank (FHLB), the Federal Home Loan Mortgage Corporation (Freddie Mac), and the Student Loan Marketing Association (Sallie Mae) issue structured notes like the ones purchased by OCIP. The FHLB ventured into this area in 1986, in the aftermath of the savings and loan debacle, to protect federally subsidized mortgages against inflation. In 1994, *Orange County held \$100 million of FHLB structured notes*. In 1992, when LIBOR was at 3.625%, the yield on these notes was 8.25%, or, \$2,687,500 in the formula just cited; by November 1994, when LIBOR had risen to 6.5%, the yield had *declined to 2.5%*, or \$1.5 million in the same formula—almost half of the former yield. John Moorlach, Citron's opponent in the campaign for county treasurer, understood this dynamic; how much others did remains an open question. Nevertheless, neither the financial instruments nor their percentage distribution in the pool were illegal in California.

Leveraging is a form of borrowing that involves the use of existing assets to purchase *additional* assets. It might be compared to a home equity loan, which is normally obtained to make home improvements, with the house itself being used as collateral. If the home equity loan were used, instead, to purchase a second house, and that second house later sold at a profit, the home equity loan could be repaid in full, with interest, and the borrower might still make a profit. Citron essentially borrowed money against existing assets in OCIP (usually government agency bonds) in order to purchase additional assets (again, government bonds). However, since public borrowing is restricted by the California constitution, he purchased *reverse repurchase agreements* to accomplish this. At the time of the bankruptcy, Citron had leveraged almost \$7.5 billion into a fund worth \$20 billion, in effect *almost tripling the size* of his investment pool.

In the latter part of 1994, when it was clear the OCIP was in trouble, Citron repeatedly referred to *paper losses* in the pool—losses that he apparently believed would *not* be realized because he believed in holding securities to maturity. By leveraging his assets, however, through reverse repurchase agreements, *Citron exposed the bonds to market fluctuations*. This market exposure could have been mitigated with sufficient cash to repurchase the bonds. Citron had started the year with over \$1 billion in cash, so he had reason to believe that the losses *might* have remained only on the books. However, in February 1994, the Federal Reserve raised the discount rate from 3% to 3.25%. While the economy in California remained stagnant, business activity in other parts of the country was expanding and the Fed grew concerned that inflation would undermine the solid gains already made. By year's end, the Federal Reserve had raised the discount rate *six times*—an *unusual* but not unprecedented action.

By October 1994, OCIP's cash reserves had declined from over \$1 billion to less than \$450 million. Matthew Raabe, Citron's assistant, alerted the county's chief administrative officer, Ernie Schneider, to the problem. Schneider employed the services of Capital Market Risk Advisers (CMRA) to assess the pool's finances. Capital Market Risk Advisers found that OCIP had already lost \$1.5 billion and that there was not sufficient cash on hand to meet the interest payments coming due in December on voter-approved bonds. The crisis was precipitated when Crédit Suisse First Boston (CSFB), holding a \$110 million FNMA note with the county's promise to repurchase, announced its intent to sell the note on the market if Orange County did not come up with sufficient cash. The county had only \$100 million on hand. Crédit Suisse First Boston then announced its intent to call in *additional loans* of \$1.2 billion. County administrators

had explored other options to liquidate their portfolio, but sunshine laws governing their meetings, and time pressures from the banks, forced them into bankruptcy to prevent a run on their remaining assets.

In 1990, the state of California faced the first of several multibillion dollar budget shortfalls. The national economy had been slowing since 1989, but the end of the Cold War, with its concomitant reductions in defense expenditures, hit Californians particularly hard. Between 1990 and 1993, California lost 400,000 jobs and its unemployment rate jumped from 5.8% to 9.4%. This not only meant that the state could no longer afford to subsidize counties for their loss of property tax revenues, but that it would be forced to take extraordinary measures to address its own fiscal crisis. The state delegated many services to counties, *increasing* their costs.

THE AFTERMATH

Orange County citizens largely blamed *government* for the bankruptcy, as demonstrated in their June 1995 vote. [Chapter 9](#) had prevented participants in the pool from withdrawing their assets and protected the county from creditors, but it had *no effect on the general obligation bonds* the county had issued, which were coming due in 1995. In order, at least in part, to honor the county's commitment to bond holders, the newly appointed county administrator, William Popejoy, recommended a *sales tax increase* of one-half cent, *limited* to ten years for the county. This would have taken the sales tax from 7.75% to 8.25%. California counties share sales tax revenues with the state, but Orange County would have received the entire half-cent increase had Measure R passed, still returning 6.0416% to the state. Measure R required a 2/3 vote. Despite extensive publicity, Measure R was defeated with only 39% of those voting endorsing the proposition. In postelection polls, 70% of those who rejected the measure said they did so because they opposed new taxes under *any* circumstances or they did not believe they should pay for *someone else's* mistakes. To Wall Street and others in the financial community, such action was tantamount to an *abrogation of contract*. Indeed, the voters' refusal to honor their debt was almost as shocking as the bankruptcy itself; it undermined the very principle of security that has long been associated with public debt and that continues, at least in theory, to justify lower yields in the public sector.

The aftermath also witnessed numerous investigations into the causes and consequences of the bankruptcy—both official and unofficial. State courts found both Robert Citron and his assistant, Matthew Raabe, guilty of misappropriating public funds and

securities fraud. With Citron, it was not only that he had lied to investors about the safety of the investment pool, but that in the months before the bankruptcy he had illegally transferred funds across accounts in an attempt to forestall the inevitable. In the end, Citron served no time in jail but was fined \$100,000 and allowed to perform community service at the county jail for ten months. In granting Citron this leniency, the court recognized the treasurer's extensive co-operation with the investigation. Matthew Raabe, the Assistant Treasurer who had initially reported the problem to the Board of Supervisors and served very briefly as county treasurer, was not as fortunate. Raabe had had little role in devising the investment strategy but had full responsibility for explaining it to investors. Raabe was also found to have been involved in the illegal transfer of interest from the investment pool to county coffers when the county's financial strategy started to unravel. In May 1997, he was convicted on *five counts* of misappropriating public funds and securities fraud, later sentenced to three years in prison, and ordered to pay a \$10,000 fine. Throughout most of the proceedings, Raabe had insisted upon his innocence, arguing that he had been a loyal subordinate following the orders of his immediate supervisor.

State Superior Court proceedings were initiated against several others, but no one was convicted or judged negligent. Non-criminal charges of willful misconduct were dropped against two members of the Board of Supervisors. The budget director, Ronald Rubino, was implicated in the misappropriation charges, but a jury failed to agree on a verdict. Later, he pleaded no contest to one misdemeanor and received 100 hr of community service. The county brought charges against Merrill Lynch in both federal criminal and civil court. After two-and-a-half years and 5000 pages of testimony, Judge Capizzi allowed Merrill Lynch to settle the criminal suit for \$30 million. A civil suit for \$2 billion was settled out of court in June 1998 with Merrill Lynch agreeing to pay \$417 million in exchange for the county's dropping all of its charges. Merrill Lynch also returned \$20 million in county funds it had been holding since the bankruptcy. Other investment firms also settled with the county. KPMG Peat Marwick L.L.P. settled for \$75 million in 1998. LeBoeuf, Lamb, Greene and MacRae agreed to restore \$12 million to the county's Community College District, making it the first depositor to have been refunded in full. Finally, CSFB settled with Orange County for \$55 million in May 1998 after CSFB had been sued by the Securities and Exchange Commission (SEC) for fraud and deceptive securities practices, hastening the bankruptcy.

The State Auditor, Kurt R. Sjoberg, focused on Citron and the county's investment strategy. Sjoberg found

that Citron had exposed the county to excessive risk by failing to observe the basic public investment priorities of safety, liquidity, and yield. This was due, as we have seen, to the leveraging arrangement accomplished through reverse repurchase agreements and the purchase of interest rate sensitive, structured notes at a time when interest rates were rising. In addition, the Auditor noted that Orange County had managed several pools: a commingled investment pool, a commingled bond investment pool, and several separate investment accounts. The State Auditor recommended to Legislators that they restrict reverse repurchase agreements to 20% of a unit's portfolio, prohibit the use of debt for speculative purposes, require regular financial reports, and establish a prudent person rule for all local investment officers.

The California Senate's Special Committee on Local Government Investments cast the blame more broadly. Committee members attributed the bankruptcy to a "lethal mixture of corporate and political culture, where the public trust has taken a back seat to corporate profit and political expediency."^[3] Specifically, they recognized that Citron had sacrificed the conventional standards of safety, liquidity, and yield for higher profits and had done so by imprudently leveraging the pool's assets. At the same time, they acknowledged that 35% of the county's *own source* revenues had derived from interest income and that there had been tremendous pressure on the Treasurer to offset the post-Proposition 13 shortfall in county revenues; accordingly, they charged the Board of Supervisors with inadequately supervising the Treasurer's activities. Members did not limit their criticism to Orange County, however; they castigated bond-rating agencies for failing to identify any risk in the county's portfolio and charged them with actively contributing to the bankruptcy. The committee also drew attention to both Moody's and Standard and Poor's responses to John Moorlach's 1994 challenge when both agencies accused him of engaging in political rhetoric. The committee found this intrusion into the political debate at the *very* least unprofessional. It also criticized securities agencies and brokerages for failing to recognize the portfolio's risk. While the advisory group did not go so far as to include financial advisers among those *responsible* for the bankruptcy, it did blame securities firms for *narrowly interpreting* their role and charged them with "insufficient regard for the special nature of public funds."^[4] Finally, the committee drew attention to the *victims* of the bankruptcy—the 187 cities, schools and special districts that had invested in the pool and whose savings were now in jeopardy. Rather than spare them, however, it charged them with abrogating their own fiduciary responsibilities by failing to raise questions about the pool and the treasurer's

investment strategy. In effect, the Special Committee suggested that officials reaped the benefits of Citron's strategy while being unwilling to delve into its details for fear of what they might discover.

The Securities and Exchange Commission conducted several investigations in the aftermath of the bankruptcy. Among the more salient were investigations into the securities activities of county officials, the influence of brokerage firms in crafting state investment legislation, and the potential conflict of interest of firms like Merrill Lynch in advising public officials. It censured Citron, Raabe, CAO Steiner, and four members of the Board of Supervisors for misleading and defrauding pool participants, although no formal admission of guilt was required or fine extracted. The parties did agree not to violate securities law in the future. While the SEC found no direct evidence of misconduct on the part of brokerage firms in the passage of state legislation, it did issue subsequent rulings prohibiting campaign contributions to officials with securities responsibility within two years of any formal transactions. It also worked with investment banks to develop *self-regulating procedures* in which the firms will disclose to the SEC information on their trades, risks and customers. Finally, the SEC developed educational packages on derivatives and strategies of risk management for state and local officials and promulgated *new rules* making brokerage firms legally responsible for taking their public clients' financial goals and *fiduciary responsibilities* into consideration when advising on investments.

The financial community responded to the bankruptcy with increased reliance on bond insurance for managing municipal debt. Prior to the bankruptcy, only municipalities with low bond ratings had been required to take out insurance; afterwards, bond insurance became almost commonplace for a while. It was the financial community's way of protecting investors against citizen abrogation of their debt responsibilities and this increased the public's cost of borrowing.

Public administration scholars responded to the bankruptcy, and the events in its aftermath, either by focusing on the risks involved in derivatives purchasing and portfolio leveraging, or by concentrating on the normative aspects of public finance and the fiduciary role that public investors must also assume. Less attention was paid to the fact that the only person actually to spend any time in jail was Ken Raabe—Robert Citron's appointed administrator in the Treasurer's Office.

REFERENCES

1. California. Legislature. Joint Legislative Budget Committee. Legislative Analyst. In *Why County Revenues*

- Vary: *State Laws and Local Conditions Affecting County Finance: An LAO Report*; Legislative Analyst's Office: Sacramento, CA, 1998; 10 pp.
2. Baldassare, M. *When Government Fails: The Orange County Bankruptcy*; University of California Press: Berkeley, CA, 1998; 71–75.
 3. California. Legislature. Senate. Special Committee on Local Government Investments. In *The Orange County Bankruptcy: Broad Repercussions, New Public Policy: Report of the Senate Special Committee on Local Government Investments*; Senate Publications: Sacramento, CA, 1995; 29 pp.
 4. California. Legislature. Senate. Special Committee on Local Government Investments. In *The Orange County Bankruptcy: Broad Repercussions, New Public Policy: Report of the Senate Special Committee on Local Government Investments*; Senate Publications: Sacramento, CA, 1995; 37 pp.

BIBLIOGRAPHY

1. California. Bureau of State Audits. In *Orange County: Treasurer's Investment Strategy was Excessively Risky and Violated the Public Trust*; California Bureau of State Audit: Sacramento, CA, 1995.
2. California Legislature. Assembly. Committee on Local Government. In *County Fiscal Problems: A Staff Summary Report for the Joint Hearing by the Assembly Local Government Committee*; Joint Publications Office: Sacramento, CA, 1987.
3. California Legislature. Assembly. Committee on Local Government. In *Summary of Proceedings, Alternatives for Reforming Local Government Finance: Summary Report of the Interim Hearing of the Assembly Local Government Committee*; Joint Publications Office: Sacramento, CA, 1990.
4. California Legislature. Assembly. Local Government Committee. In *Summary of Proceedings: Interim Hearing on Allocation of Sales Tax and Other Local Government Revenues: Developing Workable Incentives for Balanced Development*; The Assembly: Sacramento, CA, 1994.
5. California Legislature. Joint Legislative Budget Committee. Legislative Analyst. In *Why County Revenues Vary: State Laws and Local Conditions Affecting County Finance: An LAO Report*; Legislative Analyst's Office: Sacramento, CA, 1998; 1–24.
6. California Legislature. Senate. Special Committee on Local Government Investments. In *The Orange County Bankruptcy: Broad Repercussions, New Public Policy: Report of the Senate Special Committee on Local Government Investments*; Senate Publications: Sacramento, CA, 1995.
7. California. Office of the Auditor General. In *Evaluation of General Proposals to Allow State and Local Investment Authorities in California to Increase Interest Income by Temporarily Lending Investment Securities Through "Security Loans" and "Reverse Repurchase Agreements"*; California. Office of the Auditor General: Sacramento, CA, 1975.
8. Fulton, W. Marian Bergeson: Working at Home. In *Governing*; 1996, February 60.
9. Jorion, P. *Big Bets Gone Bad: Derivatives and Bankruptcy in Orange County*; Academic Press: San Diego, CA, 1995.
10. MacDonald, S.H. Alternative Responses to the Orange County Bankruptcy: An Inquiry into the Images Underlying Theory. A dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University: Blacksburg, VA, August 2000.
11. McCarthy, Ed.; Derivatives revisited. *Journal of Accountancy* 1999, 189 (5), 35–43.

Organization Development

Gerald T. Gabris

Division of Public Administration, Northern Illinois University, DeKalb, Illinois, U.S.A.

INTRODUCTION

Organization development or “OD” as most practitioners refer to it, eludes simple definition. Nonetheless, several attempts to define OD provide perspective. W. Warner Burke defined OD as “a planned process of change in an organization’s culture through the use of behavioral science technology, research and theory.”^[1] French and Bell provided a similar assessment of OD as “a top-management supported, long range effort to improve an organization’s problem-solving and renewal processes, particularly through a more effective and collaborative diagnosis and management of organization culture—with special emphasis on formal work team, temporary team, and inter-group culture.”^[2] In both definitions, OD was linked to a planned change effort in an organization’s culture using social science technology. Burke went so far as to state that when an OD effort takes place, there is no OD if no change in culture occurs.

Yet other practitioners and scholars defined OD in even broader strokes. Robert T. Golembiewski suggested that OD be best understood as a “stool.”^[3] Moreover, he goes on to say that the denotation of “stool” he is referring to involves the three-legged small-chair motif, where each supporting leg represents part of what OD does. In this perspective, OD involves attempts to modify the “structure” of an organization. This is a particularly common OD focus in the public sector.^[4] Yet, OD often involves cultural interventions in the context of working toward more responsible and mature interpersonal relationships. Finally, OD may involve the development or reformulation of organizational policies. In this sense, OD is a multifaceted discipline that may focus on multiple targets, all aimed at helping an organization address a perceived need or problem.

Frequently, but not always, OD utilizes an external consultant, known as an intervenor, who collaboratively works with the organization to diagnose problems and to help design potential solutions (known as interventions). This leads to another important aspect of OD as an organizational discipline.

OD is Value Loaded

The genesis and emergence of OD as an organizational change framework did not emerge from a vacuum but

from the rich history of human relations management beginning in the 1940s. Leading scholars and nascent practitioners such as Kurt Lewin, Douglas McGregor, Rensis Likert, Robert Blake, and Chris Argyris, to mention just a few, imbued OD with a deep-rooted appreciation of the human element and human values in the work environment. Risking some oversimplification, OD began to stand for:

- Employee dignity and respect.
- Employee participation in decision making.
- Responsible freedom and autonomy.
- Treatment of employees as mature adults.
- Stress on employee growth.
- Development of meaningful jobs.
- Recognition of human potential and value.

Given these values, OD strives to avoid conventional models of consultation, where the consultant serves merely as an external expert or as a physician to a sick patient.^[5] Instead, OD considers the role of external consultant as a collaborator, a catalyst, a role model, and as someone who will take a stand for advancing the human value of employees in the work organization. Given this value-loaded perspective, why do organizations opt for an OD problem-solving approach, and what do they hope to gain from it?

OD AS DISTINCT FROM CONVENTIONAL MANAGEMENT CHANGE MODELS

OD does not suggest it is for everyone and all organizations. Unlike many conventional approaches to organization improvement that advocate a one-size-fits-all model [see *The One Minute Manager* (Ref. ^[6])], OD utilizes an open-ended discovery process for designing potential solutions to perceived problems. In OD, this process is called “action-research.”

As W. Warner Burke avers, in practice, the words “action-research” are reversed, because research is usually conducted before action takes place.^[1] Action-research, a term originally coined by John Collier^[7] and Kurt Lewin,^[8] refers to the “methodology” common to most OD efforts. Although not

always in tidy, lock-step sequences, action research involves:

- Identification of a felt need.
- Diagnostic research to assess the need.
- Feedback between a consultant and client.
- Collaborative problem solving (intervention).
- Evaluation of results.
- Termination of effort or more diagnosis followed by more feedback and intervention.

A key point is that in OD, it is not assumed that a problem exists and the nature of a problem is not speculated until a client organization feels a strong enough need to do something about it. Once a felt need surfaces, OD consultants conduct research to obtain a better handle on the nature and depth of a particular problem, so that any treatment addresses the causes and not simply the symptoms of a nettlesome issue. Feedback is used to test the diagnosis. If feedback suggests that the problem has been identified correctly, and if the need to do something remains intense, then a collaborative design between the client organization and consultant to address the problem ensues. A collaborative design partially insures that the client organization assumes a degree of “ownership” for the solution and will actively commit to making it work. Subsequent evaluation assesses the degree of success achieved by the intervention and can provide perspective on what, if any, additional steps should be taken. According to Golembiewski, the degree of OD intervention success in the public sector for the United States approaches 85%.^[4]

WHAT KINDS OF THINGS DOES OD DO FOR A CLIENT ORGANIZATION?

Fundamentally, OD customizes planned change processes around the specific needs of the client organization. Nonetheless, several intervention technologies recur frequently as strategies for dealing with a panoply of problems or issues. It should be repeated that these technologies are not utilized randomly as panaceas to organizational problems, but instead, are only applied after careful analysis suggests the best intervention for a given situation. Further, intervention technologies have evolved over time, and new ones are continually emerging, so those described below are just the tip of the iceberg.

In the 1950s and 1960s, a primary intervention tool employed by OD’ers involved “training groups” which have also been called T-Groups or laboratory groups. Typically, managers and executives from different (stranger) organizations would become participants in a larger group led by a trained facilitator. These groups would generate their own data for

discussion (they served as their own laboratories) based on open-ended comments and observations made by group participants. No particular problem was preidentified for discussion. Groups could explore just about any issue they wanted. From these sessions, participants honed their own communication, confrontation, and interpersonal skills which could then be brought back to their home organizations. In this way, T groups could help “seed” work organizations with persons possessing progressive interpersonal skills and abilities, thus enabling their organizations to grow and adapt more effectively.

While some practitioners lament the passing of T-Groups, and in particular, the group facilitation skills these sessions imparted on those leading the sessions,^[3] other more targeted and specialized group interventions have come to predominate the tool bag used by most OD practitioners. The term used most often to describe this category of interventions is “team building.”^[9]

To the casual observer, team building may simply involve the notion of coworkers cooperating nicely on group projects. Having employees work together on joint projects, communicating openly, and developing creative solutions while avoiding conflict may certainly be one aspect of a sound “work team.” Yet in OD parlance, “team building” involves a much larger array of potential intervention strategies based on diagnostic results. Golembiewski, Munzenrider, and Stevenson^[10] provide insight into this idea by differentiating between “conflicts of agreement” versus “conflicts of disagreement.”

Suppose an external consultant asks labor and management within the same client organization the following simple questions:

1. What is going right in this organization?
2. What is going wrong in this organization?
3. What should be done to correct any problems?

Now, it may evolve where labor and management perceive the same things going right, the same things going wrong, and have in general the same solutions—but for some reason, they have not communicated these to each other. Under this scenario, they may think they are in disagreement, when in fact, they agree, but do not know it. Hence, each may act on its own, thinking the other party does not understand its position. This is a crisis of agreement. Here, just showing the labor and management groups that they actually agree may be sufficient to focus energy on joint solutions. Alternatively, the same diagnostic questions may show the two groups to be in genuine disagreement. They may hold different perceptions as to what problems exist and what should be done about them. The point is that the type of team-building strategies utilized for crises of agreement may vary substantially from crises of disagreement.

If crises of disagreement are found to exist, then third-party interventions involving conflict resolution, confrontation, and similar interaction-centered designs may be appropriate.^[10] Here, the need is to get the parties to a dispute talking and thinking about ways of mitigating conflict. Here, the intervenor may serve as an effective go-between, trying to construct an acceptable third party, neutral, consensus-based contract. These types of team building interventions may take a long time to implement, may involve several meetings, and may require detailed agreements between opposing parties.

Alternatively, if crises of agreement prevail, more task-centered team building may be appropriate. Simple examples may include strategic planning, goal setting, role analysis, visioning, and creative problem solving. Usually, interpersonal conflict is low in such situations, thus allowing the group to focus on tasks rather than on interpersonal relationships. Again, the type of team building that works best may vary from situation to situation.

OD has a much greater range of intervention techniques than T-Groups and or team building. Another whole area of intervention can be referred to as “process consultation.” Whereas team building tends to focus on concrete, tangible problems or “content,” process consultation focuses on organizational processes. The concept of process consultation was first described by Edgar Schein as a kind of consulting philosophy or attitude toward helping individuals and groups within organizations.^[5] He goes on to define process consultation as “a set of activities on the part of the consultant that help the client perceive, understand, and act upon the process events that occur in the client’s environment in order to improve the situation as defined by the client.”^[5] As such, process consultation is not a technique or set of techniques, but an approach to consulting.

Schein points out that all organizations have many processes they take for granted but that nonetheless can cause problems if not managed properly or if misunderstood. Examples of processes include communications, problem solving, decision making, and role playing. A process consultant will carefully “observe” how a client organization behaves in performing these processes. The process consultant does this by blending into the work environment in the least obtrusive manner possible, so as to gain the trust of workers and to open channels of communication. Over time, the consultant may become cognizant of patterns of behavior that have implications, say on interpersonal relationships that are either misunderstood, unknown, or misperceived by the client. When the consultant feels the client is ready for feedback, the consultant’s observations may be made known, and then, both may work toward a collaborative solution.

A concrete example of how a process may be misperceived involves group decision making. The chair of a group may summarize ideas about an issue and then ask if anyone has additional comments to make before a decision is made. Silence ensues, so the chair assumes all group members are in consensus. Later, the chair learns that two group members dislike the decision but remained quiet fearing that their views would not be taken seriously. Unless this attitude and behavior of certain group members was made known, the chair could continue making decisions under the guise of consensus, when in fact, intense disagreement may be more the case. Process consultation involves working through such troubling process-based situations over time.

Besides team building and process consultation, OD possesses many other intervention technologies and strategies, too many to describe in such a short encyclopedia entry as this. Some worth mentioning include job design and structuring,^[11] pay policies and incentives,^[12] stream analysis,^[13] and burnout and stress management.^[10,14] One could go on. For an excellent summary of potential consulting interventions, see Golembiewski’s *Handbook of Organizational Consultation*.^[15]

CONCLUSION

Organizational development is a powerful tool in the quiver of arrows available to managers for improving public organizations. Never has it claimed to be a panacea or complete solution to all problems. On the contrary, OD should be used sparingly and in situations where it has a high potential for success. The client organization has to want to do OD. There must be a strongly felt need for this type of approach to succeed. OD involves total system change, its aim is to move the organization in a planned direction and in a way that respects human value and dignity. Hopefully, this brief discussion may help some public managers have a better sense of what OD might be able to do for them. OD exists not to reinvent the status quo, but to challenge it and to open new vistas. Only the client organization under the guidance of a skillful OD practitioner may know when the time is right. Might the time be ripe to at least consider this possibility?

REFERENCES

1. Burke, W.W. *Organization Development: Principles and Practices*; Little, Brown, and Co.: Boston, MA, 1982.
2. French, W.; Bell, C. *Organization Development*; Prentice-Hall: Englewood Cliffs, NJ, 1990.

3. Golembiewski, R.T. *Organization Development: Ideas and Issues*; Transaction: New Brunswick, NJ, 1989.
4. Golembiewski, R.T. *Humanizing Public Organizations*; Lomond Publications: Mt. Airy, NJ, 1985.
5. Schein, E. *Process Consultation: Volume I*; Addison-Wesley: Reading, MA, 1988.
6. Blanchard, K.; Johnson, S. *The One Minute Manager*; Berkely Books: New York, 1982.
7. Collier, J. United States Indian administration as a laboratory of ethnic relations. *Soc. Res.* **1945**, *12*, 275–276, (May).
8. Lewin, K. Action research and minority problems. *J. Soc. Issues* **1946**, *2*, 34–46.
9. Dyer, W. *Team Building: Issues and Alternatives*; Addison-Wesley: Reading, MA, 1987.
10. Golembiewski, R.; Munzenrider, R.; Stevenson, J. *Stress in Organizations*; Praeger: New York, 1986.
11. Hackman, R.; Oldham, G. *Work Redesign*; Addison-Wesley: Reading, MA, 1980.
12. Lawler, E. *Pay and Organization Development*; Addison-Wesley: Reading, MA, 1981.
13. Porras, J. *Stream Analysis*; Addison-Wesley: Reading, MA, 1987.
14. Maslach, C. *Burnout—The Cost of Caring*; Prentice-Hall: New York, 1982.
15. Golembiewski, R.T., Ed.; *Handbook of Organizational Consultation*; Marcel Dekker: New York, 1993.

Organization Theories, Types of

Robert T. Golembiewski

Department of Political Science, University of Georgia, Athens, Georgia, U.S.A.

INTRODUCTION

One classic view of types of organization theories implies that they were much ado about little or nothing. As the doggerel of an earlier day put it:

Over forms of organization,
Let fools contest;

Whatever is best administered

Is best.

The basic approach here is more demanding and less definitive. Broadly, the present view is that forms of organization matter: given different situations and their features, different structures will have different consequences, even profoundly different consequences. At times, more or less, we will be able to urge that a specific structural model has bad→good fit with specific situational features. This case was first made in comparing conditions of turbulence versus placidity, and the fully developed comparisons were labeled “situational theory.” The bureaucratic model is better suited to placid environments, while structural innovations are appropriate for turbulent conditions. The latter case refers to more complex technologies, rapid change, and complicated product lines with shortening half-lives.

Moreover, a second kind of “fit” also relates to structure. Students of management often talk about “best practices” or “good practices,” but that is inelegant as well as incorrect. A practice may have good or bad or indifferent effects, in general; but more specific predictions will depend on the organizational structure and culture into which the practice is inserted. Commonly, even what most regard as a “good practice” may have inconsequential or even “bad” effects when incautiously inserted into the wrong structural contexts.

NOTES ON A TAXONOMY OF STRUCTURES AND THEIR CONSEQUENCES

More about such important issues associated with protean forms of “goodness of fit” will come later.

Immediate attention turns to an ideal way of dealing with structural issues in the aggregate, as well as to the present state of our reach-and-grasp in this central particular.

A General Ideal Model

Fig. 1 sketches an ideal taxonomy for present purposes, focusing on a matrix of organizational types or archetypes as well as on the “marker variables” that will provide texture and tone to the similarities and differences within or between the archetypes. A complete form of the ideal typology, that is to say, will provide two kinds of information about organizations that will permit confident and comprehensive description and prediction. Thus, that ideal will be built around an exhaustive catalog of organization types; and the typology will provide differential content for the several archetypes via a large number of “marker variables,” e.g., percentage low or high on involvement, participation, and so on. In plainer words, that ideal will identify distinct categories of organizations as well as the particular features on which the categories differ or are similar. The two cornerstones of the matrix have simple labels: they label each archetype; and these labels may be descriptive (such as High/Authoritarian–Low Participation–Low Involvement collectivities), or they may be clinical (e.g., paranoid), or they may reflect some of both.

The Present Taxonomic Status

We are a long way from having a relatively comprehensive version of the taxonomy that Fig. 1 illustrates, but several important points seem significant enough about the progress that has been made. Thus:

- The number of archetypes is not known, and perhaps there never will be a maximum: the working versions now differ widely on this central particular and, absent specificity about archetypes, cumulative research is highly impossible. Why? Directly, proportions of archetypes will differ in specific research batches in unknown ways, and often this will obscure or camouflage any associations that seem

		Organizational Archetypes			
		I	II	...	N
Marker Variables					
1	Structured Variables (e.g., centralization)	High	Low		High
2...	Style Variables (e.g., coercive)	Low	High		Low
.					
.					
N	Population Variables (e.g., authoritarian)	High	High		High

Fig. 1 Sketch of a useful taxonomy.

- to exist in one batch or another. Hence, the “mixed” character of existing research.
- Recently, statistical techniques have been developed to isolate patterns of differences and similarities for many collectivities described by numerous variables. Early aggregations now identify 10 to 12 archetypes (e.g., Ref.^[1]), but the number of archetypes may be subject to periodic expansion, as we become more capable of finer tuning.
 - Most organizational research neglects nuances distinguishing the archetypes, and focuses on associations between two or a few marker variables in an organization.^[2] This constitutes a major reason why research results are mixed.
 - The mixed results often have suggested to some that nature is chaotic and, hence, essentially unknowable. The present belief is that this goes much too far, but the truth remains: tests of associations between marker variables in relation to taxonomies have been largely restricted to the study of individuals (e.g., Ref.^[3]).

In sum, the existing confusion in the organization behavior literature is understandable, and perhaps even unavoidable. But perhaps, also, we will soon begin cumulative work toward a comprehensive taxonomy, as the next section suggests.

Some progress toward organizational archetypes

If incomplete, some noteworthy progress has been made toward isolating a panel of organizational archetypes; and both progress and limitations get attention here. Progress can be illustrated conveniently by some long-standing descriptive work on the bureaucratic and flow-of-work models. This progress gets illustrated by two kinds of efforts: orientations toward budgets and forces toward greater or lesser productivity. Large volumes of detail are available elsewhere. To rely only on this author’s research, useful summaries are possible of clinically oriented descriptions of consequences of

the two models (e.g., Refs.^[2,4–5]), as well as related to substantive issues (e.g., Ref.^[6]). In Fig. 2, conveniently, activities A+B+C yield some product or service.

Toward Two Basic Archetypes

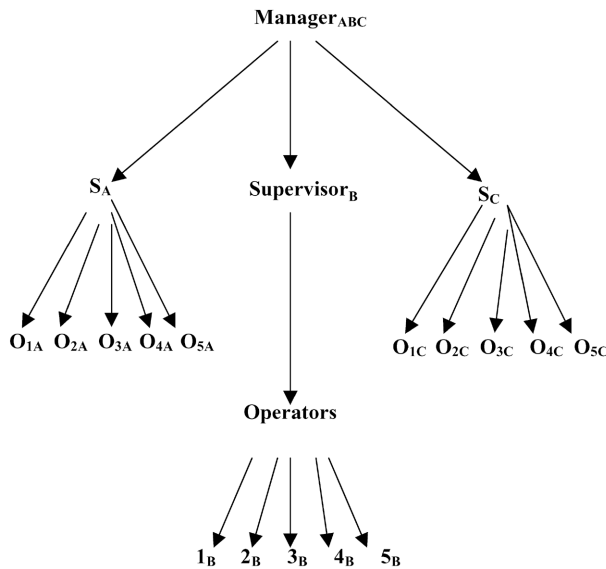
For many decades now, archetypal description has focused on two basic models. Reasonable but incomplete confidence can be placed in these alternative descriptions, and their consequences will be sampled. All in all, the samplers are perhaps best described as pre-archetypal forms or descriptions on their way to archetypal forms.

Some consequences of bureaucratic model

With predecessors extending at least from biblical days, the bureaucratic model has dominated in organizations. Fig. 2A provides skeletal details for this model. In details as well as in substance, overall, Fig. 2A can with relative safety be associated with a broad range of deleterious consequences, on average. Consider only three sets of these probabilities.

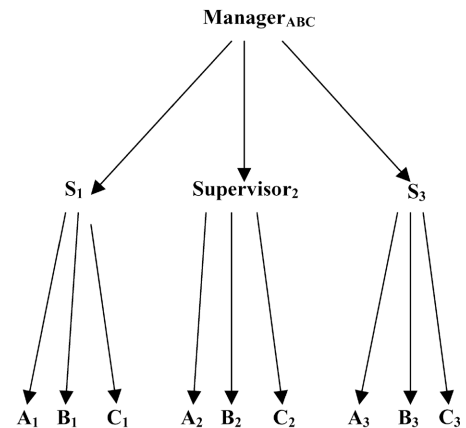
First, S in Fig. 2A tends to have particularistic biases, at the expense of systemic achievement. For example, Supervisor B in the bureaucratic model is generally motivated to increase budget and policy support for B activities, while S_A and S_C have similar motivation for focusing on A and C, respectively. And, that is the way matters should be, and will be, given only there that a minority of cases in which people can transcend self-interest over the long run. Reinforcing this tendency, M_{ABC} typically has no valid and reliable ways to assign costs to each S; and all Ss often will resist having costs assigned to their unit, and especially when any doubt exists, which is the usual case. Hence, one common feature of life in bureaucratic models: what is inelegantly but accurately described as throwing dead cats in the backyards of other units to avoid budgetary costs.

A



- departmentation is based on separate functions or activities
- reporting relationships are vertical
- span of control is small: only a few subordinates report to any supervisor

B



- departmentation is based on total flows of work
- reporting relationships are both vertical and horizontal
- span of control can be large

Fig. 2 Two pre-archetypal forms: (A) Basic bureaucratic model, (B) Flow-of-work model.

Consider, second, a related awkward feature of Fig. 2A structures: they encourage a downward orientation in output levels. To explain, only all three S units can raise output in the bureaucratic model, but any single S unit can and does in a Fig. 2A structure induce powerful forces to lower output.

The greater convenience of lowering versus raising output does not inevitably lead to reduction, but the effect is powerful, given the bureaucratic model.

Finally, with respect to an amalgam of control and responsibility, common powerful pressures in the bureaucratic model generate awkward consequences in any single S units, and especially at the beginning or end of sequences, like $A \rightarrow C$. If the S_C unit decides to reduce output, that may leave the other two units no reasonable choice but to go along or foolishly maintain higher levels. In the latter case, S_C might argue with some force that this only proves the basic point: that activities A and B are simpler, and also that the plain record justifies additional resources for C.

Some consequences of flow-of-work model

Arguments contrary to the samplers above can be generated for Fig. 2B structures; which gained great and growing popularity in the past 3 or 4 decades. At senior levels, this alternative often is called the “divisional model” (e.g., Ref.^[7]); and at low levels of

organization, this structure is associated with job enrichment and enlargement as well as with cross-training and job rotation.

Cross-training and job enrichment and enlargement are common prescriptions for organizing work today, and they are disingenuously often recommended as best or good practices while retaining the bureaucratic model. That is awkward advice, but it can only succeed despite itself and at the common cost of low levels of probability. Thus, cross-training in Fig. 2A requires the participation of all three supervisory units. This is clearly more difficult than in Fig. 2B, where each supervisory unit can engage in the project, or not, and with each having the direct opportunity to pay any costs while reaping all benefits.

Similar differences characterize each major stage of cross-training innovations, including selection of trainees, training, evaluation, and real-time application of any learning. Perhaps basically, Fig. 2A places the major burdens on each S, but with M_{ABC} the major recipient of any positive consequences, and one can easily go broke betting in favor of such an outcome. This is obviously an arrangement that requires Ss to transcend their immediate self-interests. Most undercutting is this obvious fact: in Fig. 2A, there is no direct way in which (for example) As cross-trained in B can put their greater mastery to direct use in their home unit. In Fig. 2B, in contrast, any S unit provides

sufficient scope for the full set of innovation stages, with all appropriate rewards and punishments being experienced in each involved S unit. To illustrate, direct applications of any cross-training consequences are possible in Fig. 2B structures in each S unit. Here, positive consequences of cross-training may be diffused from the positive experience of even a single S unit—possibly and even probably. In Fig. 2A, reliance often must be placed in command→obey dynamics, which have their serious limitations, and especially where bureaucratic structures encourage narrow self-interest.

Four central points

These structural comparisons could be extended broadly, but four central points will carry that burden here. First, and of greatest practical and theoretic significance, the failure of most good practices reforms is encouraged by retaining the basic structure of work. To illustrate, even desirable innovations or recommendations for (e.g.) cross-training can die as add-ons in the context of a traditional structure (e.g., Ref.^[8]).

This amounts to a bandage that leaves the basic cause in place. Basic structural change typically cannot be finessed without costs, in sum, even though this crucial detail usually gets neglected.

Second, the basic Fig. 2A versus B contrasts are here illustrated at low levels of organization, but a similar case can be developed at higher levels of organization around such flow-of-work alternatives as products (e.g., Ref.^[6]). Indeed, Fig. 2B most prominently appeared in the “division model”^[7] in research directed at large units of business organizations.

Third, the comparative relevancies of alternative structures is no Johnny-come-lately. To be sure, the higher-level limitations of the bureaucratic model were emphasized influentially in business (e.g., Ref.^[7]) and only later in government (e.g., Refs.^{[2],[5]}). But, the essential dynamics were clear at early times in both arenas (e.g., Refs.^[9–12]).

Fourth, the choice of a structural model often is influenced by considerations other than comparative dynamics, so a structural model must be accepted politically as well as generate better effects. This may seem paradoxical, but the two conditions do not necessarily occur together. For example, in U.S. Social Security operations, movement from a Fig. 2A to a 2B structure was accompanied by major increases in morale and productivity, as is to be generally expected (e.g., Refs.^{[4],[6]}). Detailed contrasts develop the point (e.g., Refs.^{[6],[13]}). While these positive effects were still being felt, however, Social Security officials became enamored of computer applications, which to them implied the need to return to a Fig. 2A structure. This is a conventional view but, for reasons far beyond the present goals, is more common than convincing or necessary.

Toward organizational archetypes: basic limitations

In sum, as the brief summaries above indicate, two preliminary archetypes permit some real progress toward comprehending structural dynamics in organizations. But, “some” is clearly not enough, as two points will illustrate for present purposes: a distinction between “mock” and “real” structural forms, and a review of what several observers see as useful elements in a full set of archetypes that eventually will comprehend organizational phenomena.

“Mock” Versus “Real” Theories

As Gouldner (Ref.^[14]) advised early on, any prescriptions can be differentiated in application as “mock” or “real,” to focus on only the simplest form of the distinction. To illustrate, a mass transit system articulated a set of reporting relationships and associated policies that could be labeled bureaucratic, but under the press of “building the railroad”—a multibillion dollar undertaking with tough schedules and tight budgets—many modifications and clear contradictions with those prescriptions were acted on, and even agreed to (e.g., Ref.^[15]). To a substantial degree and for many purposes, then, a “mock” formal infrastructure may be said to have existed but was often superseded by conscious but limited efforts at planned change.

Each of the structures in Fig. 2 can be “mock” to various degrees, and that makes for some consequential differences in the associated dynamics. For example, a “mock” version of Fig. 2A seems unlikely to generate “effective centralization,” but determined efforts by a charismatic leader might succeed in that effort even though bureaucratic structures can contribute to “chaotic localism” in ways already suggested above, e.g., in lowering rates of production or in multiple veto centers.

This is not to say that “mock” artifacts are always *faux*, or impotent, in addition. In the mass transit authority, for example, once the rail system was in operation, changing coalitions-of-interests began to vivify some of the “mock” features and to hold people responsible for what the system earlier prescribed but only incompletely acted upon (e.g., Ref.^[15]).

How Many Models is Enough?

The simple contrast above implies that Fig. 2 does not provide enough theoretical power, and that basic point can be reinforced by a brief review of other efforts toward taxonomies of organization theories. Thus, your author recognizes that two models are not sufficient, although they can support useful analysis. Elsewhere, (Ref.^[16]) your author provides substantial

detail about a third form—the matrix model—which adds usefully to the two models in Fig. 2.

Others go substantially further, and this acknowledges that Fig. 2 provides only a useful start. Thus, Mintzberg's (see Ref.^[17]) clinical synthesis provides for six structural variants:

- The “machine model” like that in Fig. 2A.
- The “professional bureaucracy,” which is much like Fig. 2A but features the addition of often complex staff and other services.
- The “entrepreneurial structure,” which focuses on a charismatic central figure who innovates and supervises an organization with little middle management support.
- The “professional organizations” in which the technically skilled employees work under flexible supervisors, ideally informed by guiding standards.
- The “innovative organization,” or *ad hoc*-racy, which relies on temporary and often-spontaneous infrastructures for specific purposes.
- A divisional structure like that in Fig. 2B.

Clearly, Mintzberg incorporates the conceptual territory covered by Fig. 2 and pushes substantially beyond it. Other exemplars employ advanced statistical methods to aid in such pushes (e.g., Ref.^[1]).

CONCLUSION

The work on organization theories, then, represents progress but only at early developmental forms of a full taxonomy. Basically, extended versions will provide increasingly powerful tools for description and prediction. The goals are that a progressively comprehensive set of archetypes will be described, and their consequences as well as covariants will be specified with substantial and growing accuracy. Beyond that, such comprehensive taxonomies can guide systemic change, as illustrated by first efforts at motivating a different structure for the U.S. federal system.^[5]

Despite some progress, versions of a complete taxonomy now lay basically in our future. Existing theories and models serve many useful purposes, but they nonetheless fall substantially short of what can be envisioned in some detail. Usefully, however, analytic ways to approach that ideal seem to have recently become available.

Note also that levels of “good fit” will have to be accommodated. Thus, this description suggests the many ways in which Fig. 2B structures have a poor fit with the bureaucratic model. At a macro level (e.g., Ref.^[5]), Fig. 2B structures also have major features of poor fit with American constitutional features. Moving beyond the bureaucratic model in

government, thus, in short, features challenges beyond those encountered in the business arena.

REFERENCES

1. Miller, D.; Friesen, P.H. *Organizations*; Prentice Hall: Englewood Cliffs, NJ, 1984.
2. Golembiewski, R.T. Organization analysis and praxis: Prominence of progress and stuckness. In *Review of Industrial and Organizational Psychology*; Cooper, C.L., Robertson, I., Eds.; John Wiley: London, 1986; 279–304.
3. Stokes, G.S.; Mumford, M.D.; Owens, W.A. *Biodata Handbook*; CPP Books: Palo Alto, CA, 1994.
4. Golembiewski, R.T. *Managing Diversity in Organizations*; University of Alabama Press: University, AL, 1995.
5. Golembiewski, R.T. Organizing public work, round three: toward a balance of political agenda and management perspectives. In *The Costs of Federalism*; Golembiewski, R.T., Wildavsky, A., Eds.; Transaction Publishers: New Brunswick, NJ, 1984; 237–267.
6. Golembiewski, R.T. *Practical Public Management*; Marcel Dekker: New York, 1995; 243–248.
7. Chandler, A.D., Jr. *Strategy and Structure*; MIT Press: Cambridge, MA, 1962.
8. Golembiewski, R.T.; Vigoda, E.; Sun, B.-C. Cacophonies in the contemporary chorus about change of public worksites. *Int. J. Public Adm.* **2001**, in press.
9. Kaufman, H.A. New york city health centers. In *State and Local Government: A Casebook*; Bock, E.A., Ed.; The Inter-University Case Program by University of Alabama Press: Birmingham, AL, 1963.
10. Simon, H. *Administrative Behavior*; Macmillan: New York, NY, 1947.
11. Golembiewski, R.T. *Civ. Serv. Manag. Work* **1962**, *56*, 961–973.
12. Golembiewski, R.T. *Men, Management, and Morality*; McGraw-Hill: New York, 1965.
13. Rainey, G.W., Jr.; Rainey, H.G. Structural overall in a government agency. *Public Adm. Q.* **1986**, *10*, 206–228.
14. Gouldner, A.W. *Patterns of Industrial Bureaucracy*; The Free Press: Glencoe, IL, 1954; 182–187.
15. Golembiewski, R.T.; Kiepper, A. *High Performance and Human Costs*; Praeger: New York, 1988; chaps. 4–5.
16. Golembiewski, R.T. Public sector organization: Why theory and practice should emphasize purpose, and how to do so. In *A Centennial History of the American Administrative State*; Free Press: New York, 1987; 433–474.
17. Mintzberg, H. *The Structuring of Organization*; Prentice Hall: Englewood Cliffs, NJ, 1979.

BIBLIOGRAPHY

1. Golembiewski, R.T. Trends in the development of the organization sciences. In *Handbook of Public Administration*; Rabin, J., Hildreth, J.B., Miller, G.J., Eds.; Marcel Dekker: New York, 1998.

Organizational Citizenship Behavior

Courtney E. Jensen
Larkin S. Dudley

Center for Public Administration and Policy, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, U.S.A.

INTRODUCTION

Organizational citizenship behavior (OCB) is a fairly recent concept that draws on a variety of disciplines including political science, psychology, economics, and sociology. This article defines OCB and describes concepts related to it; presents the organizational theories from which OCB is derived; explains the relationships between OCB and attitudes, personality, and leadership; and highlights OCB's role in organizational effectiveness.

DEFINITION AND RELATED CONCEPTS

The term OCB was coined by Dennis Organ and his colleagues in the early 1980s.^[1] It has come to be defined as “individual behavior that is discretionary, not directly or explicitly recognized by the formal rewards system, and in the aggregate promotes the efficient and effective functioning of the organization.”^[2] Actions that constitute OCB must not be required in the job definition and the rewards for such behaviors must not be contractually guaranteed. While there is no consensus as to which components define OCB, the five dimensions suggested by Organ that have received the most empirical research include:^[2,3]

1. *Altruism*: Targeting helping behavior toward a specific individual.
2. *Consciousness*: Adhering to the spirit in addition to the specifics of the rules and norms.
3. *Sportsmanship*: Not complaining about trivial matters.
4. *Courtesy*: Consulting with others, preventing problems from occurring, and mitigating them when they do.
5. *Civic virtue*: Being aware of and involved with issues that affect the organization.

The concept of OCB shares some overlap with prosocial organizational behavior, organizational spontaneity, and extra role behavior, but the term it is most closely associated with is contextual performance. Instead of concentrating on outputs or task

performance, contextual performance focuses on activities that support the social and psychological environment.^[4] The major difference between OCB and contextual performance is that the latter makes no reference to what is expected and concentrates on the social/psychological contexts that will enhance technical task performance rather than the environment that will bring about optimal organizational effectiveness.^[2]

ORGANIZATIONAL THEORY FOUNDATION

Organizational citizenship behavior can be traced back to the work of Barnard and Katz and Kahn.^[5,6] Barnard believed people's commitment or “willingness to cooperate” was essential and could not be acquired through a contractual obligation or obedience to authority. Katz and Kahn argued effective organizations have three obligations: 1) attract and retain employees; 2) ensure employees adequately function in their roles; and 3) encourage behavior that exceeds the minimal performance requirements. The third obligation creates an environment favorable for OCB as a sense of citizenship encourages a contribution beyond what is contractually required.^[2]

Other theories that have contributed to shaping OCB include the work of Roethlisberger and Dickson, who drew a distinction between the formal and the informal organization, the latter being a necessary condition for collaboration.^[7] The dual ideas of collaboration and informal organization comprise the essence of OCB.^[2] Another theory from which OCB draws is leader-member exchange theory that posits differences in subordinates abilities, causing leaders to treat them in one of two categories—hired hands and cadres. The relationship between the leader and the hired hands is contractual but the relationship with the cadres is fluid so the leader can negotiate through social exchange extra contributions that are likely to resemble what defines the behaviors of OCB.^[2] Somewhat unexpectedly, the theory of transaction cost economics that views the contract as the essence of the market exchange acknowledges behavior that fits with the concept of OCB. According to transaction theory, employees make three contributions. First, they satisfy

the agreed upon tasks of the contract and second, they contribute those actions not specified in the contract but that the employer may ask for in the future. Akin to OCB, the employee contributes spontaneous actions neither originally specified nor explicitly ordered.^[2] Finally, OCB shares some similarities with the concept of the clan in that the employees see their interests as similar to the interests of others and the organization as a whole. This leads to promoting each other's interests and a sense of community.^[2]

JOB SATISFACTION, ATTITUDES AND PERSONALITY

Organizational citizenship behavior researchers have hypothesized that job satisfaction, attitudes, and some personality factors should predict an individual's propensity toward OCB.^[2] When Barnard observed that the willingness to cooperate fluctuates, the natural question was what causes that fluctuation to occur? The concept of OCB materialized from research on one possible explanation—job satisfaction. Since OCB has the ability to take many forms and can be found in a myriad of situations, many of which are unpredictable, researchers often want to link it to job satisfaction.^[2] In trying to link job performance and job satisfaction, some researchers discovered that the connection between performance and satisfaction is weak. Others found that a relationship only emerges when satisfaction results from performance-based rewards. Propositions about job satisfaction have thus shifted to propose that improving the relationship between performance and rewards can strengthen the relationship between job performance and job satisfaction.

Attitudes and personality are good prognosticators for patterns and behaviors over time in ambiguous situations and may moderate any relationship between OCB and job satisfaction. Attitudes can come to the surface during varied situations that fail to offer compelling incentives for performance. Closely related to attitudes' moderating effect on satisfaction is the notion of trust or fairness. Actions that constitute OCB are part of a social exchange and in order for that exchange to continue to take place between the employee and the employer, the employee must perceive a sense of justice or fairness in the organization. The overlapping concepts of perceived fairness and job satisfaction are key components of the idea of morale, which in turn is a major motivational factor for OCB.^[2]

Personality's relation to OCB is best understood under the umbrella of the Big Five framework of personality capturing a wide array of personality traits in five broad categories, including agreeableness,

conscientiousness, emotional stability, extraversion, and openness to experience.^[8] The two categories that have important relevance for OCB are agreeableness, which relates to helping, courtesy, and sportsmanship; and conscientiousness, which relates to compliance and civic virtue. However, in research, personality has proven to be the weaker predictor compared to attitudes for a number of possible reasons. First, personality may have more of an indirect result by affecting attitudes that in turn effect job satisfaction. Second, it is possible personality influences the manner and motive of OCB more than its actual substance. Finally, there is the possibility that the workplace inhibits and restricts personality in some ways.^[2] Overall, there is not one dominant attitude or personality trait for predicting OCB, but in nearly every study, morale is consistently linked to an individual's OCB.^[2]

LEADERSHIP AND PERFORMANCE

One reason there is no equation to easily determine OCB is it may include the employee's desire to contribute along with his or her feeling of obligation to contribute. Leadership plays an important role in the employee's determination. Leaders can encourage OCB by clarifying what is expected, by determining how employees can achieve what is expected, and by expressing concern for the employees' personal well being.^[2] Treating employees fairly may motivate them to engage in OCB because the perception of fairness may create a sense of obligation, increase trust that contributions to the organization will be rewarded in the future, improve job satisfaction, and engender a greater regard for managers seen as kind and fair.

However, other factors may intervene negatively on the development of OCB, including organizational constraints, conditions that make it difficult for employees to perform their duties, thus reducing the ability and the motivation to perform OCB; an individualistic organizational culture, which is more likely to discourage OCB than a collectivist culture; and a widening power distance between managers and employees, which may affect the employees perception of fairness, a precondition for OCB.

From the point of view of leadership, managers tend to value OCB if they perceive it enhances organizational effectiveness. It is difficult to prove a direct correlation between OCB and organizational effectiveness but researchers note that the two may be related because OCB may affect: 1) the organizational, social, and psychological context for tasks and processes through enhancing coworker and managerial productivity; 2) the releasing of resources for more productive purposes by reducing the need for maintenance and the coordination of activities among team members

and groups; 3) the enhancement of the organization's ability to attract and retain top caliber employees; 4) the improvement of organizational stability, relationships with the environment, and acceptance of change; and 5) the creation of social capital.^[2,9]

Additionally, a body of research suggests that OCB is related to performance. However, the relationship between performance and certain forms of OCB, such as helping, appears to be stronger than the relationship with others, such as sportsmanship and civic virtue.^[2] Yet, OCB researchers have reasoned that due to the positive effects OCB may have for the manager and the organization and because managers may perceive OCB as a sign of commitment to the organization, managers are likely to repay employees who demonstrate higher levels of OCB with a higher rating on performance evaluations. OCB may not only affect the managers' evaluations of employees' performances, but their overall impressions of individuals. If the manager sees an employee exhibiting behavior he or she connects with a specific personality trait, the manager will likely consider that trait in the overall evaluation. Additionally, the more the employee conforms to a manager's vision of a good employee, the more likely positive behavior will be noticed and considered in evaluations.^[2] OCB, because it is not required, may appear especially impressive to the manager.^[2]

CONCLUSIONS

Despite the call for OCB to be utilized in the study of public administration, relatively little research has been conducted.^[10] There has been little cross-sector research on OCB but one study found higher levels of OCB in the public vs. the private sector.^[11] Studies undertaken in the public sector have shown that trust in supervisors mediated interactional justice and dimensions of OCB,^[12] job satisfaction is a predictor of altruism and compliance,^[13] contingent employees engage in OCB less than permanent employees,^[14] and organization-based self-esteem is positively related to OCB.^[15] Recent studies on OCB and the psychological contract framework have indicated perceived employer inducements were important predictors of OCB^[16] while an earlier study found perceived employer obligations explained variances in OCB better than perceived employer inducements.^[17] Despite the attention of a few scholars, the field of public administration needs to take a closer look at OCB, empirically and theoretically, to understand what insight OCB can bring to public administrators, especially in light of legal and political differences between work in the public and private sector.

The research exploring OCB in the workplace is still developing and has yet to reach undisputed

conclusions. Part of the challenge facing researchers is the multidimensional nature of OCB, which overlaps and creates confusion with a variety of similar concepts. In order for researchers to study OCB in a uniform manner, two significant hurdles need to be addressed. The first is researchers need to agree on a definitive definition of OCB that plainly distinguishes it from related concepts. The second difficulty is to draw a clear connection between OCB and organizational effectiveness. These obstacles are daunting, but if overcome, OCB could be an useful concept that would benefit research on both organizations and their citizens.

REFERENCES

1. Podsakoff, P.M.; MacKenzie, S.B.; Paine, J.B.; Bachrach, D.G. Organizational citizenship behaviors: a critical review of the theoretical and empirical literature and suggestions for future research. *J. Manage.* **2000**, *26* (3), 513–563.
2. Organ, D.W.; Podsakoff, P.M.; MacKenzie, S.B. *Organizational citizenship behavior: Its nature, antecedents, and consequences*; Sage Publications: Thousand Oaks, CA, 2006.
3. LePine, J.A.; Eriz, A.; Johnson, D.E. The nature and dimensionality of organizational citizenship behavior: a critical review and meta-analysis. *J. Appl. Psychol.* **2002**, *87* (1), 52–65.
4. Borman, W.C.; Motowidlo, S.J. Task performance and contextual performance: the meaning for personnel selection research. *Human Perform.* **1997**, *10*, 99–109.
5. Barnard, C. *The Functions of the Executive*; Harvard University Press: Cambridge, MA, 1938.
6. Katz, D.; Kahn, R.L. *The Social Psychology of Organizations*; Wiley: New York, 1966.
7. Roethlisberger, F.J.; Dickson, W.J. *Management and the Worker*; Harvard University Press: Cambridge, MA, 1939.
8. McCrae, R.R.; Costa, P.T., Jr. Validation of the five-factor model of personality across instruments and observers. *J. Personality Social Psychol.* **1987**, *52*, 81–90.
9. Borman, W.C.; Motowidlo, S.J. Expanding the criterion domain to include elements of contextual performance. In *Personnel Selections in Organizations*; Schmitt, N., Borman, W.C. and Associates, Eds.; Jossey-Bass: San Francisco, 1993; 71–98.
10. Vigoda, E.; Golembiewski, R.T. Citizenship behavior and the spirit of new managerialism: a theoretical framework and challenge for governance. *Am. Rev. Public Admin.* **2001**, *31* (3), 273–296.
11. Vigoda-Gadot, E.; Kapun, D. Perceptions of politics and perceived performance in public and private organizations: a test of one model across two sectors. *Policy Politics* **2005**, *33* (2), 251.
12. Aryee, S.; Budhwar, P.S.; Chen, Z.X. Trust as a mediator of the relationship between organizational

- justice and work outcomes: test of a social exchange model. *J. Organizational Behav.* **2002**, 23 (3), 267.
13. Tang, T.L.; Ibrahim, A.H.S. Antecedents of organizational citizenship behavior revisited: public personnel in the United States and in the Middle East. *Public Personnel Manage.* **1998**, 27 (4), 529.
 14. Coyle-Shapiro, J.A.M.; Kessler, I. Contingent and non-contingent working in local government: contrasting psychological contracts. *Public Admin.* **2002**, 80 (1), 77–101.
 15. Tang, T.L.; Gilbert, P.R. Organization-based self-esteem among mental health workers: a replication and extension. *Public Personnel Manage.* **1994**, 23 (1), 127–135.
 16. Coyle-Shapiro, J.A.M.; Conway, N. Exchange relationships: examining psychological contracts and perceived organizational support. *J. Appl. Psychol.* **2005**, 90 (4), 774.
 17. Coyle-Shapiro, J.A.M. A psychological contract perspective on organizational citizenship behavior. *J. Organizational Behav.* **2002**, 23 (8), 927.

BIBLIOGRAPHY

1. Bateman, T.S.; Organ, D.W. Job satisfaction and the good soldier: the relationship between affect and employee “citizenship.” *Academy Manage.* **1983**, 26 (4), 587–595.

Organizational Commitment

Jong-In Yoon

Department of Political Science, University of Georgia, Athens, Georgia, U.S.A.

INTRODUCTION

How are employees linked to the organization for which they work (or to which they belong)? Such a question has been with us for a long time in academia as well as in real life. Early management literature discussed the issue of employee loyalty. For example, to Barnard,^[1] loyalty is a willingness of persons to contribute efforts to the cooperative system of the organization, which he believed is an essential condition of organization. However, the controversies over the origin, nature, and creation of loyalty continued into and through the 1960s. By the early 1970s the concept of loyalty began to be discussed under a new description called organizational commitment (OC). Generally, the concept of OC contains some elements found in the concept of loyalty, but several new elements are also noticeable.

ABOUT ORGANIZATIONAL COMMITMENT

OC is generally believed to represent *an individual and psychological bond between an employee and an organization*, which includes loyalty to and identification with an organization. OC has been a popular topic during the past three decades, and a series of meta-analyses provide a summary of that literature (e.g., Refs.^[2–4]). So, what makes OC attractive to scholars and practitioners? Briefly, gaining a greater understanding of the processes related to OC has implications for employees, organizations, and society as a whole.

Basic hypotheses argue that employee commitment produces personal benefits such as non-work and career satisfaction. People need to be committed to something, and the opposite of commitment is alienation, which is obviously unhealthy. In this sense, commitment is different from workaholicism. In addition, from the perspective of an organization, commitment is viewed as a fairly stable and reliable predictor of employee behaviors such as turnover, when compared to other attitudinal constructs such as job satisfaction. Organizations also value commitment, because it is typically believed to reduce withdrawal behaviors. Moreover, committed employees are more likely to be involved in extra-role behaviors such as creativeness

and innovativeness, which are also assumed to enhance organizational competitiveness.

From a larger perspective, society as a whole also benefits from OC, because it may lower rates of costly job movement and may enhance national productivity or work quality. Overall, empirical research thus far reports modest inverse relationships between OC and turnover, intention to leave, and absenteeism.

Recently, some cynics argue that the notion of OC has faded away in the welter of work environments such as megamergers, layoffs, and contracting-outs. They also recommend that employees not become too involved in their organizations but instead look out for themselves (or their families) and prepare themselves for employment in other organizations in the event of a layoff.

However, the study of OC is neither useless nor outdated. Employees are still much committed to their organizations, in contrast to the expectations of some cynics. One survey asked IT workers to estimate the percentage of their coworkers who were “extremely committed.” Respondents reported that about 70% of their coworkers were “extremely committed” to their organization.^[5] In addition, although organizations change, they do not disappear. Rather, as organizations become leaner and smarter, those who remain in the organization become more important. Once the “fat” is gone, the remaining employees represent the “brain and muscle” of the organization.^[8]

This item addresses several aspects of OC. First, the meaning of OC, that is, the definitions and measurements of OC, is examined; second, the literature pertaining to the determinants and consequences of OC is reviewed; third, remaining issues and future research tasks are also discussed.

MEANING OF COMMITMENT

Definitions and Measurements

Although OC is one of the topics widely researched in organizational behavior, empirical literature with regard to OC is somewhat confusing, because there are many definitions and measurements of OC. For example, Mowday, Porter, and Steers^[6] reported 10 divergent definitions of OC, which reflect no consensus. This is partly due to the fact that researchers from

various disciplines (e.g., economics, psychology, sociology) have ascribed their own meanings to the construct of OC. However, these diverse definitions also reflect controversies over the nature of OC.

To illustrate, Fig. 1 shows one of the current typologies of OC, which challenges the unitary concept of OC. The distinction between a “behavioral” commitment and an “attitudinal” commitment is common and long-standing in the literature of OC, while the distinction between affective, continuance, and normative commitment is rather recent.

A behavioral approach posits that employees are committed to a specific course of action (e.g., maintaining employment with an organization). In this approach, an individual is committed through behavior. The more explicit, irrevocable, and public an individual’s prior activity has been, the more stable subsequent behavior will be. The theory of side-bets also argues that OC develops as an individual accumulates side-bets in an organization, which involves something valuable to an individual (e.g., pension, social connections, ease in doing work). The threat of losing these side-bets commits the individual to the organization.

On the other hand, an attitudinal approach focuses on an individual’s psychological attachment that links an individual to an organization. The attitudinal approach has been a mainstream of OC research. The most widely recognized definition of attitudinal commitment comes from Porter, Steers, Mowday, and Boulian,^[7] who defined commitment as “*the relative strength of an individual’s identification with and involvement in a particular organization.*” They also conceive OC as a complex concept that can be parceled into at least three factors: 1) a strong belief in and acceptance of the organization’s goals and values; 2) a willingness to exert considerable effort on behalf of the organization; and 3) a strong desire to maintain membership in the organization. In addition, the Organizational Commitment Questionnaire (or OCQ) developed by Porter and his colleagues has been widely used. In a meta-analysis, 59% of the 174 studies reviewed relied on the OCQ.^[2] The OCQ consists of 15 statements, and subsequent studies have reported that it has satisfactory test-retest and internal consistency reliability, and also has acceptable levels of predictive, convergent, and discriminant validity.

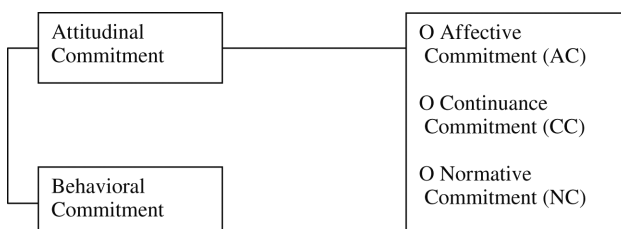


Fig. 1 A current typology of OC. (From Ref.^[11] p. 232.)

Although distinct, neither attitudinal nor behavioral approaches to OC can claim superiority. For example, a behavior may cause the development of congruent attitudes, which in turn lead to further the behaviors and vice versa. Thus, the development of OC may rely on the subtle interplay of the two types of commitment.

Distinction with Other Related Constructs

The relations of OC to major constructs also support its relevance. First, OC differs from the concept of job satisfaction, which is a positive emotional state resulting from the appraisal of one’s job. Generally, job satisfaction reflects one’s response to specific aspects of one’s job (e.g., pay, promotion) or to one’s job as a whole, whereas OC as a construct reflects one’s general affective response to the organization as a whole. OC reflects a more stable attitude over time than job satisfaction, which reflects one’s immediate affective reactions to job facets and to day-to-day transitory events. In sum, OC is more evaluative and macro than job satisfaction, and it develops more slowly over time.

Second, OC also differs from job involvement, which is the degree of absorption employees experience in their jobs. Obviously, one primary distinction between the two is that job involvement describes employees’ attachments to their jobs, whereas OC describes employees’ attachments to their organizations. However, the two constructs seem to be associated with each other to the extent that an organization provides an employee with what is desired.

Third, occupational commitment (or career commitment, career salience, professional commitment) refers to an employee’s commitment to occupation (or career, profession). Although the early literature stressed the inherent conflict between organizational and professional commitments, those two commitments may not necessarily be incompatible.

In fact, although OC is distinguished from such constructs as those mentioned above, the moderate or high correlations between OC and other constructs hint at the potential “concept redundancy” between them. Concept redundancy exists when conceptual definitions do not prove mutually exclusive, or when the link between a concept and its measurement is less than perfect. More research efforts should be paid to examine the relationships between diverse forms of work commitment.

DEVELOPMENT OF ORGANIZATIONAL COMMITMENT

The literature of OC provides “a long laundry list” of variables associated with OC, directly or indirectly, but

the majority of empirical studies investigating the determinants and consequences of OC are correlational. Thus, although this article uses such terms as “determinants” and “consequences,” few of OC empirical studies have been designed in a way that fully justifies such terms.

Literally hundreds of studies have examined the relationships between OC and its determinants, but one common theme underlying the development of OC is the notion of *exchange* between employees and their organizations. Individuals come to organizations with certain needs, desires, and skills. Organizations satisfying those needs and desires may have more committed employees. The notion of exchange is valuable not only from a conceptual standpoint in understanding the development of OC, but also from a societal standpoint in understanding that commitment is not simply a means of managerial exploitation of employees.^[6] The following discussions are drawn mainly from previous reviews.

Personal Characteristics

Overall, age and tenure have positive, albeit weak, relations with OC, but the interpretation of those results is not unequivocal. Consider whether older workers are more committed to their organizations, or experience limited job alternatives and greater sunk costs. Alternatively, older employees might actually have more positive work experiences than younger employees. Then, work experience rather than age is a true factor. When it comes to tenure, employees might become strongly attached to their organizations as tenure increases. Or it might be due to the fact that, over time, those who do not have strong OC choose to leave the organization, and only those who develop strong OC remain among the long-tenured group.

In contrast to age and tenure, the effects of gender and education on OC are less consistent. Roughly, women tend to be more committed to an organization. One explanation is that women have to overcome more barriers to attain their organizational positions. Empirical findings generally report an inverse relationship between education and OC, because highly educated employees may have higher expectations that the organization is often unable to meet and may have more alternative job opportunities than less-educated employees.

OC is also related to such dispositional variables as achievement motivation, sense of competence, and other high-order needs. Employees might be more committed to the organization that meets their high-order needs such as a need for achievement. In addition, the perceived competence has shown a large positive correlation with OC across diverse work

samples, indicating that employees confident in their abilities and achievements have higher OC than those who are less confident. An alternative explanation is that competent employees can choose higher-quality organizations that inspire strong OC.

In sum, even if unclear, the correlations obtained thus far imply that personal differences should be taken into account in any model of OC.^[6] At this point, it is difficult to confirm that OC derives directly from demographic and dispositional variables. The possibility remains that the relations between personal characteristics and OC are moderated by other variables, e.g., organizational or work experience variables.

Organizational Characteristics

Relatively recent attention has gone to the influence of organizational characteristics (e.g., organizational size, decentralization, and formalization) on OC. Researchers have found that employees experiencing greater decentralization and greater formality of written rules are more likely to feel committed to their organizations than employees experiencing these factors to a lesser extent. With regard to organizational size, some report that larger organizations tend to be less personal and harder to identify with, while others argue that larger organizations may increase the chances of promotions and other forms of intrinsic and extrinsic benefits. No one knows. Possibly, stronger relations between organizational characteristics and OC would be observed if they were examined by using an organizational level of analysis instead of (as is usual) an individual level of analysis.^[8]

Work Experiences

The vast majority of determinants studied thus far fall into the broad category of work experiences, which encompass such variables as job scope, role ambiguity, supportiveness, fairness, and leadership styles. Most variables have been measured by employees' perceptions rather than objective assessment of those characteristics. Generally, jobs that are perceived more general, complex, challenging, or enriching, have been found to lead to higher levels of OC. In addition, empirical studies have found that OC is lower among employees who are not sure about their roles (role ambiguity), or who feel that their roles are incompatible (role conflict). Similarly, role overload has been found to be inversely related to OC.

Empirical research also reports that OC is positively associated with participation in decision making, leader consideration, and organizational supportiveness and fairness. Significant positive correlations have also been reported between perceived fairness of specific

policies (e.g., pay and strategic decision making) and OC. In addition, it appears that the relations between leaders, coworkers, and employees have an impact on OC. Although less consistent, group cohesiveness has been found to be positively related to OC. Commitment also has shown positive relations to such variables as leader consideration, leader communication, and participatory leadership. Compared to personal and structural variables, in general, work experience variables have shown stronger and more consistent associations with OC.

CONSEQUENCES OF ORGANIZATIONAL COMMITMENT

As one priority, the literature of OC has focused on such variables as withdrawal behaviors (attendance, lateness, turnover) and withdrawal intentions (intention to search for job alternatives, intention to leave), and performance. Given that OC refers to a psychological bond between an employee and the organization, OC is anticipated to be negatively associated with withdrawal intentions and behaviors. Meta-analytic evidence illustrates that OC correlates positively with attendance and negatively with lateness and turnover, although these effects were relatively weak. In fact, however, OC demonstrates much larger correlations with withdrawal intentions—intention to search for job alternatives and intention to leave. This suggests that the effect of OC on turnover may be indirect through its relationship to (or mediated by) intention to leave or intention to search for another job.

Ultimately, perhaps, OC research highlights performance, but findings are weak and mixed. Why? Commitment may influence job performance, but so might factors like technology and resources. Thus, perhaps, we should not expect a strong commitment–performance linkage.

However, some optimism about the OC–performance linkage also deserves notice. Primarily, perhaps, the reason inheres in the poor validity of performance ratings, because supervisors do not take performance ratings seriously. Second, the impact of OC on performance may appear most clearly in the performance indicators that reflect employee motivation, while being relatively unconstrained by other factors (e.g., ability, access to resources), for example, employee initiative or organizational citizenship behaviors. In a related sense, for OC to influence particular performance outcomes, employees must have adequate control over the outcomes in question (e.g., cost controls). Third, most studies have examined the relations between OC and performance indicators at the individual level of analysis. Relevantly, studies adopting group or organization-level analyses have shown

stronger correlations between commitment and some performance variables. This suggests that OC may contribute to organization-level performance through ways other than individual-level performance. For example, some studies have found significant positive correlations between OC and extra-role behaviors, which go beyond and above one's job description.

CONCLUSION AND REMAINING ISSUES

The concept of OC has received intense attention during the past decades, and it will remain with us in the future. Gaining a greater understanding of how OC develops and evolves over time has significant implications for individuals and organizations, as long as organizations and employees exist. Even now, we know a great deal about the origin, nature, and consequences of OC. Clearly, much also remains unknown.

Definitions and Measurement of OC

Most obviously, continuous refinements of definitions and measurements are necessary. Recently, as noted, differentiating two or more dimensions of OC has received attention, and opinion converges on the view that *commitment is a multidimensional construct, not a unidimensional one* (see Fig. 1). They suggest that employees' dedication to an organization has a three-fold involvement: calculative (based on material rewards or side bets), affective (based on affiliation), and normative (based on obligation or value congruence).

As a further useful complication, a traditional view of commitment has come more frequently to be seen as an attachment to the organization as a whole. Emerging multifacets (or foci) of commitment involve simultaneously organization, division, coworkers, projects, or the public.

Although they need further empirical support, those two approaches—multidimensional and multifacet commitment approaches—have several advantages over traditional views of OC. Thus, multidimensional approaches open the possibility that we can understand more clearly the nature and origin of commitment. This is useful, in particular, for OC research in the public sector. For instance, the overall low OC reported among public employees (e.g., Ref.^[9]) may be largely due to low calculative commitment, regardless of high normative commitment. In particular, the concept of normative commitment may help explain any intrinsic motivation of public sector employees. This line of argument challenges the prevailing assumption in public sector HRM practices that public

employees are fundamentally self-interested and material rewards-oriented.

Relatedly, multifacet approaches fit well with current flexible and lateral organizational forms as well as with emphases on teams and groups within organizations. By investigating the foci of employee commitment, we might understand the development and consequences of OC more precisely and identify potential conflicts among multiple commitments within an organization. If anything, indeed, multifoci approaches appear to be more adequate for the study of OC in public organizations. The literature in public administration suggests that public managers have multiple masters, and that multiple constituencies affect public policies.

Model Construction and Testing

Most OC studies have been correlational and cross-sectional in design and thus have sharp limitations on causal inference, and insufficiently relate to psychological processes involved in the development and consequences of OC. Recently, causal models of commitment have received research attention, using OC as an endogenous or a mediating variable. OC has often been included as a mediator in such causal models, which link various personal characteristics and work experiences to job performance or turnover. A greater use of theory-based causal modeling techniques seems appropriate.

More longitudinal studies that investigate what factors are most important to employees at various career stages are also needed, given that OC seems a process that begins before an employee enters an organization and unfolds over time. In particular, longitudinal studies combined with multidimensional and multifacet commitment approaches may shed light on the change of the patterns (bases and foci) of commitments through an employee's career stages. In turn, this will guide organizational interventions or policies aimed at managing commitment patterns and, subsequently, their influences on employees' behavior.

Diversification of Research Methodologies

Common-method bias, a well-known problem in organizational behavior research, limits survey design in OC research. This bias occurs when measures of two or more variables are collected from the same respondents and methods, and the attempt is made to interpret any correlations among them. This bias can have extreme effects on the probability of finding significant results, even when there is no true relationship among variables. Another related concern is the social desirability bias, which occurs because questionnaire

items may prompt responses that present the person in a favorable light. Social desirability response may lead to bias in responses to questionnaires.

The OC researcher has defenses. Several research designs help resolve those problems—multitrait/multimethod approach, statistical control through partial correlation, and elimination of social desirability by using social desirability scales.

Negative Impacts of OC

So far, OC has been recognized as a positive factor for individual, organizations, and society. However, some observers raise the possibility of potential negative impacts of high levels of OC, among which are less creativity, resistance to change, stress, overconformity, and even a willingness to engage in corporate crimes. This possibility suggests that the relations between commitment and desirable outcomes adopt an inverted U curve, implying that over a certain point, high OC leads to be detrimental to individuals and organizations. Generally, the potential negative effects of OC get little attention.

Cross-Organizational and Cross-Cultural Studies

A large amount of most studies on OC involves samples from a single organization, and correlates employees' perceptions of organizational features and individual OC levels. Obviously, this generates problems, such as the restriction of variance and, subsequently, restriction of the true magnitudes of correlations between variables. Employee samples from a wide variety of organizations thus have a high priority in future research.

In a related vein, management theories and generalizations developed in one setting may not be automatically applicable to other cultures. Consider two conflicting views in the literature. A cultural determinist standpoint proposes that concepts, theories, and practices developed in one culture cannot be applied universally. In contrast, the theory of societal convergence argues that technology rather than culture plays a key role in determining worker attitudes, and thus, workers everywhere evolve similar patterns of work attitudes. This orientation suggests that work attitudes in other cultural settings may resemble those of the United States.

Research on OC outside of the United States is still at its early stage. Articles examining OC from a cross-cultural perspective have just recently begun to appear in academic journals. For example, through a search for the period from the late sixties to 1993, Randall^[10] found only 27 empirical studies investigating OC in

cross-cultural settings. To make matters worse, even most of those cross-cultural studies on OC are heavily tilted toward a few countries, such as Canada, Japan, and Israel. Obviously, more cross-cultural studies are needed.

REFERENCES

1. Barnard, C.I. *The Functions of the Executive*; Harvard University Press: Cambridge, MA, 1938.
2. Mathieu, J.E.; Zajac, D.M. A review and meta-analysis of the antecedents, correlates, and consequences of organizational commitment. *Psychol. Bull.* **1990**, *108* (2), 171–194.
3. Randall, D.A. The consequences of organizational commitment: methodological investigation. *J. Organ. Behav.* **1990**, *11*, 361–378.
4. Cohen, A.; Hudecek, H. Organizational commitment-turnover relationship across occupational groups. *Group Organ. Manage.* **1993**, *18* (2), 188–213.
5. Pittinsky, T.L. *Knowledge Nomads: Commitment at Work*; Harvard University, 2001. Unpublished doctoral dissertation paper.
6. Mowday, R.T.; Porter, L.W.; Steers, R.M. *Employee-Organization Linkages: The Psychology of Commitment, Absenteeism and Turnover*; Academic Press: New York, 1982.
7. Porter, L.W.; Steers, R.M.; Mowday, R.T.; Boulian, P.V. Organizational commitment, job satisfaction, and turnover among psychiatric technicians. *J. Appl. Psychol.* **1974**, *59*, 603–609.
8. Meyer, J.P.; Allen, N.J. *Commitment in the Workplace: Theory, Research and Application*; Sage Publication: Thousand Oaks, CA, 1997.
9. Buchanan, B. Government managers, business executives, and organizational commitment. *Public Adm. Rev.* **1974**, *35*, 339–347.
10. Randall, D.M. Cross-cultural research on organizational commitment: a review and application of hofstede's value survey module. *J. Bus. Res.* **1993**, *26*, 91–110.
11. Brown, R.B. Organizational commitment: clarifying the concept and simplifying the existing construct typology. *J. Vocat. Behav.* **1996**, *49*, 230–251.

BIBLIOGRAPHY

1. Becker, H. Notes on the concept of commitment. *Am. J. Sociol.* **1960**, *66*, 2–40.
2. Morrow, P.C. *The Theory and Measurement of Work Commitment*; JAI Press: Greenwich, CT, 1993.
3. Randall, D.A. Commitment and the organization: the organization man revisited. *Acad. Manage. Rev.* **1987**, *12* (3), 460–471.

Organizational Culture

Craig Lundberg

School of Hotel Administration, Cornell University, Ithaca, New York, U.S.A.

INTRODUCTION

The phenomena referred to as organizational culture, after early charges of mysticism and faddism, has become both popular and legitimate, and has a growing literature. Regardless of its evolving sophistication and contemporary stature as a field of study by scholars and as a source of inspiration for practitioners, organizational culture continues to be sidetracked by numerous, vigorously debated questions. Is culture something organizations are? Is organizational culture an independent or a dependent variable? For an organization, is there one culture, a set of subcultures, or both? Can organizational culture be known through objective assessment or only through interpretation or inference? Is each organizational culture unique, or does each have features in common that permit comparative analysis? Behind many, if not most, of the debated questions is a more fundamental, double-barreled one: is organizational culture amenable to change? If so, can organizational culture change be managed? To date, responses have been polarized. Some answer that cultures are changeable and of course they can be managed, whereas others respond that cultures have a life of their own that cannot be planned for or intervened with. The stance taken here is that such responses are overstated, glossing both the phenomena and our thinking about it.

The significance of organizational culture—that is, why it has drawn the attention of scholars and practitioners alike—is probably because it raises to consciousness features of organizations and the behavior of their members heretofore relatively ignored. Cultures after all are essentially about meanings, deep, symbolic, and otherwise, and meanings are always emotionally laden—two features largely taboo in organizational life. Cultures also serve as a major source of identity, provide nuanced behavioral guidance, reduce uncertainty, permit claims and justifications, and are of significance to people joined together in pursuit of common goals. Even the ways of understanding organizational cultures have utility. For example, we are counseled to look closely at the details of cultural specifics and to stand back and see cultures as wholes—to alternate careful description and cautious interpretation—ways of knowing that are both insightful and have built-in correctives. In a relatively random, often

ambiguous, and largely unknowable world, it can be argued that cultures are essential for collective activity and human health, development, and survival. In a world of change—large and small, swift and slow, but ubiquitous and continuous—the influence of organizational cultures on change, and the influence of change on cultures, merits close attention.

Regardless of the reasons for engaging in culture work, consequences are inevitable. As already mentioned, performing culture work (i.e., surfacing) can enhance socialization and team development. Also, as members become more aware of their organization's culture, this awareness contributes to the body of information on which everyday decisions and actions are based. Such cultural awareness often leads to subtle as well as intentional self-correcting behaviors—reinforcing patterned behaviors seemingly congruent with the culture and altering those that are not. Culture work can and usually does prepare members for deliberate organizational diagnosis and redesign of formal structures, systems, and processes (i.e., knowing what is culturally feasible). Thus, when formal changes are proposed (e.g., goal setting, and strategic/tactical planning), culture work assesses if it is consistent with the culture. Sometimes, probably rarely, culture surfacing and assessment may mobilize the concern and energy that leads to a new vision for a desired culture and efforts to craft it.

To argue, as is the intent of this article, that some aspects of cultures, under some circumstances, are amenable to intentional management change suggests that we go beyond the contemporary culture debates, as well as the piece-meal quality of the contemporary culture literature, to attempt to see more and see differently. Thus, this article outlines the assumptions and common understanding of the phenomena of culture, as well as culture's functions and significance, and offers a framework for cultural understanding.

WHAT IS ORGANIZATIONAL CULTURE?

The phenomena referred to as organizational culture, after early charges of mysticism and faddism, has become both popular and legitimate (e.g., Ref.^[1]). Although the construct of culture has been with us for a long time and has been enormously elaborated

(e.g., Ref.^[2]), the focus and the field known as organizational culture is of fairly recent origin.^[3] Although the phenomena were emphasized by a few early writers on organizations, “culture,” organizationally conceived, pushes us into territory left insufficiently explored by ideas that came before it. Group norms seemed sufficient to explain such phenomenon as why participants of off-site training reverted to their former behaviors and attitudes once they returned to their work setting. Organizational climate, although lending itself to measurement and association with many variables of importance, like norms, ultimately did not allow an explanation of the deeper casual aspects of how organizations function. By the 1960s, a growing emphasis on whole organizations began and concepts such as “system” were evoked to describe the patterns of norms, attitudes, and values organizationwide. Needed, moreover, were explanations for variations in climate, norms, and organizational behavior, and levels of stability in group and organizational behavior.^[4,5]

Although the phenomena of organizational culture have always existed, its popularity only began in the 1980s. This popularity undoubtedly began with Deal and Kennedy's^[6] *Corporate Cultures* and continued with Davis^[7] in *Managing Corporate Culture*. However, the literature has grown swiftly. A succession of more carefully conceived and empirically based efforts has provided elaboration of what organizational culture is, how it may be studied, its relationship to strategy, structure, and management practices, as well as several cautions (e.g., Refs.^[8–11]).

As with any growing field of study, however, the concept of organizational culture has many definitions. This is partly because culture is one of the most familiar things in our experience—we have grown up in a culture, intuitively know it and use it every day—and this excessive familiarity makes culture a difficult phenomenon on which to focus, much less to crisply conceptualize. Although definitions abound, almost all of them contain one or more central ideas such as values, shared meanings, assumptions, and beliefs. All definitions, however, subsume a distinctive set of assumptions about organizations and human behavior:

- Most of what happens in organizations is substantially more uncertain or ambiguous than in prevailing theories.^a
- Ambiguity and uncertainty undermine common, traditional approaches to analysis, problem solving, coordination, and control.

^aPrevailing theories assume that organizations are designed, purposive, more or less formalized systems consistent with a paradigm of intentionality where there is an unquestioned presumption of logic. (Ref.^[12].) and a compelling climate of rationality. (Ref.^[13].)

- People naturally create and perpetuate symbols of all types to combat chaos, provide direction, increase predictability, resolve confusion, and the like.

A distillation of the definitions and themes in the literature provides useful conceptual understanding of the phenomenon. There is a growing convergence that culture may be understood as a layered phenomenon, that it is composed of two or three interrelated levels of meanings from those relatively observable to those mostly invisible. Behind the many definitions, conceptual frameworks, and points of view about organizational culture, several widespread understandings of the phenomena can be discerned:

- It is a shared, common frame of reference (i.e., it is largely taken for granted and is shared by some significant portion of members).
- It is acquired and governs (i.e., it is socially learned and transmitted by members and provides them with rules for their organizational behavior).
- It endures over time (i.e., it can be found in any fairly stable social unit of any size, as long as it has some history).
- It is symbolic (i.e., it is manifested in observables such as language, behavior, and things to which are attributed meanings).
- It is at its core typically invisible and determinant (i.e., it is ultimately comprised of a configuration of deeply buried values and assumptions).
- It is modifiable, but not easily so.

Popular accounts of organizational culture emphasize the exotic and uncommon—heros, unusual labels, dramatic stories, peculiar ceremonies, and special events (e.g., the pink Cadillacs of Mary Kay Cosmetics, R and D skunk works, a chief executive officer “pitching in” on the production floor, the garage origins of a computer firm). Although such colorful anecdotes easily capture our attention, it is easy to miss their point; namely, that they are symbolic and their meaning is found in how they relate to a configuration of deeper elements.

If we look closely at the many definitions, common themes, conceptual schemes, and perspectives for discovering organizational culture, we can notice that there are essentially just two components: observable patterned behavior such as norms, language, customs, etc., and inferred cognitive-emotional ideas such as assumptions, beliefs, and other shared meanings. We further notice that these two components vary by level of generality or scope, from the specific to the philosophical, and are linked. The ideas serve as frames that cue the observed activity: “A cue in a frame is what makes sense, not the cue alone or the frame alone.”^[14] Meaning is therefore relational. Organizational culture

is thus understood as an organized set of ideas that draws attention to the phenomenon of observable patterned behavior (activities and events) and gives them meaning. It is the nature of the relationship between culture ideas and their phenomenon that is crucial.

The significance of organizational culture—that is, why it has drawn the attention of scholars and practitioners alike—is probably because it raises to consciousness features of organizations and the behavior of their members heretofore relatively ignored. Cultures after all are essentially about meanings—deep, symbolic, and otherwise—and meanings are always emotionally laden. These two features are largely taboo in organizational life.

What does organizational culture contribute to an organization and its members? Organizations share common issues and tasks—from survival in their external environments to managing their internal affairs, from handling crises to inculcating new members, from dealing with growth or decline to maintaining morale, from measuring performance to renewing product/service offerings. Organizational culture in one sense provides the organization’s solutions to these issues and tasks. What has been found to have consistently worked over time is symbolically expressed and maintained through patterned behaviors and devices as guides to future actions. Culture, therefore, is what is taught and reinforced, typically unconsciously, to members as the proper way to perceive, think, feel, and act in relation to the organization’s ongoing and recurring issues and fundamental tasks. As such, culture becomes both the source and the vehicle of organizational meanings.

Through its meaning-infused symbols and patterned activities, organizational culture provides much that is invaluable to organizations and members. It is, as mentioned, that which defines the uniqueness of the organization and serves as an organizational identity, or the meaning of who we are collectively. It also provides that which enables members to create their organizational identity. Organizational culture is also the basis for order, direction, and coherence—the meanings of where we are going and how we are related.

Culture, as central meanings, provides the rationale and supportive reasoning for an organization’s mission, strategy, and policy—in fact, for all required activities and responsibilities. It says, in effect, what we are jointly about and why we bother doing it. As cultural meanings are acquired by members they serve to mobilize, guide, and control human energy. Members learn what is generally expected of them as members—their understandings of what is appropriate and inappropriate behavior. Organizational culture as meaning also fills the void of uncertainty, too, and is therefore the primary source of anxiety reduction for members and others (e.g., the organization’s

stakeholders). Organizational culture is functional in that it provides meaningful response to the sorts of questions that pervade human existence: Who are we? What are we about? How should we behave?

LEVELS OF MEANING APPROACH

The investigation and understanding of organizational cultures is formed by a “levels of meaning” approach as outlined in Table 1. Let us briefly examine this framework. At the heart of any ongoing organization is a configuration of organization-specific *values and assumptions* (and associated ideology) that constitute its essential character. This core serves as the set of premises or precepts that strongly affect the vast majority of organizational thought and action. Values are the collective sense of what should be striven for or avoided, and the real ideals and standards and sins of the organization. Assumptions are the shared premises on which the dominant coalition bases its world views. Assumptions, therefore, refer to very basic things: the nature of human relationships, the nature of truth and time, the nature of human activity, and whether certain classes of people should be given preferential treatment. A central value and assumption always has to do with change, whether it is really possible or desirable, as well as who should initiate and manage it. This core level of meaning is essentially “taken for granted” and, to know it, we infer its components from other, more visible things.

A middle level of meaning comprises organization-specific *strategic beliefs*. Strategic beliefs refer not to the long-range plans or strategic pronouncements of organizational spokespersons, but rather to the fundamental “oughts” in the minds of the influential organizational leaders. There are four basic types of strategic beliefs.^[15]

- 1. Beliefs about strategic vision (i.e., what the organization can become and do, and what it will not attempt)

Table 1 “Levels of meaning” framework for understanding organizational culture

Manifest level
Symbolic artifacts
Language (jargon, sayings, slogans)
Stories (myths, sagas)
Ritualistic activities (rituals, ceremonies)
Patterned conduct (norms, conventions, customs)
Strategic level
Strategic beliefs
Core level
Ideologies
Values
Assumptions

Source: From Ref.^[20].

2. Beliefs about capital market expectations (i.e., convictions about what is necessary to keep lenders and investors satisfied)
3. Beliefs about product–market competition (i.e., understanding how and why the organization can succeed in its task environment and industry)
4. Beliefs about internal managing (i.e., the appropriate operating practices that support the other beliefs)

Strategic beliefs reflect, on the one hand, the cultural core and, on the other hand, actively condition the most surface or visible level.

Every organization or significant subpart that has some history will have developed a set of organization-specific *manifest elements*, which are relatively visible in that some careful, systematic listening and observing will reveal them. There are:

1. Symbolic artifacts (i.e., meaning-infused objects)
2. Unique language elements—a unique vocabulary (jargon), slang (argot), and sayings and slogans—and their non-verbal counterparts
3. Organizational stories about significant persons and events—some true and some elaborated, such as fictional myths and heroic sagas
4. Repeated social events—rituals that reaffirm and mark what is important, and ceremonies that signify key transitions or enable something deemed important to happen more easily
5. Patterned conduct—the norms, conventions, and customs that inform members about appropriate behavior and the preferred style of relating

Collectively, these manifest elements predispose members to experience organizational events, activities, and things in certain ways, to identify problems in similar ways, to perform within acceptable limits, and to strive toward common purposes. Experiencing the manifest culture, members over time come to acquire and share a unique common psychology familiarly understood as organizational “know-how.”

Four additional points require emphasis. The first point is the fallacy of believing that there is just one organizational culture. Although most organizations do have a more or less well-developed “umbrella” culture, it is also common that most organizational subsystems also have cultures (e.g., divisions or departments, major layers of the organization), as do the major occupations of the organization.^[16] Also, there is always the question of congruence among subsystem cultures, and between them and the umbrella culture.

The second point to emphasize builds on the first. Organizational and subsystem cultures may vary considerably along a continuum of intraculture clarity and consensus. Some cultures, regardless of size, have

meanings that are consistent and agreed upon (i.e., so-called “strong cultures”). Other cultures may hold differentiated and multiple meanings. Martin^[17] suggested the following cultural alternatives: “integration” (where the meanings of the dominant coalition are shared organizationwide), “differentiation” (where meanings overlap, collide, or coexist across subsystems), and “fragmentation” (where meanings are usually multiple and more ambiguous across all units of analysis from person to organization). The point here is not to assume either an intra- or intersystem consensus of meaning.

The third point to emphasize is that the examination of levels of meaning and the identification of meanings is quite difficult because such interpretive sense making always reflects the cultural allegiances of the investigators.

Last, organizational culture research tends to appear in one or the other of two major pairs of perspectives, the functionalist and the interpretive,^[18] and either the insider (emic) or outsider (etic) perspective^[19]—each important and useful.

The early debates about whether organizational culture can be managed and whether there is a single best perspective have largely been settled. Today, few would argue either that culture is usefully viewed as an independent variable (and hence is not manageable) or that culture is most usefully viewed as an objective phenomena (social constructivism now predominates). The continuing debates revolve around preferences for studying organizational culture (i.e., whether up close, such as ethnographically, or from afar, such as with questionnaire surveys) and preferences for describing cultural meanings (i.e., as contextually unique or in terms of cross cultural generalizations).

CONCLUSION

As we enter into the twenty-first century, organizational culture work needs to proceed on at least two fronts. The first front is conceptual. For example, organizational culture needs to be clearly differentiated from several closely associated constructs, such as climate and identity,^[17] elaborated in terms of previously underdiscussed components, such as emotion and time, and clarified in terms of encompassing cultural units, such as national and industry cultures, and cultural boundaries. The second front is empirical. Organizational culture needs to be studied longitudinally for antecedents and consequences in emerging contexts, such as multinational organizations spatially dispersed and virtual organizations, and whether it leads or follows organizational strategy and structure, and organizational transformation. Organizational culture, however, is the nexus of a symbolic meaning-centered paradigm that usefully

contributes to our thinking about the governance of postbureaucratic organizations.

REFERENCES

1. Cooper, C.L., Cartwright, S., Earley, P.C., Eds.; *The International Handbook of Organizational Culture and Climate*; John Wiley and Sons: Chichester, England, 2001.
2. Kroeber, A.L.; Kluckhohn, C. *Culture: A Critical Review of Concepts and Definitions*; Random House: New York, 1963.
3. Schein, E.H. Organizational culture. *Am. Psychol.* **1990**, *45*, 109–119.
4. Hofstede, G. *Culture's Consequences*; Sage: Beverly Hills, CA, 1980.
5. Schein, E.H. *Organizational Culture and Leadership*; Jossey-Bass: San Francisco, 1985.
6. Deal, T.; Kennedy, A. *Corporate Cultures: The Rites and Rituals of Corporate Life*; Addison-Wesley: Reading, MA, 1982.
7. Davis, S. *Managing Corporate Culture*; Ballinger: Cambridge, MA, 1984.
8. Frost, P.J., Moore, L.F., Louis, M.R., Lundberg, C.C., Martin, J., Eds.; *Organizational Culture*; Sage: Beverly Hills, CA, 1985.
9. Frost, P.J., Moore, L.F., Louis, M.R., Lundberg, C.C., Martin, J., Eds.; *Reframing Culture*; Sage: Thousand Oaks, CA, 1991.
10. Kilmann, R.H., Saxton, M.J., Serpa, R., Eds.; *Gaining Control of the Corporate Culture*; Jossey-Bass: San Francisco, 1985.
11. Martin, J. *Cultures in Organizations: Three Perspectives*; Oxford University Press: New York, 1992.
12. Weick, K. *The Presumption of Logic in Executive Thought and Action. Prepared for the Symposium, Functioning of the Executive Mind*; Case Western Reserve University, 1982.
13. Staw, B.M. Rationality and justification in organization life. In *Research in Organizational Behavior*; Staw, B.M., Cummings, L.L., Eds.; JAI Press: Greenwich, CT, 1980.
14. Weick, K. *Sensemaking in Organizations*; Sage: Thousand Oaks, CA, 1995; 110 pp.
15. Lorsch, J.W. Strategic myopia: culture as an invisible barrier to change. In *Gaining Control of the Corporate Culture*; Kilmann, R.H., Saxton, M.J., Serpa, R., Eds.; Jossey-Bass: San Francisco, 1985.
16. Trice, H.M. *Occupational Subcultures in the Workplace*; ILR Press: Ithaca, NY, 1993.
17. Martin, J. *Organizational Culture: Mapping the Terrain*; Sage: Thousand Oaks, CA, 2002.
18. Smircich, L. Concepts of culture and organizational analysis. *Adm. Sci. Q.* **1983**, *28*, 339–358.
19. Louis, M.R. An investigator's guide to workplace culture. In *Organizational Culture*; Frost, P.J., Moore, L.F., Louis, M.R., Lundberg, C.C., Martin, J., Eds.; Sage: Beverly Hills, CA, 1985.
20. Lundberg, C.C. Surfacing organizational culture. *J. Manag. Psychol.* **1990**, *5*, 19–26.

Papers on the Science of Administration

Mark L. McConkie

*Graduate School of Public Affairs, University of Colorado at Colorado Springs,
Colorado Springs, Colorado, U.S.A.*

R. Wayne Boss

*Leeds School of Business, University of Colorado at Boulder,
Boulder, Colorado, U.S.A.*

INTRODUCTION

Recognizing that “the President needs help,” Franklin D. Roosevelt appointed the Committee on Administrative Management, charged to make recommendations for improving the efficiency of the Presidents’ office. The Committee was popularly known as The Brownlow Committee, named after its chairman, Professor Louis Brownlow, who had been prominent in developing city management as a profession. The two other members of the Committee were Charles Miriam of the University of Chicago, and Luther Gulick of Columbia University and the Institute of Public Administration in New York City.

They subsequently shared their basic principles with the reading public. Gulick joined with Colonel Lindal Urwick and published in 1937, as a report to the Committee, *The Papers on the Science of Administration*.

IMPORTANCE OF THE “PAPERS”

The *Papers on the Science of Administration* is a collection of 11 essays or “papers” on how to more artfully manage modern organizations. They set forth the thesis that certain “principles” govern in effective management, and that by defining and living in harmony with those principles, managerial effectiveness and organizational productivity are increased. These principles are found in public and private settings. In explaining this phenomenon, Urwick wrote: “there are principles which can be arrived at inductively from the study of human experience of organization, which should govern arrangements for human association of any kind. These principles can be studied as a technical question, irrespective of the purpose of the enterprise, the personnel composing it, or any constitutional, political, or social theory underlying its creation.”^[1] These papers and this article are important, because they defined a theory of management that became

the focal point in the debate over the shape of public administration for years to come.

The papers were written in an era when Frederick W. Taylor’s *Principles of Scientific Management*—which argued that by applying scientific principles to management, the “one best way” of doing things might be discovered—had captured the imagination and practice of many managers in the public and private sectors.^[2] The *Papers on the Science of Administration* marked what many have called the “high noon of orthodoxy” in the practice of public administration, and it was recognized as the clearest exposition of its principles. Gulick asked the question, “What is the work of the chief executive?” and responded by asserting “The answer is POSDCORB,” a pithy anagram which summarized the executive’s role of Planning, Organizing, Staffing, Directing, Coordinating, Reporting, and Budgeting.^[2] At the time of their publication, the nation was pulling out of the great depression and was soon to be engulfed in World War II. POSDCORB gave government managers a system to approach the planning, organizing, staffing, and so on central to managing these great crises, which was done so well that Newland identified this period as “the golden age” of American public administration.^[3]

POSDCORB was also important because it began to pull together the intellectual fabric for the study of public administration and provide a common view point for the academics “who founded such key institutions as the Public Administration Clearing House (1930), the American Society for Public Administration (1939), and the Public Administration Review (1940), which significantly advanced research and teaching agendas for the field. POSDCORB further provided a cohesive and comprehensive academic rationale for setting up graduate programs advancing research as well as its applications” at federal, state, and local levels of government.^[4] Thus, it was that the “principles of administration” played a signal role in the growth and development of the field of public administration.

THE DEBATE OVER PRINCIPLES

Perhaps the most important impact of POSDCORB and the *Papers on the Science of Administration* was in the debate they triggered, which further defined the field of public administration. Following World War II, Simon, Dahl, and Waldo, in particular, argued that the principles identified by Gulick and Urwick were not principles at all, but merely “proverbs,” buffeted by all the subjective nuances of perception, values, and bias inherent in the human condition.^[5–7] Simon, in particular, argued that for every “principle” of administration an opposing “principle” existed, some of which were mutually contradictory, and that therefore, the idea of “principles” was moot.^[8] Still, the *Papers on the Science of Administration* are important for the impact they have had in shaping the field of public administration, in focusing attention on the role of upper-level management in government and in the insights proffered by individual authors of the articles.

REFERENCES

1. Urwick, L. Organization as a technical problem. In *Papers on the Science of Administration*; Gulick, L., Urwick, L., Eds.; Augustus M. Kelley, Publishers: Clifton, NJ, 1937; 49 pp.
2. Gulick, L. Notes on the theory of organization. In *Papers on the Science of Administration*; Gulick, L., Urwick, L., Eds.; Augustus M. Kelley, Publishers: Clifton, NJ, 1937; 13 pp.
3. Newland, C.A. *Public Administration and Community: Realism in the Practice of Ideals*; Public Administration Service: McLean, VA, 1984.
4. Stillman, R.J. *Public Administration: Concepts and Cases*; Houghlin Mifflin: Boston, 2000; 21 pp.
5. Simon, H. The proverbs of administration. *Public Admin. Rev.* **1946**, 6 (1), 53–67.
6. Dahl, R. The science of public administration: three problems. *Public Admin. Rev.* **1947**, 7 (1), 1–11.
7. Waldo, D. *The Administrative State*; Ronald Press: New York, 1948.
8. Simon, H. *Administrative Behavior*; The MacMillan Co.: New York, 1947.

Parental Rights and Obligations

Michael W. Austin

*Department of Philosophy and Religion, Eastern Kentucky University,
Richmond, Kentucky, U.S.A.*

INTRODUCTION

In the past 30 years, advances in reproductive technology and social change have led to increased reflection on the nature of parenthood. Because of these advances, many people who in the past would not have been able to gain access to the joys and challenges of parenthood can now gain such joys. These technologies and their implications also raise important moral and legal questions about the nature of the parent–child relationship. How should we think about being a parent? If parents have rights, how extensive are they? What obligations do parents have to their children? Should we require prospective parents to be licensed before we allow them to procreate and raise children? These are some of the theoretically and practically important questions related to the parent–child relationship.

PARENTHOOD: ABSOLUTISM AND LIBERATIONISM

Historically, some have proposed an absolutist view of the rights and obligations of parents.^[1] Absolutists hold that parents have complete and final authority over their children and over decisions that have an impact on them. Some absolutists have even argued that parents hold the power of life and death over their children. An absolutist view of parenthood often arises from the belief that children are somehow the property of their parents, and as property-owners have control of their property, parents, as such, have control over their children. One reason that absolutists offer for believing that parents have this level of authority is that such authority discourages vice in children and encourages them to develop respect for the laws of the state. Children are to learn obedience in the family, so that they can learn obedience in society, and be subject to the laws of the commonwealth. While absolutism is no longer a widely-held view among scholars, it is still the case that many people conceive of themselves as having near absolute control over their children.

Many justifications have been offered for the paternalistic treatment of children by their parents and society. For example, some hold that children are the property of their parents, as noted earlier. Others maintain that parents have authority over their children simply because they created them. Lastly, some argue that it is because children are incapable of taking care of themselves that parents and society can treat children paternalistically.

The children's liberationist rejects all of these arguments, and liberationism contrasts sharply with absolutism. Some liberationists argue that children should be able to do, without permission, whatever adults are allowed to do.^[2] Supporters of liberationism argue that children should have the same moral and legal status as adults, and so should be granted the same rights and freedoms possessed by any other member of society, such as the right to vote, the right to self-determination, the right to design their own education, the right to forgo education altogether, the right to sexual freedom, the right to alternative home environments, and the right to select guardians other than their parents.

The extreme absolutist holds that children essentially have no rights, whereas the extreme liberationist holds that children have all the rights that adults possess. There is much logical space between these two positions, which allows for both parents and children to be possessors of rights, *qua* parents and children. It is the case, however, that the greater the extent of the rights that children possess, there will be a corresponding reduction in the extent of the rights that parents possess, *qua* parents.

PARENTS AND CHILDREN: SOME RELIGIOUS APPROACHES

While it must be emphasized that it can be unclear what a particular religion teaches about the rights and obligations of parents, given the diversity of viewpoints within different religious traditions, it is the case that adherents to many of the major world religions address many of the moral and legal issues relevant to the parent–child relationship. For example, the

halakhah, i.e., the Jewish law, regulates the parent–child relationship via a system of mutual obligations that devolve on parents and children.^[3] The obligation to provide support for children falls solely on the father under the halakhah. Fathers have authority over their households, but are also obligated to provide for a child's daily needs, to educate the child in the ways of Judaism, and to teach male children a trade, among other duties. Also, fathers do not have the right to forfeit their parental obligations, whereas mothers may do so, in which case the children become wards of the community. In situations of divorce, parents are obligated to retain custody of their children. This obligation is grounded in the child's welfare, which generates a right of the child to be cared for by his or her parents. This emphasis on the moral and legal claims of children that arise out of their vulnerability is a widely-held view in both religious and nonreligious conceptions of the rights and obligations of parents.

Christian teaching with respect to the parent–child relationship is also diverse. Jean Bodin's absolutist view was grounded in his understanding of divine law—he believed that God granted absolute authority to parents over their children. Of course, many Christian thinkers agree that parental authority is derived from or modeled upon the authority that God has over humanity, but reject absolutist views of the extent of parental authority. Both Saint Augustine and Saint Thomas Aquinas wrote on the family, basing their views on Christian considerations, and Aquinas is quite clear in his rejection of absolutism.^[4] Augustine holds that parents have responsibilities concerning their children. Parents should tend to the welfare of their children, but not seek to satisfy all of the wishes of their offspring. Children need to learn discipline from their parents, and to learn religious faith from them. Aquinas sees parents as superiors, in the sense that they are to supervise and instruct their children. For Aquinas, parents do have distinct parental rights, but these rights are not absolute, because children also possess rights owing to the fact that they are individual human beings who as such are objects of justice. The rights of parents are constrained by the obligations of parents to educate their children in the physical, mental, moral, and religious realms, and to provide for their daily needs.

In many non-Muslim cultures, the common understanding of a Muslim family ethic includes the absolute (or nearly so) authority of men, as husbands and fathers, over their households. While there is of course oppression of women and children in the name of Islam, this is true of other faiths as well. Some Islamic teachers emphasize the complementarity of men and women.^[5] By nature, it is believed, men

are physically stronger than women, and women are superior to men in their capacities for love and affection. Given this, the duty of safeguarding the family falls upon the man, whereas the duties of childrearing fall upon the woman. If families live according to this natural order, then all will experience satisfaction in family life. Of course, there are Muslims who disagree with the complementarity view, as well as others who would emphasize egalitarian considerations when describing the moral aspects of Islamic family life.^[6]

PARENTHOOD AND PUBLIC POLICY

The ubiquitous nature of cases of child abuse and neglect raises important questions for public policy as it relates to the family. Consider the case of Tanner Dowler. In February of 2004, Joseph Dowler was sentenced in Boulder Colorado District Court to 60 years in prison for child abuse leading to the death of his infant son, Tanner.^[7] In his nine weeks of life, Tanner Dowler suffered horrific injuries. Tanner endured 11 broken ribs, fractures in every limb, and oxygen deprivation that led to severe brain damage. He also suffered second-degree burns on the soles of his feet. At the sentencing hearing, Boulder District Attorney Mary Keenan read a statement from Tanner's perspective, saying that "I will never get my first tooth, sit up alone, crawl, or take my first step. I will never get to learn that neglect and rejection and fear and pain are not the way other people live their lives."

What can be done to prevent such abuses? One answer that has been proposed by Hugh LaFollette^[8] is that the state should legally require people to obtain a parenting license before they are allowed to raise a child. This sort of program would be aimed at preventing people like Joseph Dowler from having custody of a child, and so it would seek to prevent the suffering and death of children that occurs because of their parents. Children are among the most vulnerable among us, and so their vulnerability could demand protection from harm via the legal system when parents and other caretakers fail to fulfill their responsibilities. LaFollette's claim is that we should require parental licenses to prevent such failures from occurring in the first place.

LaFollette argues that if an activity is potentially harmful to others and requires a certain level of competence, which can be demonstrated via a moderately reliable test, then it should be regulated. For example, we require that physicians obtain medical licenses from the state to ensure their competency owing to the potential harm to citizens from medical malpractice. To drive an automobile, a level of

competence must be demonstrated because of the potential harm to others that can be done by incompetent drivers. Crucial to the argument in favor of licensing parents is that parenting shares these features with other activities for which we require licenses: It is potentially quite harmful and requires a certain level of demonstrable competence.

Parents can harm their children through abuse, neglect, and lack of love or respect. The results are physical and psychological damage, leading to adults who are neither well adjusted nor happy, and this often leads to cyclical patterns of abuse. Additionally, many people lack the competency required to parent children in an adequate way, owing to factors such as an unsuitable temperament, a lack of knowledge, a lack of energy, or instability. If the state requires that parents be licensed, then those who lack the competency required to be a parent and are therefore likely to harm their children in the ways described earlier would be prevented from doing so. Those who fail to obtain a parental license and yet bear a child would face having that child removed from home and given to qualified adoptive parents.

If we deny that parental licensing is justified, then, by parity of reasoning, we have no grounds to require licensing for physicians or automobile drivers, given that we require licenses for those who engage in these activities because they are potentially harmful and require a level of demonstrable competence. Moreover, we require that adoptive parents demonstrate a level of competence before being allowed to adopt a child. LaFollette concludes that a state program for licensing parents is desirable, justifiable, and feasible. This proposal is of course very controversial, insofar as the restrictions it places on those who may raise children are more stringent than those currently in operation.

Parents: Owners or Stewards?

How should parents conceive of themselves? Are they owners of their children, or stewards of them? To contemporary ears, applying the term “ownership” to the parent–child relationship can be morally problematic. William Irvine^[9] has developed a stewardship conception of parenthood that sharply contrasts with the ownership model. Irvine’s view is that parents should see themselves as trustees, or stewards, of their children’s lives.

To clarify the concept of stewardship, imagine that your neighbor is on a trip, and asks you to take care of his lawn while he is away. During the time he is away, you are a steward of your neighbor’s lawn. You mow it, water it, and seek to care for it as needed. Similarly, says Irvine, parents should see

themselves as stewards of their children. A parent can see himself/herself as caring for the body and mind of his/her child and acts as if the adult human person whom the child will become is away on a trip, a trip she will return from in approximately 20 years. At that time, the parent’s stewardship is finally and fully transferred over to the child, and the child who is now able to govern his or her own life is free to exercise control over that life. The steward-parent focuses his or her parenting on the child and the needs of the child, and is willing to sacrifice for the good of the child.

One could agree with this, but still hold that parents have moral and legal claims as well, given the goods that parenting makes possible, such as love, intimacy, and meaning in life. These goods might only be pursuable in the private sphere. So, the right to privacy also would cover family life, giving parents the right to rear their children based not only on the welfare of those children but also on the possibility for the satisfaction of desires for love, intimacy, and meaning on the part of the parents.

CONCLUSIONS

The parent–child relationship is one of the most significant relationships that human beings experience, as it can provide a deep sense of well being, intimacy, and love in the lives of both parents and children. Moreover, the impact of the quality of family life on civic life is great. Cycles of crime and violence are often perpetuated via families. So, parents, children, and the state have interests in play with regard to family life. While no consensus has been reached on the existence and extent of the rights of parents, or on the nature of parental obligations, these are crucial issues that merit further attention.

REFERENCES

1. Bodin, J. *Six Books of the Commonwealth*; Barnes and Noble: New York, 1967.
2. Farson, R. Birthrights. In *The Children’s Rights Movement*; Gross, B., Gross, R., Eds.; Anchor Press: Garden City, NY, 1977; 325–328, Harris, J. Liberating children. In *The Liberation Debate: Rights at Issue*; Leahy, M., Cohn-Sherbok, D., Eds.; Routledge: New York, 1996; 135–146.
3. Wigoder, G., Ed. *The New Encyclopedia of Judaism*; New York University Press: New York, 2002; 588–589, 270–272.
4. Blustein, J.B. *Parents and Children: The Ethics of the Family*; Oxford University Press: New York, 1982; 53–62.

5. Rippin, A. *Muslims: Their Religious Beliefs and Practices. Volume 2: The Contemporary Period*; Routledge: New York, 1993; 120–123.
6. Rahman, F. *Islam*, 2nd Ed.; University of Chicago Press: Chicago, 1979; 212–234.
7. http://www.dailycamera.com/bdc/lafayette_news/article/0,1713,BDC_2425_2643088,00.html (accessed May 18 2004).
8. LaFollette, H. Licensing parents. In *Morality and Moral Controversies*; Arthur, J., Ed.; Prentice Hall: Upper Saddle River, NJ, 2002; 637–645.
9. Irvine, W. *Doing Right by Children*; Paragon House: St. Paul, MN, 2001.

BIBLIOGRAPHY

1. Archard, D. *Children, Family, and the State*; Ashgate: Burlington, VT, 2003.
2. Ladd, R.E., Ed. *Children's Rights Re-Visioned*; Wadsworth: Belmont, CA, 1996.
3. Narayan, U.; Bartkowiak, J.J., Eds. *Having and Raising Children*; The Pennsylvania State University Press: University Park, PN, 1999.
4. Purdy, L.M. *Their Best Interest? The Case Against Equal Rights for Children*; Cornell University Press: Ithaca, NY, 1992.

Participatory Policy Analysis

Virginia Eubanks

*Department of Women's Studies, University at Albany, State University of New York,
Albany, New York, U.S.A.*

INTRODUCTION

Participatory policy analysis (PPA) is a set of techniques, gaining popularity in the professional literature and practice, for expanding the range of actors involved in public decision making through deliberative dialogue. PPA arose from two sources: the new social movements of the 1960s and 1970s and recent critiques of “positivist” policy making. Scholars and advocates of more participatory approaches argue that PPA can help governmental bodies make better decisions, resolve intractable issues, empower and engage citizens, and promote justice and equity. Nevertheless, PPA has been subject to intense scrutiny and criticism, particularly in the international development community. Postpositivist critiques of policy making have not always resulted in convincing postpositivist methods or analytical instruments, and concerns about the wide uptake of PPA have taken several forms. These include lack of attention to social structure, implication in devolution and neoliberalism, and subjugation disguised as empowerment. This article discusses three examples of PPA in action. Examples of PPA are now fairly common in the professional literature, so these examples are neither exhaustive nor representative but do serve as an indicator of the breadth of current practice.

POSTPOSITIVISM AND POLICY ANALYSIS

In recent years, a “postpositivist” or “postempiricist” critique of traditional policy analysis has gained currency among policy makers, scholars, and activists.^[1–7] This point of view argues that positivist approaches to decision making—based on highly specialized, quantitative, instrumental, and “rational” processes derived largely from economics—have historically dominated the professional practice of the policy sciences at the expense of more democratic considerations.^[8] Positivist approaches to policy analysis, critics argue, tend to privilege elite forms of knowledge, discount stakeholders most affected by policy outcomes, perpetuate a false fact/value dichotomy, and marginalize policy analysis by narrowly conscribing the stages in the policy process at which analysis is

deemed appropriate. As such, technocratic policy analysis (and the instrumental rationality that underpins it) “destroys the more congenial, spontaneous, egalitarian and intrinsically meaningful aspects of human association...is antidemocratic...represses individuals...is ineffective when confronted with complex social problems...makes effective and appropriate policy analysis impossible...[and is based in] inappropriate and unfruitful social science instruments and methods.”^[3] These instrumental processes result, as critics note, in elite-oriented policy analysis, systemic and power-laden distortions of communication between analysts and their clients, and declining democratic participation. At best, the failures of positivist policy analysis are unintentional—the result of institutionalization of “hard” scientific norms and lack of attention to the social, political, and cultural aspects of decision making. At worst, false objectivity is used to compel the precarious consent of the governed—a kind of “policy public relations” intended to reproduce the *status quo* and maintain power of elites under the guise of efficiency.

Discussion of PPA has arisen as part of this more general critique, but the practice can be linked to participatory democratic techniques that originated in the new social movements of the 1960s and 1970s.^a In the most general terms, PPA is an approach to policy making and evaluation that “expands the range of actors...involved in the making and execution of public policy in a discursive or deliberative mode.”^[9] Alternatively, Igor Mayer defines PPA more concretely as:

[A] practical discipline which contributes to policy-making by designing policy-analytic fora, establishing conditions for participation, and facilitation and support of the policy-relevant debate and argumentation between the participants which take place within the forum... The analyst in PPA is an organizer of participatory fora and a facilitator of the interactions that take place within this forum.^[6]

^aThese techniques include, among others, consensus conferences, participatory action research, participatory analysis, participatory environmental planning, participatory rural appraisal, scenario workshops, and science shops.

In this model of PPA, professional policy analysts serve a quite new (and potentially uncomfortable) role as animators or facilitators of ordinary citizens' local knowledge, seeking more diverse forms of expertise to inform, and even *transform*, the political process through citizen engagement.

THE BENEFITS OF INCREASED CITIZEN PARTICIPATION IN POLICY MAKING

deLeon and Longobardi call PPA "policy analysis for the good society," claiming that the case for integrating PPA into common practice is fairly straightforward. "Citizens deserve a greater say in governance," they write, "based upon the knowledge that they are best able to articulate their special 'needs.' Given this voice as a basis for action, government can be more informed and responsive, thus promoting a more involved, engaged, and, in many perceptions, a better citizenry."^[10] Scholars and advocates of more participatory approaches argue that PPA can help governmental bodies make better decisions, resolve intractable issues, empower and engage citizens, and promote justice and equity.

Making Better Decisions

Proponents of PPA claim that greater citizen participation in policy making increases the complexity of analysis, integrates varied bodies of knowledge and experience, and enhances the relevance and timeliness of decision making. Even at the level of consultation, increased citizen participation may help policy makers make better, or at least better-informed, decisions.^[4,9] This evidentiary impact is a primary goal of PPA, particularly when a stakeholder model—one that emphasizes the importance of contributions of communities directly affected by the decision in question—is followed.^[11–13] Scholars have noted that the failures of positivist policies often result from policymaker's inability to integrate social norms, values, and the realities of living in political communities into their decision-making processes (see especially Ref.^[7]). A key benefit of PPA, therefore, is developing "a fair and impartial representation of all citizens' values and preferences"^[12] (quoted in Ref.^[14]).

Resolving Intractable Issues

Particularly in the domain of environmental issues, scholars and practitioners of PPA have argued that collaborative, participatory decision-making structures

are more likely to resolve intractable issues than decision-making structures centered on conflict and win/lose mentalities.^[15] Although they acknowledge that the quality of the deliberative process, communication between agencies, commitment to the process, and shared jurisdiction all have major impacts on the outcomes of participatory processes, Beierle and Konisky show considerable evidence that PPA does bring different parties involved in environmental planning closer to resolution, by developing common issue frames.^[16] The process of articulating common interests is important in overcoming NIMBY (not in my back yard) problems, which are often characterized by myopia and entrenched self-interest. Similarly, Broome has shown that participatory planning and design can help build consensus and develop a sense of community in protracted conflict situations, such as citizen peace-building work in Cyprus.^[17]

Empowering and Engaging Citizens

There are important benefits to PPA aside from the somewhat instrumental goals of making better decisions and resolving intractable issues, however. There is considerable evidence that participatory program design increases citizen participation and faith in governance. Recent scholars of "policy feedback" argue that as clients and recipients participate in government programs, they learn lessons about how citizens and governments relate, and these lessons have political consequences beyond the domain of the immediate agencies and decisions.^[18,19] If one of the normative goals of policy making is to create positive attitudes about government, increase citizen interest in public affairs, and facilitate participation in public life, PPA may provide an important strategy to enrich democratic practice.

Promoting Justice and Equity

At its most sweeping level of impact, PPA incorporates an argument against *containment* of people's power to make their own decisions, whether by professional experts or well-intentioned "representatives" of their interests. This argument rests on the assumption that oppression and marginalization are maintained by disparities of wealth and power and must be challenged on two fronts, by 1) directly confronting the structural maldistribution of economic and social resources, knowledge, power, and authority and 2) "animating" people to develop a collective critical consciousness in order to agitate on their own behalf. This fundamentally radical critique, drawn from traditions of popular education, assumes the capacity of everyday citizens to

generate valid analyses of their own situations and recognizes that some knowledge, values, and interests may be fundamentally incompatible (for example, the interests of the poor to a life of material and spiritual well-being as vs. the interests of elites in maintaining the status quo).

CRITIQUES OF PARTICIPATION AS THE “NEW ORTHODOXY”

Despite these potential benefits, PPA has been subject to intense scrutiny and criticism for at least a decade. Participatory methodologies have become hot topics, because these practices currently serve to fill the gap left by the withdrawal of the state under globalized neoliberal regimes. Bastian and Bastian and Cooke and Kothari, for example, argue that participatory strategies have become “the new orthodoxy” in international development practice.^[20,21] Critics argue that myopic focus on local economic regeneration or community decision making leaves scholars and activists unable to articulate a “holistic analysis of inequality and injustice...which can leave us misguidedly supporting policies which encourage participation as a further erosion of rights in favour of responsibilities, the end result of which is a continued transfer of resources from the poorest communities” (on page 175 of Ref.^[22]). Postpositivist critiques of policy making have not always resulted in convincing post-positivist methods or analytical instruments (e.g., Refs.^[6,16]), and, in this uncertain environment, concerns about PPA have taken several forms.

Lack of Attention to Social Structure

Some scholars and practitioners argue that participatory practices displace analysis of larger, structural workings of power by focusing on individual and “local” empowerment. This focus on microlevel intervention can obscure, and indeed sustain, broader macrolevel inequalities and injustice. The success of PPA for emancipatory goals relies on the transformation of major institutions, but participatory process often myopically sees the “local” as the only site available to intervention and change, while “extralocal” agencies are unmarked, universal, technical and scientific, and therefore given. The result is a grossly oversimplified view of power, which can obscure the operation of the state and other powerful interests to organize social relations in ways that reproduce social inequality. By acting as a stand-in for social justice, poorly developed or cynically practiced participatory methods easily facilitate the illegitimate and/or unjust

exercise of power (become tyrannical), acting as a screen behind which the levers of business or other powerful interests remain intact.

Implication in Devolution and Neoliberalism

It is perhaps not coincidental that attention to participatory practices has coincided with the rise of devolution and neoliberalism, particularly in the United States and United Kingdom. Ledwith argues that, since the 1970s, the twin political economic trends of devolution and neoliberalism “transform rights into responsibilities by transferring the collective responsibility of the welfare state to the individual, the family and community as moral responsibility” (on page 171 of Ref.^[22]) by hijacking the language and practices of liberatory social movements. PPA may work to reproduce this dynamic if “local” interactions take priority over the “extralocal” structures that shape them. Without attention to scale change, these techniques can serve to obliterate possibilities for dialogue, suppress opportunities for social learning, and delegitimize broad-based methods for ensuring equity, such as elections, legislative interventions, trade unionism, and solidarity politics.^[23]

Subjection Disguised as Empowerment

At worst, participatory strategies become a technology for enrolling the oppressed into their own subordination, as when participatory strategies are cynically used to overcome implementation barriers. A participatory focus can mask a wide variety of practices. Some practices simply extract a contribution of unpaid labor from the “beneficiaries” as “exchange” or in order to build “independence.” Some participatory strategies attempt to promote self-reliance, but only because project leaders assume that the participants’ dependency is the reason for their marginalization and exploitation. This can be the case with “Welfare to Work” programs in the United States that “offer impoverished women ‘choices’ that aren’t really choices, such as ‘become self-sufficient or get off welfare,’ and represent any failure as failure of the individual.”^[24] Participatory strategies can also be a simple means to manage projects efficiently or increase sustainability, a market research tool that creates the desires it purports to discover, or a public relations strategy, playing a greater role in the representation of projects to funders than in implementation. In this way, participatory strategies allow government agencies, NGOs, or other social institutions to pass hidden costs on to, while requiring contributions of labor, cash, or kind from, project participants.

PPA IN ACTION

Examples of PPA are now fairly common in the professional literature (see, e.g., Refs.^[6,11,16,17,23,25]). The few that are listed below are not intended to be a representative sample, but rather an indicator of the breadth of PPA projects.

Highlander Research and Education Center

The Highlander Folk School (later Highlander Research and Education Center), a residential self-education program, was formed in 1932 by Myles Horton, Don West and Jim Dombrowski to “train community leaders for participation in a democratic society” and to help spread democratic principles to all human relationships in every political, economic, social, and cultural activity.^[26] Their programs over the last 74 years—including workshops on labor organizing and education (1930s and 1940s), desegregation in the public schools (1950s), citizenship and voter registration (1950s and 1960s), civil rights organizing and leadership (1960s), strip mining and toxic waste dumping in Appalachia (1970s), economic globalization issues (1980s), and multilingual organizing, inter-racial coalitions and youth leadership in the American South (1990s–present)—have been incredibly ambitious. Nevertheless, Highlander’s “method” is quite simple. Highlander staffers actively recruit community organizers, bring them together in a neutral space, and facilitate them through a three-step mutual education process: 1) State and analyze your community’s problems; 2) choose a few central problems to discuss with the aid of consultants; and 3) choose a strategy and plan for action when you go back home.^[27] Highlander has sustained some outstanding community leaders (such as Rosa Parks, Esau Jenkins, Septima Clark, and Bernice Robinson) and pioneered a number of revolutionary educational techniques (for example, citizenship schools).

Consensus Conferences

The Consensus Conference is a technique, pioneered by the U.S. National Institute of Health but most popular in Denmark, which convenes a group of “non-expert” citizens to guide technological deliberations and to make policy recommendations based on consensus procedures. The Danish Board of Technologies (DBT) for example, sends 2000 invitations to a random sample of Danish citizens, and conference participants are chosen to reflect the age, gender, educational, occupational, and geographic diversity of Denmark. Participants gather for an informal set of meetings to receive background on the technology under

investigation, formulate an approach to the topic, and draft questions for a panel of experts. The conference itself begins when a variety of experts are called to make presentations, outlining their understanding and recommendations for the issue at hand. The citizen panel questions and cross-examines them and then, guided by a professional facilitator provided by the DBT, spends several days deliberating, coming to a consensus, and preparing a report outlining the issues that bear on the topic and their recommendations. Finally, they make a public presentation of their findings and the resulting report to government officials, policy makers, other citizens, and the media.

Parables to Policy

The Southern Rural Development Initiative (SRDI), a collaboration of 32 community-based institutions operating in persistently poor rural communities across 11 southern states in the United States, created the “From Parables to Policy” project in order to take advantage of the storytelling skills and traditions of rural Southerners.^b *Parables to Policy* takes community-based stories and transforms them into web-based, multimedia “parables” (stories with a purpose), intended to “serve as grist for policy dialogues that recognize and respond to rural community needs and visions.”^[25] SRDI believes that these parables serve three purposes: 1) creating opportunities for community self-discovery; 2) stimulating peer learning and discovery across the South; and 3) providing distant policy makers with alternative “hearings” to develop better understandings of southern rural poverty.

CONCLUSION

In closing, PPA may offer important resources for enriching the policy- and decision-making processes. Although policy makers, scholars, and activists should be deeply suspicious of participatory practices that displace macrosocial analysis, neglect the extralocal, eviscerate the state’s commitment to social welfare, or heighten emphasis on personal or individual responsibility under the guise of “empowerment,” broad-based grassroots citizen participation should be a feature of any genuine democracy. These techniques are particularly important to explore now, as formal political participation in the United States reaches new lows, sectarian conflict and violence grow worldwide, and economic globalization increases economic, political, and social inequalities. In this context, PPA

^b<http://www.srdi.org/parables/>

can offer new ways to foster understanding between citizens, redistribute the risks and benefits of government programs, and increase the accountability of decision makers.

ARTICLES OF FURTHER INTEREST

Economic Development, Citizen Participation and, p. 569.

Postpositivist Perspectives in Policy Analysis, p. 1530.

Citizen Participation in the Budget Process: Exit, Voice, and Loyalty, p. 305.

REFERENCES

1. Danziger, M. Policy analysis postmodernized. *Policy Stud. J.* **1995**, 23, 435–450.
2. deLeon, P. *Democracy and the Policy Sciences*; SUNY Press: Albany, 1997.
3. Dryzek, J.S. *Discursive Democracy*; Cambridge University Press: New York, 1990; 5–6.
4. Fischer, F., Forester, J., Eds.; *The Argumentative Turn in Policy Analysis and Planning*; UCL Press: London, 1993.
5. Fischer, F. *Reframing Public Policy: Discursive Politics and Deliberative Practices*; Oxford University Press: London, 2003.
6. Mayer, I. *Debating Technologies: A Methodological Contribution to the Design and Evaluation of Participatory Policy Analysis*; Tilburg University Press: Tilburg, Netherlands, 1997; 25 pp.
7. Stone, D. *Policy Paradox*, 2nd Ed.; W.W. Norton: New York, 1997, Personal correspondence, July 2007.
8. deLeon, P. Democracy and the policy sciences: aspirations and operations. *Policy Stud. J.* **1994**, 22 (2), 200–212.
9. deLeon, P. Participatory policy analysis: prescriptions and precautions. *Asian J. Public Admin.* **1990**, 12, 29–54, 30.
10. deLeon, P.; Longobardi, R.C. Policy analysis in the good society. *Good Soc.* **2002**, 11 (1), 37–41, 39.
11. Durning, D. Participatory policy analysis in a social service agency: a case study. *J. Policy Anal. Manage.* **1993**, 12 (2), 297–322.
12. Renn, O.; Webler, T.; Rakel, H.; Dienel, P.; Johnson, B. Public participation in decision making: a three-step procedure. *Policy Sci.* **1993**, 26 (3), 189–214.
13. Torgerson, D. Between knowledge and power: three faces of policy analysis. *Policy Sci.* **1986**, 19 (1), 33–59.
14. Wagle, U. The policy sciences of democracy: the issues of methodology and citizen participation. *Policy Sci.* **2000**, 33, 207–223, 214.
15. Susskind, L.; Cruickshank, J. *Breaking the Impasse: Consensual Approaches to Resolving Public Disputes*; Basic Books: New York, 1987.
16. Beierle, T.C.; Konisky, D.M. Values, conflict, and trust in participatory environmental planning. *J. Policy Anal. Manage.* **2000**, 19 (4), 587–602.
17. Broome, B.J. Participatory planning and design in a protracted conflict situation: applications with citizen peace-building groups in Cyprus. *Systems Res. Behav. Sci.* **2002**, 19, 313–321.
18. Mettler, S. *Soldiers to Citizens: The G.I. Bill and the Making of the Greatest Generation*; Oxford University Press: London, 2005.
19. Soss, J. Lessons of welfare: policy design, political learning and political action. *Am. Pol. Sci. Rev.* **1999**, 93 (2), 363–380.
20. Bastian, S.; Bastian, N. *Assessing Participation: A Debate from South Asia*; Konark Publishers: New Delhi, 1996.
21. Cooke, B., Kothari, U., Eds.; *Participation: The New Tyranny?* Zed Books: London, 2001.
22. Ledwith, M. Community work as critical pedagogy: re-envisioning Freire and Gramsci. *Commun. Dev. J.* **2001**, 36, 171–182.
23. Campbell, N.; Eubanks, V. Making sense of imbrication: popular technology and ‘inside-out’ methodologies. In *Proceedings of the Participatory Design Conference 2004*; CPSR: Cambridge, MA, 2004; 65–73.
24. Fitzgerald, J. Empowerment-Or-Else. Unpublished paper delivered at the National Women’s Studies Association Conference, Minneapolis, MN, June 2001, Personal correspondence, July 2007.
25. Managing Information with Rural America (MIRA). *A Project of the W.K. Kellogg Foundation: Parables to Policy*. <http://mira.wkkf.org/grantees/srdi/srdi.htm>. Nd (accessed January 2, 2007).
26. Horton, M. Transcript of Tape Recording on Leadership (in preparation for the Highlander workshop on *World Problems, the United Nations, and You*). 1954. Highlander Research and Education Center Archives Group 2, Series I, Box 7, File 117.
27. MacLean, K. It’s A Miracle-I Still Don’t Believe It: An Interview with Myles Horton. *Phi Delta Kappan*, May **1966**. Highlander Research and Education Center Archives, Group 1, Series IV, Box 1, File 35.

Patronage and Spoils

Donald E. Klingner

*Graduate School of Public Affairs, University of Colorado at Colorado Springs,
Colorado Springs, Colorado, U.S.A.*

INTRODUCTION

Patronage is a public personnel system characterized by maintenance of political or personal loyalty through legislative or executive approval of individual hiring decisions; spoils is the practice by which successful electoral candidates reward individuals for personal or political loyalty, and secure their continued cooperation during their term of office by awarding them public jobs. Political appointees may be fired at any time; they serve at the discretion of those who appoint them.

This article discusses the following: 1) the role of patronage in democratic government; 2) patronage as a stage in the evolution of public personnel systems; 3) the relationship between patronage and political culture; and 4) maintaining patronage within acceptable limits.

THE ROLE OF PATRONAGE IN DEMOCRATIC GOVERNMENT

Patronage and spoils have long been considered essential to democratic government. But, they are also a threat to it. The relationship between patronage and civil service is intermittently marked by intense conflict, for both systems represent powerful and legitimate values. Civil service systems base decisions about whom to hire, reassign, or promote on applicants' and employees' competencies, rather than political loyalty. Just as the Constitution protects citizens from unfair actions of government officials, so merit systems protect public employees' rights to job security and due process from inappropriate partisan political pressure.

PATRONAGE AND THE EVOLUTION OF PUBLIC PERSONNEL SYSTEMS

In the United States

Public jobs were first allocated primarily among elite leaders—the small group of upper-class property owners who had led the fight for independence and

established a national government. Next, the emergence of political parties created a patronage system that rewarded party members and campaign workers with jobs once their candidate was elected. The spoils system expanded as the functions of government and the number of government employees grew after the Civil War (1861–1865). Political “machines” developed in big cities to nominate candidates and win elections. With electoral victory came the opportunity and obligation to dispense public jobs as “spoils” to those who had worked hardest for the party. The “machine” verified party loyalty, cleared applicants, and collected “voluntary” assessments from job holders to pay the party officials who had provided the job and to finance future election campaigns.

The patronage helped millions of immigrants' transitions to life in a complex urban society. Effective party machines, partisan political leaders (often serving as role models, because they were of the same culture and language as the immigrants they represented), and local precinct workers brought government to these people. In some cases, “government” was the help they needed to survive—baskets of food when they had none, buckets of coal in the winter, help in paying for weddings and funerals, help filling out immigration and voter registration forms, and sometimes a job with the city gas company or street repair department when one became available. While it does not necessarily result in the selection of highly qualified employees or provision of efficient government services, the patronage system enables elected officials to achieve political objectives by placing loyal supporters in key administrative positions. And, it exists today because elected officials get elected, and reelected, by providing information and help to citizens who feel that civil service “bureaucrats” are not able or willing to help them.

Much of the history of public personnel management can be viewed as efforts to reconcile civil service and patronage systems at an operational level. The Pendleton Act (1883) created the civil service system at the federal level, leading eventually to the development and implementation of civil service systems for a majority of professional and technical positions. The period between 1883 and 1937 is important in the development of public personnel administration

Table 1 Evolution of public personnel systems and values in the United States

Stage of evolution	Dominant value(s)	Dominant system(s)	Pressures for change
One (1789–1828)	Responsiveness	“Government by elites”	Political parties + Patronage
Two (1828–1883)	Responsiveness	Patronage	Modernization + Democratization
Three (1883–1933)	Efficiency + Individual rights	Civil service	Responsiveness + Effective government
Four (1933–1964)	Responsiveness + Efficiency + Individual rights	Patronage + Civil service	Individual rights + Social equity
Five (1964–1980)	Responsiveness + Efficiency + Individual rights + Social equity	Patronage + Civil service + Collective bargaining + Affirmative action	Dynamic equilibrium among four competing values and systems
Six (1980–now)	Responsiveness + Efficiency + Individual Accountability + Limited government + Community responsibility	Patronage + Civil service + Collective bargaining + Affirmative action + Alternative mechanisms + Flexible employment relationships	Dynamic equilibrium among competing and interacting values and systems

based on merit principles, at least partly in opposition to the disadvantages of patronage and spoils. The Civil Service Reform Act (CSRA) of 1978, passed almost a century later, was designed to maintain bureaucratic responsiveness but still protect federal career civil service from political interference (Table 1).

In Developing Countries

The modernization process appears to be relatively uniform, because pressures for modernization and democratization tend to parallel though lag behind those in the Western world;^[1,2] administrative reforms are introduced by Western consultants or exposure to the West; and Western lenders often mandate administrative reforms as a condition of continued credit.^[3]

In the first stage, the elite leaders of successful independence movements establish new nations. The transition to patronage follows as these emergent nations strive to strengthen the conditions in civil society that underlie effective government (such as education, political participation, economic growth, and social justice) by refining their constitutions, developing political parties, and creating public agencies. The third stage, if it

occurs, is a transition from patronage to merit marked by passage of a civil service law, creation of a civil service agency, and development of personnel policies and procedures. It happens due to internal pressures for efficiency (modernization) and human rights (democratization). Often, international lenders and donor governments add external pressure by tying funding to increased government capacity, transparency, and citizen participation (Table 2).

THE RELATIONSHIP BETWEEN POLITICAL CULTURE AND PATRONAGE

The transition from patronage to civil service and other public personnel systems reflects a historical process of evolution and changes in underlying values.

In the United States, the Irish bosses who ran big-city political “machines” epitomized patronage. These were based on dominant values of social conservatism and amoral familism.^[4] The exchange for votes often led to widespread electoral corruption. But the amoral familism of the community appeared to tolerate such irregularities despite exposure by the media and crusades by reformers.^[5] The demise of the Irish machines

Table 2 Evolution of public personnel systems and values in developing countries

Stage of evolution	Dominant value(s)	Dominant system(s)	Pressures for change
One	Responsiveness	“Government by elites”	Political parties + Patronage
Two	Responsiveness	Patronage	Modernization + Democratization
Three	Efficiency + Individual rights	Civil service + Patronage	Responsiveness + Effective government
Four	Responsiveness + Efficiency + Limited government	Patronage + Civil service + Collective bargaining + Privatization	Dynamic equilibrium among pro- and anti-governmental values and systems

is attributed to the challenge of new ethnics in the 1930s as well as New Deal federal programs that provided employment and benefits outside of the patronage system. By the early 1950s, most of the machines were gone or badly weakened, with Dailey's Chicago machine an anachronistic exception.^[6] By the 1960s, not only had most Irish-Americans moved away from amoral familism, but many had also become champions of mainstream values (e.g., honesty, efficiency, and impartiality) and even the modern conservative movement.

It has taken the United States over two centuries to develop an effective balance between patronage and civil service. The development of public personnel management has occurred within a context of almost two centuries of democratic government under a single Constitution, and within a civil society widely considered controlled by laws rather than individuals. Though our policy-making process is costly, complex, and tortuous, it results in outcomes that are generally considered to be transparent, effective at maintaining government authority, and politically responsive to the will of the electorate. Our tax system functions well, in spite of (or because of) its voluntary nature. While our society is deeply affected by conflicts

based on race, ethnicity, and class, it provides great opportunity for personal growth and economic advancement. Our political and administrative processes are generally open to public scrutiny by a free press and public records laws. Patronage may no longer be a major issue in most levels and sectors of government, but cynics would respond that it is less serious only because it has been replaced by corrupt contracts as a more effective means of exchanging campaign contributions for access to public officials.^[7] Thus, public personnel systems developed through successive (and successful) fights against the excesses of patronage, and against social pressures to be the "employer of last resort," in a well-developed economy that provides ample jobs outside government. Though our conflicts with corruption, cronyism, and nepotism are not completely resolved, we expect that government will provide services efficiently, using honest and qualified employees.

By contrast, less-developed countries may be characterized by factors—mostly beyond their control—that make it difficult to establish conditions of statehood we in the United States take for granted: a national identity, the rule of law, and a self-sufficient economy. Even the development of stable patronage

Table 3 How country conditions affect public personnel systems

Negative indicators	Positive indicators
<i>1. Successful Transition from Independence to a Functional Patronage System</i>	
• High reliance on charismatic leadership • Restricted freedom of speech and press • High emphasis on export of agricultural products and raw materials • Capital flight • Discrimination (race, ethnicity or class) • Inadequate electoral process	• Stable political parties • Open information and free media • Balanced, domestically-focused economic growth, including professional/technical • Domestic reinvestment of capital • Social justice (minorities and women) • Functioning electoral process
<i>2. Successful Transition from Patronage to a Functional Civil Service System</i>	
• Government process considered low on effectiveness, rationality, transparency • Widespread patronage appointments, and job retention based on salary "kickbacks" • High unemployment or under-employment • Public sector the "employer of last resort" • Underpaid, under-qualified public employees • Widespread employment discrimination based on race, gender or ethnicity • High degree of administrative formalism • High role of the military in civil society and government • Government reforms generated mainly by international economic and political pressure	• Government process considered high on effectiveness, rationality, transparency • Civil Service law, public personnel agency, and policies and procedures • Low unemployment or under-employment • Balanced economic growth/development • Adequately paid, qualified civil service • Low level of employment discrimination based on race, gender, or ethnicity • Low degree of administrative formalism • Reduced role of the military in civil society and government • Government reforms due mainly to domestic political, social and economic pressure

systems may be hampered by societal conditions (e.g., non-functional justice systems, inability to meet even minimum standards of education and health care, political leadership based on “cults of personality” rather than true pluralist political parties, and overly centralized and authoritarian political systems). These conditions generally impede the evolution of rational administrative structures and systems.^[8] For example, organizations in many less-developed countries share common structural and managerial attributes that differ from those typically found in North America, Europe, and Japan: *low* levels of role specialization, formalism, and morale; and *high* levels of centralization, paternalism, authoritarian leadership, rigid stratification, and dysfunctional conflict.^[9] Developing countries that lack a strong cultural tradition of the “public interest,” the rule of law, or a professional public service may also suffer from administrative formalism: laws, agencies, and practices that look fine “on paper” may not function well in reality.^[10] And clearly, the most significant difference is that we in this country were able to first progress from patronage to civil service, second integrate them into an effectiveness model that combined efficiency and patronage, then integrate affirmative action and collective bargaining into the mix, and finally establish the boundaries between public personnel management and emergent market-based techniques like privatization and service contracting.^[11] Developing countries today usually face these pressures simultaneously (Table 3).

LIMITING PATRONAGE AND SPOILS WITHIN APPROPRIATE BOUNDS

The relationship between patronage and civil service is intermittently marked by intense conflict, for both systems represent powerful and legitimate values. The principles of merit and political neutrality reflect the civil service ideal—a competent, committed workforce of career civil servants is essential to the professional conduct of the public’s business.^[12] Patronage diminishes organizational effectiveness and efficiency, and individual rights, in that elected officials consider political loyalty the most important selection criterion, regardless of the applicants’ qualifications. The civil service model views personnel management as a neutral administrative function; the patronage model views it as a policy- or management-oriented function under the executive branch.

Constitutional Checks

While popular political opinion focuses on the politicization of judicial opinions, in fact, a significant and

growing body of case limits political patronage in the United States.^[13]

Legal and Structural Checks

Within the U.S. national government, the Civil Service Reform Act of 1978 established that public personnel management agencies had at least two contradictory objectives—protecting employee rights and making agencies politically responsive—that required that the Civil Service Commission (CSC) be split into two agencies, the Merit Systems Protection Board (MSPB), and the Office of Personnel Management (OPM).

Conflict Among Public Personnel Systems and Values Tends to be Self-Limiting

Conflict among values and public personnel systems is limited and regulated by the dynamic realities of the competition. Because jobs and resources are finite, jobs allocated through one system cannot be allocated through others. Each value, carried to its extreme, creates distortions that limit the effectiveness of human resource management, because other values are suppressed. So, attempts by each system or value to dominate lead inevitably to stabilizing reactions and value compromises. Responsiveness carried to extremes results in the hiring of employees solely on the basis of patronage, without regard for other qualifications; or in the awarding of contracts based solely on political considerations (graft and corruption). Elected officials ensure the continued political responsiveness of civil service systems by patronage appointments of agency heads responsible for managing the bureaucracy, legislative control over agency budgets and staffing levels, and executive control over hiring decisions.

REFERENCES

1. Klingner, D.E. Public personnel management and democratization: a view from three Central American Republics. *Public Adm. Rev.* **1996**, 56 (4), 390–399.
2. Bekke, H., Perry, J.L., Toonen, T., Eds.; *Civil Service Systems in Comparative Perspective*; Indiana University Press: Bloomington, 1996.
3. Salgado, R. *Public Administration for Results: Choice, Design and Sustainability in Institutional Development and Civil Service Reform*; DPP Working Paper Series; Interamerican Development Bank: Washington, DC, 1997; 106 pp.
4. Kenny, K. *The American Irish: A History*; Longman: Harlow, England, 2000.
5. Reedy, G.E. *From the Ward to the White House: The Irish in American Politics*; Charles Scribner’s Sons: New York, 1991.

6. Cohen, A.; Taylor, E. *American Pharaoh. Mayor Richard J. Daley: His Battle for Chicago and the Nation*; Little, Brown and Company: Boston, 2000.
7. Darrough, M.N. Privatization and corruption: patronage vs. spoils. *Int. Public Manag. J.* **2000**, *2* (2), 273–298.
8. Ruffing-Hilliard, K. Merit reform in Latin America: a comparative perspective. In *Handbook of Comparative and Development Public Administration*; Farazmand, A., Ed.; Marcel Dekker, Inc.: New York, 1991; 301–312.
9. Kettl, D. The global revolution in public management: driving themes, missing links. *J. Policy Anal. Manage.* **1997**, *16* (3), 446–462.
10. Kearney, R.C. Spoils in the Caribbean: the struggle for merit-based civil service in the Dominican Republic. *Public Adm. Rev.* **1986**, *46* (2), 66–80.
11. Klingner, D.E.; Sabet, M.G. Contemporary public human resource management: patronage, civil service, privatization, and service contracting. In *Encyclopedia of Life Support Systems*; Tummala, K., Ed.; UNESCO: New York, 2001.
12. Heclo, H. *A Government of Strangers*; Brookings Institution: Washington, DC, 1977; 20 pp.
13. Hamilton, D.K. The continuing judicial assault on patronage. *Public Adm. Rev.* **1999**, *59* (1), 54–62.

Pay for Performance and Merit Pay

Joan E. Pynes

Department of Government and International Affairs, University of South Florida,
Tampa, Florida, U.S.A.

INTRODUCTION

Public organizations need to rethink their traditional structures and reward systems to recruit, attract, and retain talented employees. Today's employees seek to be challenged and want to be provided with opportunities for skill building, a rewards system based on performance, and autonomy to complete work assignments.^[1-3] Many job seekers have an expanded view of public service; they define public service as work that provides a tangible benefit to society. Today, Americans are willing to move across the government, private, and non-profit sectors.^[4,5] Some of the common criticisms of working for the public sector are time-in-grade requirements, lower salary, and lack of variety in assignments. The public sector often does not pay competitive salaries, especially for education and experience. More traditions and expectations exist in the public sector than in the private or non-profit sectors. Public employees have come to expect lifetime employment, secure benefits, and pay raises associated with longevity. Although those expectations may have been attractive for a different generation of workers, to compete in today's labor market, public organizations need flexible options that enable them to strategically use rewards. As a way to reward talented and productive employees, various pay-for-performance/merit-pay techniques have become more prevalent in the public sector.

THE CONCEPT OF PAY-FOR-PERFORMANCE SYSTEMS

Pay-for-performance systems are grounded in the belief that individuals should be paid according to their contributions; increases are rewarded on the basis of performance rather than seniority, equality, or need.^[6] As logical as that may sound, research over the years has indicated that merit-pay systems have not achieved the expected and desired results.^[6-9]

Critics claim that the pay-for-performance evaluation process is subjective, that employees are rated by instruments that do not reflect their actual job competencies, that supervisors lack skills to develop performance standards and provide feedback, and that

comparing individuals with one another sets up a competitive environment that can be destructive to department/unit cohesion.^[10] An additional criticism is that adequate financial resources are not always allocated. For example, New York city signed a contract with the American Federation of State, County and Municipal Employees (AFSCME) District Council 37 that provides a pay raise of 8% and gives the mayor some merit-pay flexibility. However, union officials recognize that, because the contract was more generous than what city officials had originally planned for, the city would not be able to find much more in its budget for large merit payments on top of the across-the-board raises.^[11]

Even when pay rewards are not restricted, the small percentage of difference between high and low performers typically found in merit systems does not encourage improved performance or reward outstanding employees.^[6] Merit-pay systems have been condemned for focusing on compensation, rather than on improved performance. Research has found that when pay and performance are discussed, employees fail to address the developmental issues and, instead, focus on not receiving a pay increase or on receiving an increase lower than expected. When provided with constructive feedback in a training and development context, employees are likely to accept the information. However, employees process the information differently when feedback is tied to pay increases. They tend to get defensive, believing that the rater is taking something away from them by not granting a pay increase. Other research indicates that when performance appraisal results determine pay, employees often set lower goals so that they can achieve them.^[12,13]

The concepts of procedural justice and distributive justice must be considered when developing and administering pay-for-performance systems. *Procedural justice* focuses on the perceived fairness of the evaluation procedures used to determine performance ratings and/or merit increases. For example, what procedures and/or instruments are used to guarantee a link between pay and performance? *Distributive justice* focuses on the perceived fairness of the rating and/or increase received relative to the work performed.^[14] For example, are the ratings and/or increase congruent to the performance inputs? Merit systems that are not

developed with these principles in mind will lack the integrity and credibility necessary for employees to perceive that the system can discern and will reward differences in performance.

Research by Heneman indicated that for employees to perceive the process as just, five components must exist: 1) performance must be clearly defined; 2) rewards must be communicated to employees; 3) rewards must be made contingent upon desired performance; 4) opportunities such as training to improve performance must exist; and 5) the perceived relationship between rewards and performance should be viewed as important as the actual relationship.^[6] The U.S. Merit Systems Protection Board also includes the need for transparency and includes that there be checks and balances in the system (procedural and distributive justice), there be sufficient funding to provide high-performing employees with meaningful pay increases and bonuses, and the pay-for-performance systems be evaluated regularly and modified when necessary.^[15]

To be successful, pay for performance must be linked to the strategic mission of the organization and upper-level management must support the plan. Employees should participate in the development of the plan. This increases their understanding, commitment, and trust in the plan. Organizations must provide training to the raters and hold them accountable for the accuracy of their ratings.^[6,8,9,15]

The community of Aurora, Colorado, failed in its attempt to introduce a merit-pay plan. The city's supervisors were not trained in making evaluations, they were not sure what the standards were, and they did not know how to provide performance feedback. The system produced uneven results and, therefore, lacked credibility among the employees. City Manager Jim Griesemer acknowledged, "We hadn't made the investment in our supervisors that was necessary to give them the tools to do this as well as they were capable."^[16]

The less-than-favorable endorsement of merit-pay programs has not discouraged the public sector from attempting to institute them, and there have been some successes. A number of federal agencies have implemented pay-for-performance systems such as the Navy Demonstration Project at China Lake, the National Institute of Standards and Technology, the Naval Research Laboratory, the Department of Commerce, the Naval Sea Systems Command Warfare Centers, and the Civilian Acquisition Personnel Demonstration Project.^[17,18] The pay-for-performance system introduced in the Navy Demonstration Project at China Lake was funded by money reserved for step increases, quality step increases, and promotions between grades included in a pay band. An additional 0.8–1.0% was reserved for cash bonus. Employees were evaluated

on a five-level rating system, linked to the system of pay increases. Evaluation of the Navy Demonstration Project at China Lake found that money was a motivator, performance was rewarded and the rewards were considered fair, and employees believed that good and bad performances were distinguished and that the process was fair.

The City of Denver has a pay-for-performance plan. City employees' overall performances are rated as *needs improvement*: failed to meet critical performance standards; *successful*: consistently achieved performance standards; or *exceptional*: consistently surpassed performance standards. Employees evaluated as *needs improvement* are not eligible for merit increases. The amount of increase depends on the employees rating of *exceptional* or *successful* and their steps and classifications in the salary schedule (Brown, D.; personal communication, August 28).^[19]

The Town of Cary, North Carolina, developed a pay-for-performance program, in response to the Town Councils' direction to develop a program that places more of the focus on funding and rewarding the best performers. In fact, the pay-for-performance policy states the following: to support the Town's mission by encouraging employees to develop and contribute to the maximum of their potential by compensating those employees based on job performance.^[20]

The program recognizes various forms of performance through varying levels of compensation. Market adjustments to the pay plan keep the plan competitive. No automatic raises are provided, employees receive base and/or lump sum increases. The amount that employees are eligible for depends on their performance level and where they fall within the salary range. Movement beyond the market rate of pay for a pay range is reserved for employees who receive ratings of proficient or exemplary. Lump sum special performance awards of any dollar amount up to and including 10% of base salary may be awarded with the prior approval of the Town Manager. These awards may be made to individuals or groups/team members, responsible for singular achievement or accomplishments that bring special recognition to the town or generate substantial savings or create significant organizational improvements.^[20,21]

GAIN SHARING

Gain sharing is a team bonus program that measures controllable costs, such as improved safety records or decreases in waste or units of output. Teamwork is encouraged, and all team members are rewarded for controlling costs or improving productivity. Formulas are used to measure costs that are controllable, and these costs are then compared with the costs of a

historical base period. When performance improves relative to the base period, a bonus pool is funded. When performance falls short, no bonus pool is created. Employees keep a percentage of the bonus pool, and the organization keeps the rest. Charlotte, North Carolina, began a gain-sharing program in 1994. Each year, the city manager sets a citywide savings goal. If the goal is met, one-half of the savings are set aside to create an employee incentive pool and the balance is returned to the general fund.^[22-24] One-half the pool is paid, in equal shares, to all employees, and payouts from the other half depend on the achievement of department goals. Each department establishes two to five specific goals, which are selected from customer satisfaction, productivity, quality, time standards, and/or safety. If a department has three goals and meets one, employees receive 33% of the second half of their share of the pool.

Hamilton County, Ohio, has a gain-sharing program titled Project Gain. The rationale and the policy are provided below.

Hamilton County desires to reward county employees for exceeding specific performance goals above and beyond the normal scope of duties. An eligible employee participating in an approved program will be compensated based on enhanced revenues, improved efficiencies, or improved service quality that generates sustainable savings. Project Gain employee participation is voluntary. Rewards are shared by groups, not individuals, encouraging employee/employer partnership for improvement (Irwin, S.; personal communication, August 25, 2006; Stephenson, E.; personal communication, August 29, 2006).

Two projects were recently rewarded under Project Gain. Employees designed and developed a court case management application and support system that met the needs of the Probate Court. Costs associated with retaining an outside vendor to do the same were eliminated. Between \$135,000 and \$150,000 in support costs were saved. Employees received monetary awards ranging from \$800 to \$5000. The second project successfully completed was that Facilities and Public Works employees reduced the cost of snow removal at five county sites. County employees removed snow from parking lot areas, driveways, and areas previously under contract with an outside firm. This resulted in a savings of \$33,349. Employees received awards of \$1000.00 (Irwin, S.; personal communication, August 25, 2006; Stephenson, E.; personal communication, August 29, 2006).

As these examples illustrate, not all gain-sharing plans are the same. The formulas and participative management features need to fit each other and the organization. However, some common critical elements are necessary for any plan to succeed. First, there must be a credible and trusted development

process. Second, employees must believe that improved performance and decreased costs will lead to bonuses, the bonuses must be understandable and large enough to influence performance, and employees must recognize how their behavior can influence the size of the bonus. Third, employees need to be involved in the process; they must have influence over the measures used to calculate the bonus. Fourth, there must be appropriate measures, and they must focus on all the controllable costs. Measures such as units of output, materials, and supplies must all be addressed; otherwise, employees may focus on one cost, leading to its reduction but also to increases in other costs. For example, data-processing clerks may produce a greater number of records, but if the number of errors on the records has increased, the effort has been counterproductive. Or public works employees may be able to maintain and landscape more of the city's property in less time, but if the increased productivity results in equipment breakdowns and expensive repairs, then that is counterproductive. Fifth, the program must be maintained; because missions and environments change, gain-sharing formulas and programs must also change and stay relevant (Irwin, S.; personal communication, August 25, 2006; Stephenson, E.; personal communication, August 29, 2006).^[13,22-24]

SKILL-BASED PAY OR PAY FOR KNOWLEDGE

In skill-based or pay-for-knowledge pay plans, the number of tasks or jobs or the amount of knowledge an employee masters determines pay. It is a compensation system, based on paying for what employees can do and for the knowledge or skills they possess. Under skill-based pay, employees can be expected to perform a broad range of duties. Benefits attributed to skill-based pay from an organizational standpoint include developing a cross-trained and more flexible workforce, improving the flow of information throughout the organization, placing an emphasis on the work to be done rather than on the job itself, encouraging the acquisition of skills needed to perform various jobs, and increasing employees' interest in and commitment to their work. Benefits from the employees' perspective include higher motivation, increased job satisfaction, and greater opportunities for increased pay.^[25]

The Government Accounting Office (GAO), now Government Accountability Office, implemented a competency-based performance system to create a clear management linkage between employee performance and core values. Competencies are the knowledge skills and attitudes needed by any individual employee to carry out the job effectively. Individuals who contribute to the overall values and objectives of an organization are rewarded.

Twelve competencies tied to meaningful performance in GAO were identified. They are:

- Achieving results.
- Maintaining client and customer focus.
- Developing people.
- Thinking critically.
- Improving professional competence.
- Collaborating with others.
- Presenting information orally.
- Presenting information in writing.
- Facilitating and implementing change.
- Representing GAO.
- Investing resources.
- Leading others.

Employees are placed in one or five pay categories, based on their demonstrated competencies, performance, and contributions to organizational goals.^[26]

Organizations that implement skill-based pay need to be aware that wages and salaries will increase as employees learn new skills. Despite the strategic focus on skill-based training, however, many employees unfortunately will resist its implementation.^[27]

Thompson and LeHew provided a comprehensive and critical review of skill-based pay for public organizations.^[28] They developed a theoretical model that suggested broadening the framework to include a discussion of environmental contingencies, such as threat/crises, competition, and performance pressures; the proximate environment, such as the various stakeholders, unions, and elected officials; the design features of the plan, such as how radical it is, training opportunities available for employees, and the certification process that is involved; and the relevance of those components to organizational contingencies, such as workforce characteristics, managerial practices and attitudes, and congruence to skill-based pay.

CONCLUSION

There are many acknowledged problems with performance appraisals and their linkages with pay for performance/merit pay. Concerns about the accuracy of performance evaluation ratings, procedural fairness in the implementation of the reward systems, increasing competition among employees, and greater conflict between management and employees are all legitimate points.^[10,29-31] However, the risk is too great if public organizations are not willing to experiment with innovative reward programs to attract, motivate, and retain quality performers. Although the lion's share of the research on pay for performance/merit pay tends to emphasize the difficulties and problems in implementing new reward systems, it is important to note that

there have also been some successes. For new reward systems to work, the systems need to fit the organization and its values. They must be as objective as possible and designed to reward top performers, rather than to penalize satisfactory performance. To be successful, pay-for-performance/merit-pay plans need the long-term commitment from multiple stakeholders. Support must come from the elected officials responsible for appropriating funds; executives, middle managers, and supervisors; the support of the exclusive bargaining representatives in unionized environments; and financial resources to make the pay plans meaningful.^[29-31]

REFERENCES

1. Green, M.E. Beware and prepare: the government workforce of the future. *Public Pers. Manage.* **2000**, 29 (4), 435-444.
2. Lachance, J.R. International symposium of the international personnel management association. *Public Pers. Manage.* **2000**, 29 (3), 305-313.
3. Langan, S. Finding the needle in the haystack: the challenge of recruiting and retaining sharp employees. *Public Pers. Manage.* **2000**, 29 (4), 461-464.
4. Light, P.C. The new public service. *Gov. Exec. Mag.*, 2001; <http://www.govexec.com/features/0100/0100s2.htm> GovExec.com. [WWW document, URL.]
5. Light, P.C. *Search of Public Service*; Brookings Institution: Washington, DC, 2003.
6. Heneman, R.D. *Merit Pay: Linking Pay Increases to Performance Ratings*; Addison-Wesley: Reading, MA, 1992.
7. Kellough, J.E.; Lu, H. The paradox of merit pay in the public sector. *Rev. Public Pers. Admin.* **1993**, 13, 45-64.
8. Perry, J.L., Hays, S.W., Kearney, R.C., Eds.; *Compensation, merit pay, and motivation*. In *Public Personnel Administration: Problems and Prospects*; Prentice-Hall: Englewood Cliffs, 1995; 121-132.
9. Risher, H., Fay, C.H., Perry, J.L., Risher, H., Fay, C.H., Eds.; *Merit pay: motivating and rewarding individual performance*. In *New Strategies for Public Pay: Rethinking Government Compensation Programs*; Jossey-Bass, Inc.: San Francisco, 1997; 253-271.
10. O'Donnell, M.; O'Brien, J. Performance-based pay in the Australian public service. *Rev. Public Pers. Admin.* **2000**, 20 (2), 20-34.
11. Greenhouse, S.; Saulny, S. City's municipal workers agree on contract and 8% raise. In *N. Y. Times*, 2001; <http://www.nytimes.com/2001/04/12/nyregion/12UNIO.html> [WWW document, URL.]
12. Cascio, W.F. *Applied Psychology in Personnel*; Prentice-Hall, Inc.: Englewood Cliffs, NJ, 1991.
13. Lawler, E., Gomez-Mejia, L.R., Eds.; *Pay for performance: a strategic analysis*. In *Compensation and Benefits*; Bureau of National Affairs: Washington, DC, 1989; 136-181.
14. Greenberg, J. *The Quest for Justice on the Job: Essays & Experiment*; Sage: Thousand Oaks, CA, 1996.

15. U.S. Merit Systems Protection Board. *Designing an Effective Pay for Performance Compensation System*; Washington, DC, 2006.
16. Witt, E. Sugarplums and lumps of coal. *Governing* **1989**, 2, 28–33.
17. U.S. General Accounting Office. *Human Capital: Implementing Pay for Performance at Selected Personnel Demonstration Projects*; GAO-04-83: Washington, DC, 2004.
18. Schay, B.W., Risher, H., Fay, C.H., Eds.; Paying for performance: lessons learned in fifteen years of federal demonstration projects. In *New Strategies for Public Pay: Rethinking Government Compensation Programs*; Jossey-Bass, Inc.: San Francisco, 1997; 207–230.
19. Rule 13 Pay for Performance. <http://www.denvergov.org/admin/template3/forms/RULE52013.pdf> (retrieved August 30, 2006).
20. Town of Cary, North Carolina Agency Awards for Excellence. October 2004 IPMA-HR News 13-14.
21. Town of Cary, NC, Policy Statement 31 Pay for Performance Policy. <http://townofcary.org/depts/tcdept/policies/031.htm> (retrieved August 30, 2006).
22. Risher, H. Can gain sharing help to reinvent government? *Public Manage* **1998**, 80 (5), 17–21.
23. Risher, H. Merit pay can be a hard sell. *Public Manage* **1999**, 81 (7), 8–13.
24. Risher, H. Are public employers ready for a “new pay” program? *Public Pers. Manage* **1999**, 28 (3), 323–343.
25. Towers Perrin. In *Why Did We Adopt Skill-Based Pay?* Towers Perrin: New York, 1992.
26. U.S. General Accounting Office. *Results-Oriented Cultures: Modern Performance Management Systems Are Needed to Effectively Support Pay for Performance*; GAO-03-612T: Washington, DC, 2003.
27. Feuer, D. Paying for knowledge. *Training* **1987**, 24, 57–66.
28. Thompson, R.; Lehew, C.W. Skill-based pay as an organizational innovation. *Rev. Public Pers. Admin.* **2000**, 20 (1), 20–40.
29. Bowman, J.S. Performance appraisal: verisimilitude trumps veracity. *Public Pers. Manage* **1999**, 28 (4), 557–576.
30. Clifford, J.P. The collective wisdom of the workforce: conversations with employees regarding performance evaluation. *Public Pers. Manage* **1999**, 28 (1), 119–155.
31. Gabris, G.T.; Ihrke, D.M. Improving employee acceptance toward performance appraisal and merit pay systems. The role of leadership credibility. *Rev. Public Pers. Admin.* **2000**, 20 (1), 41–53.

Pay-As-You-Go Financing

David L. Baker

School of Public Affairs, Arizona State University, Tempe, Arizona, U.S.A.

INTRODUCTION

Some governmental budgets aggregate appropriations into one unified, or consolidated, budget (e.g., the federal government). Other governmental budgets are divided into 1) an operating budget and 2) a capital outlay budget. Within the operating vs. capital outlay distinction, an operating budget represents a financial plan that proposes expenditures based on estimated revenues to finance them for a fixed period (usually a defined fiscal year). An operating budget addresses everyday activities of an entity. A capital outlay budget represents a plan for investing in capital assets (also referred to as fixed assets) that are non-recurring and endure beyond the operating budget cycle. Examples of capital assets with a long-term character include land, buildings, infrastructure, equipment, and furniture.

Pay-as-you-go financing (PAYGO) refers to public financing strategies that pay for spending increases or revenue reductions from current resources and without increasing a deficit or debt financing.^[1-3] PAYGO financing serves as a means to achieve budgetary balance and to control deficits for consolidated and operating budgets. For capital budgets, it functions in contrast to pay-as-you-use (debt) financing whereby capital outlays are paid for as they are used through debt service payments. PAYGO financing offers versatile features for consolidated, operating, and capital budgeting that require consideration within a public agency's overall budgetary and financial context.^[4,5]

The following review of PAYGO financing unfolds through several sections. First, PAYGO financing for consolidated and operating budgets is discussed. Second, the review examines the role of PAYGO financing for capital outlay budgeting in comparison with debt financing. This analysis includes capital outlay examples and an example of transitioning PAYGO financing from capital to operational costs. Third, potential PAYGO financing sources are discussed. Fourth, and finally, the review closes with a summary regarding the usefulness of PAYGO financing.

PAY-AS-YOU-GO FINANCING FOR CONSOLIDATED AND OPERATING BUDGETS

Governmental entities operating under a consolidated budget employ PAYGO financing for budgetary balance and deficit control.^[3] Federal budgeting serves as an example of this. The 1990 Federal Budget Enforcement Act (BEA) and subsequent amendments to the act establish PAYGO rules for spending and revenue legislation. Generally, congressional action is prohibited from adding to the budget deficit. Spending increases must be offset with other spending reductions or revenue increases. Conversely, revenue reductions must be offset by revenue increases or spending decreases. If full offsets are not made, funds are sequestered from certain expenditure areas.^[6]

"PAYGO is a multiyear control enforced one year at a time."^[6] Congressional budgetary actions are examined annually under the BEA on an aggregate basis, not by individual legislative bills. Federal PAYGO financing does not control spending influenced by population growth and inflation under existing law. Provisions are designed to limit growth in budgets and deficits by linking new spending to other expense reductions or revenue increases. PAYGO financing provisions shift the consequences for budgetary decisions to the stage in the process when spending increases or revenue reductions are first proposed.^[7]

PAYGO financing also refers to the operating budget strategy of assuring that public agencies deal with financial liabilities as they occur. For example, PAYGO financing is at work when pension programs routinely fund the actuarially determined contribution to cover the pension liability accrued for each payroll. This avoids accumulating an unfunded pension liability. PAYGO financing mitigates the temptation to perceive resources as available for new initiatives by allocating funds to meet known obligations as incurred. Additionally, the political desirability of proposed spending confronts the task of selling other program cuts or tax increases to offset the new costs.^[8] PAYGO financing acts as a cost-control measure.

PAY-AS-YOU-GO FINANCING FOR CAPITAL OUTLAY BUDGETS

Public agencies often rationalize PAYGO financing by the purpose of the appropriation. Deciding who will pay for a capital asset tends to be influenced by both the political and economic environments.^[9] Financing methods also introduce conflict over opportunity costs because the amount spent on a capital outlay becomes unavailable for another use.^[1] On the one hand, debt-financing advocates encourage capital assets to be purchased over time (usually the expected use cycle of an asset) by those who benefit from them. This promotes intergenerational equity in a community rather than permitting free riders to receive future capital asset benefits with no financial contribution. This also recognizes that capital assets may be needed today to support tomorrow's growth while such growth generates the revenue stream to retire the debt. On the other hand, PAYGO financing makes sense when short-term and long-term cost savings are desirable because of future fiscal uncertainty or the public agency already services substantial annual debt. Additionally, this strategy may be prudent where no significant concern exists about some citizens benefiting later while avoiding early cost participation.

PAYGO financing is the least expensive way of funding public agency capital outlays. Several relative, but material, agency cost components associated with debt financing are avoided through this financing method. The cost components identified in Table 1 are avoided in PAYGO financing. Costs identified do not include the varied, yet substantial, public agency staff and policy-maker costs associated with borrowing. These include the cost of the time involved in initiating capital asset financing, deal structuring, compiling supporting documentation, managing the project, tracking funds, and filing required financial statements throughout the life of the financing.

In addition to cost avoidances associated with debt financing, other reasons explain why public

agencies turn to PAYGO financing. First, PAYGO financing enables public agencies to manage their debt profile. A public agency's debt profile affects credit ratings.^[1] A higher credit rating translates into lower interest costs on future financings, while a lower credit rating means higher interest costs.^[10] Second, a PAYGO approach generates a certain amount of support. Because current residents often choose a specific capital outlay, there are sentiments that "the cost should not be imposed on future residents who have no say in the matter."^[11] Third, PAYGO financing encourages government to live within its means. Fourth, PAYGO financing constrains the growth of interest costs as a portion of an agency's annual budget. Fifth, and finally, the fiscal capacity of a public agency may be sufficient to fund a needed capital asset from current revenue without burdening future fiscal years.

COMPARISON WITH DEBT FINANCING

Advantages of PAYGO financing for capital outlays require evaluation against the advantages of pay-as-you-use (debt) financing. Common policy and financial considerations with debt financing include the following:

- Does a public agency have the fiscal capacity to provide the necessary resources for essential capital assets from current revenues?
- Are there timing elements that must be factored into the decision calculus of PAYGO financing vs. debt financing (e.g., state- or federal-mandated requirements, pending litigation, or an advantageous, time-limited opportunity)?
- Can debt repayment schedules accommodate the community's ability to pay with synchronized payments aligned with the enjoyment of the benefits?^[11]
- Can a public agency capitalize on low interest rates to stretch precious funds further?
- Does prudent debt financing maximize a public agency's options to invest current proceeds, to address other pressing priorities, and/or to meet unanticipated needs?
- Does borrowing avoid significant year-to-year funding variations that otherwise would be difficult for a jurisdiction to handle?
- Does excessive commitment to PAYGO financing prevent a public agency from capital outlays essential to the community?^[4]
- Does debt financing allow the cost per capita for a capital asset acquisition to decrease over time, assuming population growth? Is this a desirable feature from the community's standpoint?

Table 1 Cost avoidance in pay-as-you-go financing for capital outlays

Debt interest on funds borrowed
Bond counsel
Disclosure counsel
Financial advisor
Trustee/pay agent
Official statement printing and distribution
Rating agency
Underwriter discount or premium included in the securities purchase
Insurance premiums
Credit enhancements

Public finance literature dealing with capital outlays occasionally refers to PAYGO financing as though it is an inviolate, overarching policy of particular public agencies.^[12] A public agency either follows a PAYGO financing policy or it does not follow such a policy in this frame of reference. This portrayal diverges from the more common practice of public agencies employing a mix of financial strategies to meet capital outlay needs. Each capital outlay calls for individual assessment of financial and policy issues germane to that particular capital outlay.

PAY-AS-YOU-GO FINANCING CAPITAL OUTLAY EXAMPLES

Three PAYGO financing examples illustrate the utility of this mainstream public financing strategy. Each involves a public agency investigating PAYGO financing plans for the acquisition and development of a community park. An estimated \$3.0 million is needed for the project.

Variation Examples

1. The public agency may appropriate \$3.0 million, the total cost of acquisition and development, in one fiscal year offsetting the cost with increased revenues.
2. The public agency may appropriate \$3.0 million over multiple years, assuming constant costs, in equal increments or non-equal increments that match available resources by fiscal year. Planned PAYGO financing appropriations may be advanced, suspended, or adjusted in future years based on available resources. Year-to-year funding may be assembled and preserved for a specific capital outlay purpose using an approved reserve (i.e., Good City community park reserve). Accounting treatments may vary by jurisdiction.
3. The public agency may budget and expend \$3.0 million for community park acquisition and development on a PAYGO financing basis in definable phases that coincide with each fiscal year's funding availability. For example, the first year estimated land acquisition price of \$700,000 and design development cost of \$200,000 might be budgeted and expended. In the second year, park development and landscaping, estimated at \$1.2 million, may be funded. The third year might include an estimated \$900,000 for parking, picnic grounds, and specialty areas (i.e., tot lots, volleyball courts, and nature center). At the federal level, this is referred to as "budgeting for stand-alone stages of larger projects."^[13]

Pay-as-you-go financing may be used in conjunction with other funding. Typical examples include: 1) PAYGO financing for some portion of a capital outlay while the remainder is debt-financed; 2) debt-financing a facility while cash-funding the furnishings that have a shorter useful life than the facility financed; and 3) PAYGO financing to match a grant or some other funding source dedicated to the capital outlay.

TRANSITIONING FROM CAPITAL TO OPERATIONAL COSTS

A more elaborate example demonstrates how PAYGO financing can be transitioned from a capital outlay to operating costs associated with a critically needed facility. In this instance, the capital outlay involves a \$6.0 million expansion to an existing county jail. Upon occupancy, the jail expansion affects the operating budget in a dramatic manner.^[14] Once operational, the project calls for approximately \$2.0 million in new, annual operating costs for staffing, food, utilities, clothing, and supplies. Furthermore, operational costs will escalate at an estimated rate of 12–15% annually.

Several steps are taken to finance this facility and its operational cost. First, the public agency decides to use PAYGO financing because 1) it experiences a significant, ongoing increase in property tax revenue caused by development related to population growth and 2) this financing strategy minimizes the agency's debt profile because the jurisdiction anticipates a hospital expansion project requiring \$50–75 million in debt financing within several years. Second, a \$2.0 million surge in revenues is appropriated in two places for the first year of the project: 1) a reserve designated for jail expansion construction and 2) an initial construction project allocation to cover environmental work, architect and engineering fees, and project management. Third, in project year two, the continuing dedicated revenue stream of \$2.0 million is split between building the reserve (\$1.8 million) and ongoing preconstruction costs (\$0.2 million). Fourth, in project year three, a third \$2.0 million increment is budgeted along with the previously accumulated reserve to fully fund the construction. Fifth, once construction is completed, the annual \$2.0 million dedicated revenue stream is applied to cover the ongoing operational cost of the jail expansion.

In the meantime, over the 4-year time span, the incremental growth on the base revenue stream of \$2.0 million aids in addressing the annual operating cost increase of 12–15%. This overall strategy allows a smooth transition from capital outlay to funding the ongoing operational costs as part of the total budget. Graphically, this PAYGO financing is summarized in [Table 2](#) with appropriation elements.

Table 2 Transitioning pay-as-you-go capital outlay funding to operational costs

	Project years			
	1	2	3	4
Revenue (in millions)	\$2.0	\$2.0	\$2.0	\$2.0
Appropriation (in millions)				
Capital outlay	0.6 ^a	0.2 ^a	5.2 ^b	0
Reserve	1.4	3.2 ^c	0	0
Operating costs	0	0	0	2.0 ^d

^aPreconstruction capital outlay costs (i.e., environmental work, architect and engineering fees, and project management costs).

^bComprised of \$3.2 million accumulated reserve from year two plus \$2.0 million revenue increment.

^cComprised of \$1.4 million reserve from year one, plus \$1.8 million revenue increment after covering ongoing preconstruction capital outlay costs.

^dAnticipated annual operational cost.

POTENTIAL PAY-AS-YOU-GO FINANCING SOURCES

Consolidated and operating budgets most frequently obtain PAYGO financing from three sources. First, revenue increases, both non-recurring and recurring, may provide funds for PAYGO financing. Second, spending decreases, both non-recurring and recurring, fortuitous or forced, can fuel a PAYGO strategy. Third, governmental entities may take action to rebudget funds from a lesser priority to a higher priority.^[15]

For capital budgets, PAYGO funds most frequently come from non-recurring and recurring revenue streams. However, there are other sources from which to derive such funds. Common sources after revenues include: 1) use of unanticipated fund balance; 2) non-recurring and recurring savings; 3) grants; 4) cancellation and rebudgeting of a designated reserve; and 5) rebudgeting from a lower priority to a higher priority appropriation.

CONCLUSION

PAYGO financing is an important tool in the public agency toolbox. It offers budget balancing and deficit control features for consolidated and operating budgets. PAYGO financing serves as a cost-effective, sensible alternative to debt financing. Use of this financing method for a particular policy objective requires analysis in the overall financial context of a given public agency. PAYGO financing demonstrates versatility and can be integrated with other strategies.

Finally, PAYGO financing may be utilized flexibly with a range of budgetary sources.

REFERENCES

1. Bland, R.L.; Clarke, W. Budgeting for capital improvements. In *Handbook of Government Budgeting*; Meyers, R.T., Ed.; Jossey-Bass Publishers: San Francisco, 1999; 653–677.
2. Mikesell, J.L. *Fiscal Administration: Analysis and Applications for the Public Sector*, 6th Ed.; Wadsworth Publishers: Belmont, CA, 2003.
3. Voorhees, W.R. Budgeting and deficit controls. In *Encyclopedia of Public Administration and Public Policy*; Rabin, J., Ed.; Marcel Dekker, Inc.: New York, 2003; 98–102.
4. Forrester, J. Municipal capital budgeting: an examination. public budgeting and finance. Summer **1993**, 13 (2), 85–101.
5. Steiss, A.W. *Financial Management in Public Organizations*; Brook/Cole Publishing Company: Pacific Grove, CA, 1989.
6. Schick, A. *The Federal Budget: Politics, Policy, Process*; The Brookings Institution: Washington, DC, 1995; 100 pp.
7. Thurber, J.A.; Durst, S.L. Delay, deadlock, and deficits: evaluating proposals for congressional budget reform. In *Federal Budget and Financial Management Reform*; Lynch, T.D., Ed.; Quorum Books: New York, 1991; 53–88.
8. White, B. Examining budgets for chief executives. In *Handbook of Government Budgeting*; Meyers, R.T., Ed.; Jossey-Bass Publishers: San Francisco, 1999; 462–484.
9. Temel, J.W. *Fundamentals of Municipal Bonds*, 5th Ed.; John Wiley & Sons: New York, 2001.
10. Miller, G.J. Capital asset financing. In *Public Budgeting and Finance*, 4th Ed.; Golembiewski, R.T., Rabin, J., Eds.; Marcel Dekker, Inc.: New York, 1997; 633–651.
11. Horler, V.L. *Guidelines to Public Debt Financing in California*; Rev. Ed. Packard Press: San Francisco, 1987; 12 pp.
12. Moody's Municipal Department. *Moody's on Municipals: An Introduction to Issuing Debt*; Moody's Investor Services, Inc.: New York, 1987.
13. Posner, P.L.; Lewis, T.V.; Laufe, H. Budgeting for federal capital. Public Budg. Finance **1998**, 18 (3), 11–23.
14. Bland, R.L.; Nunn, S. The impact of capital spending on municipal operating budgets. Public Budg. Finance **1992**, 12 (2), 32–47.
15. Forrester, J.P.; Mullins, D.R. Rebudgeting: the serial nature of municipal budgetary processes. Public Adm. Rev. **1992**, 52 (5), 467–473.

Pension Under-Funding: Generational Accounting Perspective

Paul J.M. Klumpes

Liyan Tang

Tanaka Business School, Imperial College London, London, U.K.

INTRODUCTION

Public pension and health-care plans and over generationally sensitive programs of many national and local governments cover the substantial majority of their population and command a substantial proportion of their governments' capital investment obligations. However, unlike privately funded health-care insurance, but similar to other types of publicly managed pension funds and social insurance, these plans are likely to encounter major long-term financial problems. This is because public health-care systems, and most other government-provided social insurance systems, rely on a "pay-as-you-go" ("PAYG") funding arrangement, whereby promised benefits and services are usually paid for by the current active workforce in the form of government taxes. However, the political will of governments to continue funding these promises via PAYG is coming under threat, because the populations of most advanced capitalist economies are ageing. This raises concern whether governments' existing and potential revenues will be sufficient to sustain current levels of public health care into the future.^[1]

Currently, there is considerable actuarial discretion over the amount, timing, and incidence of reported pension liabilities. For example, GASB Statement Nos. 43 and 45 stipulated for other postemployment benefits (OPEBs) have not resolved the issue as to ameliorating the reporting relevance and comparability of funding status of OPEB, given the long-term funding attribute of these benefits.

What is the problem with reporting pension obligations under accrual accounting systems? On the one hand, reporting funded status to a single period in the past confines the relevant information for financial statement users to the ex-post funded status at the balance sheet date that is subject to past investment performance and funding experiences. In securing the reliability of financial reporting, the funded status does not encompass the economic value of pension liabilities that capture information about long-term future cash flows. On the other hand, under the GASB, the unfunded actuarial accrued liability (UAAL) represents the excess of the actuarial accrued liability over the actuarial value of assets. A quirk of these statements is that they permit six actuarial cost methods

to be utilized in determining the portion of future normal cost allocated to the reporting period. Despite the fact that a government may use the same actuarial method for both funding and financial reporting, the comparability of funded status across various OPEB plans is considerably impaired, as the choice of the actuarial method is left to the discretion of the government. In addition, that the actuarial valuation is used for pension assets is because fair value can be volatile in the short run with fluctuations of gains and losses.

CRITICISMS OF CURRENT RULES

The short-term emphasis of current accrual-based reporting for pensions or OPEBs assume that post-employment benefits should effectively be treated as gratuities over a spot contract, where only the benefits, which "legally" accrue to the reporting period when services are rendered, should be disclosed. An underlying assumption of this perspective is the plan termination assumption, which is also known as a "wind-up" assumption. It assumes the precise amount that the firm would require to discharge its "explicit" legal obligation if the plan were terminated.^[2] Consequently, the corporate finance perspective overlooks the long-term employment expectation of pension members under an ongoing plan assumption.^[3]

Concerns have been raised regarding the perceived deficiencies in the accounting for postemployment benefits because of the long-term nature and magnitude of the defined benefit arrangement.^{[4]a} In consideration of the long-term nature of the arrangement, the intergenerational equity perspective is more consistent with the going concern principle in a lifetime contract. Generational accounting is based on the intertemporal constraint of funding postemployment benefits under a defined benefit arrangement. It generates a set of accounts for existing members as opposed to future members of the postemployment plan. Current generational accounts record the present value of cumulative contributions against cumulative benefits from

^aB13, Appendix B, Statement of Financial Accounting Standards No. 158, *Employers' Accounting for Defined Benefit Pension and Other Postretirement Plans*.

funding existing members. In constructing generational accounts, assets and liabilities of the plan are assumed to accrue to a future time when the plan is ceased for the group of existing members, as its last member exists the plan. Hence, funding status is projected to a future date. Table 1 exhibits the difference between current GASBS under the accrual accounting and the generational accounting with respect to various qualitative characteristics of financial reporting.

Note that under the generational accounting, liability is depicted as the sum of cumulative benefit obligations and unfunded benefit obligation. For OPEB, this is based on the definition of the annual required contribution of the employer for the period where the OPEB contributions should cover the normal cost and the UAAL to be amortized in the current period, under the GASBS. The present value approach is consistent with the estimate of fair value codified in GASBS31, Par. 74, considering market prices for similar investments and using discounted cash flow analysis.

In general, taxpayers, employees, and government agencies should be concerned about the buildup of assets and liabilities affecting PAYG-funded health-care systems for several reasons. First, the adoption of accrual-based GAAP highlights the extent to which the private and public sectors are exposed to unfunded pension obligations. Second, population's age, all else being equal, the growth rate of pension, and health-related liabilities will eventually outstrip the growth rate of assets. Consequently, over time, these systems can represent a major form of public borrowing against the future, which has important accounting implications.^[5] They may also powerfully affect future revenue-raising potentials of governments and the health-care premiums established by private health-care firms. Finally, as a matter of government public policy to alleviate social exclusion of the vulnerable and poor, PAYG-funded public health-care systems necessarily involve both distributional inequity, through

the provision of benefits to dependants without additional contributions, and intergenerational inequity.^[6] However such inequities are not recognized by either cash or accrual-based accounting principles used in other public-sector contexts, such as pensions and health-care provision in local communities.

OBJECTIVES OF GENERATIONAL ACCOUNTS

Appleby suggests that a major issue in the accountable financial management of any public PAYG funded system is to provide information about: 1) who pays; 2) how are these payments allocated to taxpayers; and 3) how are they used in return for a standard level of community treatment and care.^[7] However, prior researchers, evaluating the accounting control and financial management of health-care systems, have until now paid scant attention to the accountability implications of underfunded public health-care services or other services provided by local government, the department of social security, or public-sector pension funds.^[8–10]

Generational accounting is a new method of long-term financial planning and analysis. It seeks to answer the following questions: how large a government's pension funding burden does current policy imply for future generations? Is pension funding policy sustainable without major additional sacrifices on the part of current or future generations or major cutbacks in government activities? What alternative policies would suffice to produce generational balance—where future generations face the same burden as to current generations when adjusted for growth? How would different methods of achieving such balance affect the remaining lifetime funding burdens—the generational accounts—of those alive.

Generational accounts are defined as the present value of net taxes (contributions paid less benefits

Table 1 Major conceptual differences between generational and accrual-based perspectives in government financial reporting

Conceptual framework	Accrual-based perspective	Generational-based perspective
Objective	Report on interperiod equity (GASB, on page 22)	Provide information about intergenerational equity (AAS 31, on page 10)
Intended users	Present and potential taxpayers	Current and future generations of taxpayers
Recognition criteria	For each entity, revenues when earned and expenses when related good or service used	For each generation, net payments and receipts over lifetime
Applicable financial reports	Balance sheet and income statement	Statement of intergenerational equity
Assumed pension obligation	Past service costs to date, based on termination or "spin-off" calculation of accrued benefit obligation	Present valuation of future pension benefits or continuation of economic benefit obligation

received) that individuals of different age cohorts are expected, under current funding policy, to pay over their remaining lifetimes. Adding up the generational accounts of all currently living generations gives the collective contribution of those now alive toward the payment of government bills. This refers to the present value of government pension funds' current and future investments plus its net unfunded debt (i.e., its financial liabilities minus its financial and real assets, including those of public-sector enterprises). Any unfunded debt relating to past service obligations to current generations must be paid by future generations.

We now explain the relationship between accrual- and generational-based accounting perspectives in explaining the funding status of the public-sector pension fund and then briefly outlines how the generational accounts are calculated. The methodology underlying these calculations is consistent with that used in the relevant generational accounting literature.

Generational accounting explicitly recognizes generational imbalance between three inter-related variables affecting PAYG funding: 1) required contributions; 2) actual annual health-care contributions; and 3) expenditures and benefit payments. The implications of these obligations for accountability of public-sector entities to constituents is then explored in terms of the performance reporting and unfunded liability commitments, on a "fair valuation" basis. A statement of generational equity is a means to explain the incidence of these commitments across a range of taxpayers, contributors, and beneficiaries.

The concept of generational accounting, and its implications for exploring alternative basis for public sector reporting to those currently prescribed by accrual-based GAAP principles, can improve the understanding of the ramifications of incorporating the full intergenerational equity implications of continuing reliance of the underfunded long-term health-care and pension commitments.

Prior to the 1990s, the public sector in many countries did not, as a matter of practice, produce publicly available accrual-based accounts, showing a statement of their financial resources based on domestic or internationally credible GAAP. However, in 1998, the U.K. government committed itself to the implementation of accrual-based "whole of government" accounting principles as from the 2000 financial year (although most other local governments have still not yet made such a commitment). Under these proposed principles as set out by U.K. Treasury,^[11] liabilities for national insurance and other social insurance programs, or even governmental pension commitments to public-sector employees, are not recognized, because they are assumed to represent moral rather than legal obligations of government (paragraph 90 of Ref.^[11]).

Currently, the International Federation of Accountants has issued "international GAAP" for government entities that do not incorporate these implicit obligations. Instead, to date, the adoption of accrual accounting in the recording of these expenditures means that the vast proportion of these expenditures is recorded as "recurrent": capital expenditure and consumption appear to be immaterial. However, this analysis omits the future obligations likely to be imposed on future taxpayers from privately financed capital expenditure papers required to maintain public infrastructure, which Mayston suggests may be quite substantial in the public sector.^[12] Barton develops similar criticisms concerning the application of accrual-based accounting principles to the Australian public sector.^[13]

GENERATIONAL ACCOUNTING: PROS AND CONS

To the extent that real increments or reductions in a deficit indicates that future generations must help pay for the provision of generationally sensitive public-sector services to the current generation, a full accrual-accounting-based set of financial statements provides only part of the information required to assess issues of intergenerational equity. In order to assess the full intergenerational equity impact of PAYG-financed public social security programs and health-care expenditures on government financial management of countries facing demographic transition, Auerbach et al. propose a system of generational accounting that traces financial flows to and from different generations.^[14] Generational accounts indicate, in present value terms, what a typical individual stakeholder of each generation can expect to pay, now and in the future, in terms of net contributions. Thus, generational accounting indicates not only what existing generations will pay, but also what future generations must pay, given current policy and government's intertemporal budget constraint. This constraint requires that government bills not paid by current generations must ultimately be paid by future generations.^[15]

A major problem with generational accounts is that they are not broadly disseminated to the public arena, since they generally fall outside the scope of reported pension and insurance obligations, in accordance with accrual-based accounting principles. Klumpes develops a statement of intergenerational equity for underfunded health-care and public pension provisions.^[16] The analysis provides detailed intergenerational and intragenerational inequities in the funding and provision of these services, which, if present trends continue, would require increases in contributions or

taxes or decreases in service and/or benefit levels. Klumpes also develops generational accounts for specific insurance-related funds for welfare and pension provision, which demonstrates the sensitivity of public-sector accounts to the demographics of ageing populations.^[17] Tang and Klumpes exploit a fair value-based generational accounting system to investigate the effect of underfunded PAYG pensions on the financial dimensions of Chinese SOEs.^[18]

Financial statements are required to provide supplementary disclosures on the pension liability, assets, and resulting surplus or deficit on a fair value basis. U.S. companies have been reporting on this basis. The FASB proposes that the most relevant and reliable information available about pension liability or asset is based on the fair value of plan assets and a measure of the present value of the obligation using current, explicit assumptions (Summary, *Employer's Accounting for Pensions*).^[4]

The recent FASB statement as to fair value measurement, SFAS 157, clarifies and incorporates the guidance of using cash flow information and present value techniques to estimate fair value.^[4] In particular, present value techniques are identified as one of the valuation mechanisms by income approach, which converts future amounts (e.g., cash flows) to a single discounted present amount. It also implies that public-sector pension and health-care obligations should be calculated on an economic, rather than an accrual legal basis, as currently required.

Generational accounting is consistent with the present value techniques, as recognized by FASB to estimate fair value. The results, therefore, provide supplementary perceptions with regard to value relevance (herein defined as the association between financial statement measures and public stakeholders' future cash flow expectations) and disclosure implications of present value techniques in predicting unfunded pension liabilities via a generationally accountable perspective. This is in accord with the FASB's objective of including risk in the estimate of fair value, in order to replicate the market's behavior toward assets and liabilities with uncertain cash flows. Since pension assets are not always invested in risk-free assets, assumptions may be needed concerning the underlying economic conditions and monetary assets, must be identified and which have maturity dates or durations that mirror the period covered by the cash flows.

CONCLUSION

The concept of generational accounts links to the need for new reporting models in the public sector for performance reporting. The analysis is expected to raise serious questions over the reliability of (short-term)

performance measures that are now increasingly being used to evaluate the financial management of publicly financed public-sector services. This is because ageing populations imply that severe intergenerational inequities can be involved in PAYG financing of these services, which, in turn, may also underlie calculated cost and profit efficiencies that affect the long-term demand for, and supply of, these services. It is unlikely that these issues can be resolved by simply drawing up fixed short-term contracts that affect service delivery (e.g., agreements with third parties to provide recurrent services or private sector finance to fund major capital expenditures).

The calculations are based on far-reaching assumptions of life expectancy, incomes, economic growth, tax rates, and costs of providing services, inflation rates, and discount rates. Thus, these models provide a basis for judging whether the continuation of current policies should enable each generation to fund the provision of generationally sensitive public-sector services to it or whether their provision depends on future generations contributing to them as well as their own benefits.

Generational accounting is applicable to a range of other public policy settings than has been covered here, such as education- and environmental-related management. Future research is also needed to address the generational equity implications arising from the increasingly complex interactions between public- and private-sector entities in the delivery of many types of public services.

REFERENCES

1. Organisation for Economic Cooperation and Development. *Ageing Populations, Pension Systems and Government Budgets: Simulations for 20 OECD Countries*; OECD: Paris, 1996.
2. Pesando, J.E. Economic effects of private pensions. In *Private Pensions and Public Policy*; OECD: Paris, 1992; 165–182.
3. Bulow, J. What are corporate pension liabilities? *Q. J. Econ.* **1982**, 97: 435–452.
4. Financial Accounting Standards Board. *Employers' Accounting for Defined Benefit Pension and Other Post-retirement Plans*. Statement of Financial Accounting Standards No. 158. Norwalk, 2006.
5. Loeb, S.E. Population aging: some accounting considerations. *J. Acc. Public Policy* **1987**, 6, 1–14.
6. Danzon, P. Can our social insurance systems survive the demographic shifts of the twenty-first century? In *Demography and Retirement: The Twenty-First Century*; Rappaport, A.M., Schieber, S.J., Eds.; Praeger: New York, 1993.
7. Appleby, J. *Financing Health Care in the 1990s*; Open University Press: Ballmoor, Buckingham, 1992.

8. Chua, W.F.; Preston, A. Worrying about accounting in health care. *Acc. Aud. Accountability J.* **1994**, *7* (3), 4–17.
9. Lapsley, I. Constructing health care accountants: melding calculation and care. *Manage. Acc. Res.* **1998**, *8*, 31–35.
10. Lawrence, S.; Alam, M.; Lowe, T. The great experiment: financial management reform in the NZ health sector. *Acc. Aud. Accountability J.* **1994**, *7* (3), 68–96.
11. Treasury. *Whole of Government Financial Reporting*; HMSO: London, 1998.
12. Mayston, D. The private finance initiative in the national health service: an unhealthy development in new public management? *Finan. Accountability Manage.* **1999**, *15* (3/4), 249–274.
13. Barton, A. Public and private sector accounting—the non-identical twins. *Aust. Acc. Rev.* **1999**, *9* (2), 22–31.
14. Auerbach, A.J.; Gokhale, J.; Kotlikoff, L.J. Generational accounts: a meaningful alternative to deficit accounting. In *Tax Policy and the Economy*; Bradford, D., Ed.; MIT Press: Cambridge, 1991.
15. Auerbach, A.J.; Kotlikoff, L.J. Generational accounting: a meaningful way to evaluate fiscal policy. *J. Econ. Perspect.* **1992**, *8* (1), 73–94.
16. Klumpes, P.J.M. Australia's unfunded Medicare: a generational accounting analysis. *Finan. Accountability Manage.* **2000**, *17* (2), 171–190.
17. Klumpes, P.J.M. Determinants of underfunded pension liabilities in the OECD. *Eur. Acc. Rev.* **2003**, *12* (3), 489–513.
18. Tang, L.; Klumpes, P.J.M. Measuring the PAYG pension funding of Chinese listed SOEs: a generational Accounting Approach, mimeo, 2007.

Performance Appraisal

Dennis M. Daley

*Political Science and Public Administration, North Carolina State University,
Raleigh, North Carolina, U.S.A.*

INTRODUCTION

Performance appraisal systems require a specific appraisal instrument; there are no generic appraisal processes. Each performance appraisal process is intricately involved with and dependent on the instrument it employs (as well as with the organizational and individual behaviors entailed). Focusing on the traditional question of what appraisal type or instrument to use (even if other factors are also now considered to be important) remains a central issue of concern among performance appraisal scholars and advocates.

In appraising individual performance, the search for specific performance appraisal instruments runs the gamut from the subjective (essay, graphic rating scale, and checklists) through interpersonal (rankings and forced distribution) to the objective [behaviorally anchored rating scales (BARS) and management by objectives (MBO)]. While history has, more or less, passed the subjective and interpersonal approaches by as meaningful solutions to the appraisal dilemma, reality still sees them employed quite extensively. So, an examination of these types of appraisal not only outlines the intellectual development that has led to the introduction of objective appraisal systems but, rather unfortunately, also is a look at systems—however inadequate they may be—that are still in use.

DESIGNING APPRAISAL SYSTEMS

In order to correctly employ the performance appraisal process, the answers to a series of five questions need to be explored—Why do we appraise?, What do we appraise?, When do we appraise?, Who does the appraising?, and How to appraise? These questions focus on the fundamental elements that make up the performance appraisal process.

Why Do We Appraise?

Performance appraisal is a conscientious effort at formally, rationally, and objectively organizing our assessments of others. In doing so, it is focused on the task of enhancing job relatedness. Eliminating measurement and rating errors and structuring the

decision-making process in order to accomplish this are the dual foci of appraisal research.

The purposes in which performance appraisal can be employed are numerous. However, these can be grouped into two broad categories—judgmental and developmental. While developmental and judgmental appraisals have enhanced productivity as their goal, they approach it in two quite distinctive fashions.

Development focuses on an individuals' potential rather than on their current level of skills and capabilities. Hence, it is essential in such assessments to consider the question of growth for what? Whether viewed from an organizational or individual perspective, the goal toward which this potential is directed needs examination.

Judgmental purposes follow the management systems or command and control model of authority, yet, there are serious questions as to the cognitive validity entailed in judgmental appraisals. Judgmental appraisals are also quite explicitly linked to extrinsic rewards and punishments. In fact, the existence and adequacy of the reward structure is an important subsidiary question with regard to their effectiveness. This has proved an important limitation in their use among public sector agencies.

What Do We Appraise?

Job relatedness is the chief standard by which the acceptability of a performance appraisal measurement is judged. Established through the Uniform Guidelines (1978) in response to the Supreme Court's decision in *Griggs v. Duke Power Company* (1971), the courts have repeatedly reaffirmed and explained the requirement for job relatedness in subsequent rulings (*Brito v. Zia Company*, 1973; *Ramirez v. Hofheinz*, 1980; *Zell v. United States*, 1979). Job relatedness poses a twofold requirement for organizations—criterion must enable supervisors to discriminate between employees solely in terms of their job performance, and the organization must be able to prove or demonstrate the existence of that relationship.

Job relatedness encompasses both internal, task content and external, organizational context. In addition to the traditional, essential responsibilities associated with completing a task, performance appraisals also

measure the social capital networking skills necessary for that task to be integrated into other organizational efforts.

When Do We Appraise?

In most circumstances where an annual cycle is indeed employed, the timing decision is reduced to a choice between the anniversary date and focal-point methods. Under the anniversary-date approach, appraisals coincide with the anniversary of each individual's date of employment. The anniversary-date appraisal cycle spreads the supervisor's workload over the course of the entire year. Given the large number of employees for which an individual supervisor may hold appraisal responsible, this can be advantageous. While the number of appraisals is the same under both approaches, the supervisor's ability to manage them is improved under the anniversary-date approach. With the supervisor no longer facing an avalanche of appraisals, the quality of individual appraisals should be improved.

Who Does the Appraising?

Traditionally, the immediate supervisor is most often the one responsible for appraising subordinates. Supervisory appraisal of subordinates is the predominant method and is widely suggested to occur in approximately 90% of the cases. This is the management individual deemed most knowledgeable as to the employee and the job. In fact, performance appraisal is often viewed as a key management system tool in establishing a supervisor's command and control authority. Contrastively, from a more humanistic perspective, the appraisal process is also viewed as being designed to strengthen the employee-supervisory relationship through the encouragement of mutual understanding.

Yet, there are some potentially interesting answers to this question other than that found in supervisory-based appraisal. The combination of a number of these information sources (especially supervisor, subordinate, peer, and even self-ratings) is the basis for 360° feedback. With its low hierarchy and professional staff, the modern organization is, on the one hand, forced to seek out nonsupervisory sources for appraisal and, on the other, is blessed with highly knowledgeable employees. Hence, 360° feedback promises to provide a more balanced form of appraisal. Of course, the 360° appraisal is designed to serve as a developmental instrument. When it is transformed into a judgmental appraisal, 360° appraisal loses its effectiveness (and often engenders employee distrust).

SUBJECTIVE TECHNIQUES

The use of subjective techniques is not inherently wrong. They are, after all, the foundation upon which the more objective techniques are developed. The problem with subjective appraisals centers is the basic vagueness or the idiosyncrasy, which is their essence. Subjective appraisals are the unique assessment of a specific individual. It is the resultant possibilities for inter-rater differences and errors that create the problem.

The Essay

Essays, graphic rating scales, and checklists are three of the formats that are basically subjective appraisals. The accuracy of the assessments derived when each of these formats is employed may prove exceedingly high. However, this accuracy flows more from the interactive combination of organization and individual rather than from the merits of the specific instrument being employed.

The essay appraisal format is a *tabula rosa*. Supervisors have a blank space on which they are free to write. Essay appraisals (along with the more modern audio or video log equivalents) are descendant from the traditional duty or fitness report. Almost all appraisals, including today's objective techniques, include an essay component.

The Graphic Rating Scale

The subjective graphic rating scale is perhaps still the most pervasive form of performance appraisal. Even though the trend is clearly toward the introduction of more objective systems, the graphic rating scale remains in widespread use. Rating scales remain popular, because they give the illusion of objectivity while involving little in the way of monetary costs.

Basically, a graphic rating scale consists of a set of items addressing personal traits (trustworthy, loyal, helpful, friendly, courteous, kind, obedient, etc.) and job activities (communication skills, sets realistic goals, keeps files and records up to date, adheres to policies and procedures, knowledge of job, etc.). Employee "performance" on these items is then rated using another set of adjective evaluations (poor, acceptable, fair, good, exceptional, etc.), which are invariably linked to a system of numeric scores. This enables the calculation of an average or overall, summary numeric.

Checklists and Forced Choice

Checklist or forced choice appraisals include sets of items linked to the performance of specific jobs; they

also include items for which no established relationships have been previously documented. In conducting a checklist performance, appraisal supervisors are asked to pick from a series of lists of four, those items in each set deemed to be most like and least like an employee. These are then compared against a code sheet, and only those that match validated relationships are tabulated into a Mixed Standards.

These problems can be somewhat alleviated with the use of a mixed standard scale. Mixed standard scales are checklists based on behaviorally anchored items. Specific job-related measures representing good, average, and poor levels of performance are constructed for each job responsibility or task. These items provide an added dimension and depth to the checklists. Mixed standard scales add a degree of objectivity to the format of checklist appraisals. Since supervisors are not made aware of what items “count,” distrust is fostered. Furthermore, this means that supervisors are unable to explain the basis for the employee’s appraisal.

INTERPERSONAL COMPARISON TECHNIQUES

Interpersonal comparisons, such as rankings or forced distributions, may be based on subjective or objective criteria. However, even when initially based solely on objective, job-related evidence, they experience serious shortcomings. Central to all interpersonal appraisal systems is the comparison or assessment of the individual against other individuals, rather than with the specific job to be done.

Ranking

One method for interpersonal comparisons is that of ranking. This is approached in a wholistic manner, wherein an organization’s employees are graded from best to worse. This requires complete knowledge and understanding of the entire organization—purpose and people. While perhaps feasible only in small organizations, a number of gimmicks can be used in order to extend its application to larger units.

Forced Distributions

Forced distributions are another means for making interpersonal comparisons. Grading on the curve is not a new notion and, unlike rankings, can easily be applied to large organizations. However, it is just as prone to error. Forced distributions assume that employee performance fits some external model or distribution, usually envisioned along the lines of something like a normally distributed, bell-shaped

curve. Since employees are not randomly hired, this assumption distorts the actual performance picture.

OBJECTIVE TECHNIQUES

The techniques BARS, competency-based appraisals, and MBO essentially involve the same components but approach them with a slightly different focus in mind. Hence, the objective components, which are common in all three approaches, are introduced into the appraisal process in a somewhat different order.

The BARS appraisals work best with large groups and subgroups of individuals with job descriptions that can be standardized; MBO, on the other hand, is suited such that it can be tailored to each individual job. MBO is best when it is focused on the results to be expected from job performance; BARS handles behavioral processes where outputs are more identifiable and assurable than outcomes. Both employ variations on participative management in order to guarantee their effectiveness. A somewhat more passive approach to participation guides BARS, while a more proactive style is found in MBO. Competency-based appraisal (which may follow a BARS or MBO format) substitutes the measurement of underlying skills, knowledge, and abilities for the behaviors and results that they help produce.

Behaviorally Anchored Rating Scales

The BARS are extensions of the subjective graphic rating scale. They are a clear attempt to translate the graphic rating scale into an objective appraisal system. They address and correct for many of the subjective issues that cloud the validity and inhibit the use of graphic rating scales.

Both the BARS and the MBO approaches emphasize detailed job analyses. Ideally, performance appraisal should be able to work off the same job analysis system used in the development of an organization’s position descriptions and position classification system (and be employed as a guide in the selection process and for designing training programs). Unfortunately, many organizations, especially among those in the public sector, employ different systems of job analysis when it comes to selecting people to perform a job and when it comes to assessing their performance on that job.

The critical incident technique forms the central component in the BARS system. A job analysis, complete with questionnaires and confirmatory desk audits, is conducted on similar positions. Personnel analysts and subject matter experts, i.e., incumbent job holders and supervisors, sift through the resultant

job responsibilities, activities, or behaviors to identify those that are deemed as most critical to its performance. It is assumed that these critical incidents will involve quality and quantity considerations.

MBO Appraisal

The MBO are more focused on results; however, it obviously can also be adapted to situations in which outputs or processes are more involved than outcomes. It originated as a means for managers to translate their strategic plans into implementable programs. It is a basic command and control management system for implementation and monitoring.

It is quite difficult for an MBO approach to assess performance and simultaneously identify potential. While simply having subordinate "acceptance" of performance standards is often considered to be a workable approach, MBO actually assumes that objectives are arrived at through a more actively participative process. Supervisors and employees are envisioned as discussing and negotiating performance standards that are mutually acceptable. Standards are arrived at in an atmosphere of understanding and not in one of imposition. MBO, like the BARS, approach functions better in a more participative environment.

Finally, MBO may overstate the demand for results without focusing on or directly assisting with the means for achieving them. The conditions or resources necessary for successful implementation are assumed to be provided for, automatically. Private sector organizations more readily assume that requisite resources will be forthcoming when goals and objectives are agreed upon than, unfortunately, is often the case in the public sector.

Competency-Based Appraisal

Similar to BARS, competency-based appraisal assesses an employee on the underlying "competencies" (similar to skills, knowledge, and abilities) necessary to complete the behaviors or tasks associated with specific job content. Assessing competencies replaces the measurement of behaviors and results. Since underlying competencies are used in performing many, different behaviors, validating competencies should prove to be overall easier than the more numerous individual behaviors. Since results may be due to the effort of multiple participants, the assignment of individual ratings may prove difficult, whereas the competencies necessary for achieving those results may be more easily measured. Competency-based appraisals that fail to validate the job relatedness of their competencies are simply another form of graphic rating scale.

APPRAISAL ERROR

Performance appraisal is a human process. While the tendency to focus attention on the tools used in the appraisal process can draw attention away from this, it remains the essential aspect of performance evaluation. The development of psychometric accuracy has produced performance appraisal instruments of complex sophistication. Yet, the resultant objective BARS and MBO appraisal systems are only as good as the people who use them. For all their advantages, they are still only tools for aiding us in making our decisions. Rater errors have been extensively treated in the performance appraisal literature.

Organizational Error

Much of what passes under the rubric of rater error is in reality supervisory adjustments to organizational demands. While the impact of these adjustments may be deemed as negative, they are neither accidental nor totally within the control of the supervisor to correct. Goals may be unclear or misunderstood due to communication problems. A hidden agenda may desire to use performance appraisals as means of controlling employees rather than for encouraging productivity. The expectations of what can be done can simply be unrealistic. Finally, results may be due to activity of groups rather than of individuals. The supervisors endeavor to coordinate workers and obtain productivity within this system. As such, the performance appraisal is part of the organization's overall management control system.

Rater Error

Rater errors fall into four categories: 1) job responsibility errors; 2) contrast errors; 3) unidimensional errors; and 4) interpersonal errors. In many cases, these errors are corrected through the employment of objective appraisal instruments. In other instances, more thorough supervisory training is recommended. Even though supervisors strive to objectively evaluate employees, rater errors prey on the weaknesses inherent in the cognitive process. It is this readiness with which an organization can slip into error that calls for the constant monitoring of the appraisal process.

Errors are committed whenever the responsibilities inherent in the job are substituted for a measure of the incumbent's job performance. Similarly, individuals working in a critical unit may benefit from the perceived centrality or significance of their part of the organization.

Contrast errors arise through interpersonal comparisons. Individuals are not assessed on their job performance, but on their performance compared to someone else's performance, or, as is more often the case, someone else's personal traits and characteristics.

Unidimensional errors abound. In these instances, one item dominates the evaluation process to such an extent that other critical factors are ignored. Unidimensional errors can stem from substantive or mechanical concerns.

Interpersonal biases introduce intentional distortions into the appraisal process. The extent to which a supervisor's own performance and career are dependent on a subordinate's performance, the more likely it is that favorable ratings will be awarded. This interdependence creates a mutual need for maintaining a harmonious relationship.

Training individuals in the use of these tools is just as important as the development of objective appraisal techniques. Supervisory training requires care. Supervisory training can encompass organizational and employee considerations as well as those related to the appraisal process. Performance appraisal is part of an overall performance management system. As such, its interaction with the other systemic aspects is just as important as part of its functioning as are the mechanics of the appraisal process.

CONCLUSION

Performance appraisal is the central instrument used for calibrating and monitoring employee behaviors and results. It is used to verify recruitment and selection

techniques and the appropriateness of job analyses. Performance appraisal is used in making judgments regarding pay and promotion and in exercising developmental options such as feedback and training.

While many organizations still rely upon subjective techniques or interpersonal comparisons, these fail to meet legal or objective criterion set or effective appraisal processes. Only through the employment of objective techniques, such as that found in BARS and MBO approaches will the strategic benefits of the performance appraisal process be achieved.

Even objective systems are subject to appraisal error. Organizational and structural errors arising from a misalignment or uncertainty on strategic goals and objectives can lead to distortion. Rater errors originate in cognitive behaviors that evolved for making quick decisions in simpler hunter and gatherer societies. Vigilance and continuous training are required to overcome these limitations.

BIBLIOGRAPHY

1. Daley, D.M. *Performance Appraisal in the Public Sector*; Quorum Westport: CT, 1992.
2. Grote, R. *The Complete Guide to Performance Appraisal*; AMACON: NY, 1996.
3. Milkovich, G.T., Wigdor, A.J., Eds.; *Pay for Performance*; National Academy Press: Washington, DC, 1991.
4. Mohrman, A.M., Jr.; Resnick-West, S.M.; Lawler, E.E., III. *Designing Performance Appraisal Systems*; Jossey Bass: San Francisco, CA, 1989.
5. Murphy, K.R.; Cleveland, J.N. *Understanding Performance Appraisal*; Sage Thousand: Oaks, CA, 1995.

Performance Audits

Brendan F. Burke

Department of Political Science, Bridgewater State College, Bridgewater, Massachusetts, U.S.A.

Wendy Haynes

Bridgewater State College, Bridgewater, Massachusetts, U.S.A.

INTRODUCTION

In general, the auditing function involves “a systematic examination of accounts or programmatic activities, so as to ascertain their accuracy; (or as) a means of verifying the detailed transactions underlying any item in a record.”^[1] Typically, a governing body engages audit professionals to obtain an independent assessment of expenses and programmatic activity to test the alignment between intended goals and actual activities. The most important dimensions—including the professionals who perform the work and their organizational relationship to the entity being reviewed—involve three primary elements: 1) the outside or third party role of auditors, serving between the principals who authorize governmental activity and the agents who carry it out; 2) the adherence to professional standards at every stage of the process; and 3) the system of its focus, wherein the auditor is bound in his or her efforts to an initial plan and a method for carrying the plan through.^[2] The ultimate goal of auditing is almost invariably found in the assessment of the effectiveness, efficiency, and economy of the use of governmental resources, the core values that underpin the executive branch of government, though additional value-based goals have been added in recent years.

This article describes the history of performance auditing and distinguishes between performance and compliance audits before exploring the performance auditing profession more specifically. The Government Accountability Office’s (GAO) standards for the performance auditing profession as well as a typical audit process are described as parts of the consistent nature of the audit profession. Subsequently, the varying components of performance auditing are described, including the collaborative work that specifies performance audit targets and a range of performance audit techniques. The study closes with a comment about the nature and disposition of the successful performance auditor.

HISTORY AND CURRENT CONTEXT OF AUDITING

Although its roots in the accounting profession can still be seen, performance auditing has moved a long

way from the stereotypical “green eyeshades,” furrowed brow, and dusty ledgers of accounting-oriented auditors in earlier times. Auditing may not be the oldest profession, but it has a long history dating back, some would say, to tick marks on Egyptian granaries in 3500 b.c. The notion of accountability forms the fundamental core of all auditing. The need grew out of the separation of a sovereign or owner from his or her riches and the consequent obligation of his or her minions to account for how those riches had been used and increased. No doubt most of the sovereign’s representatives, there to prevent embezzlement, were illiterate; hence, the annual audit was quite literally the “hearing” of accounts.^[3]

Developments around the 18th century—including the French Declaration of the Rights of Man and the American Declaration of Independence—changed the accountability relationships. Whereas the masses had been accountable to the sovereign for the use and propagation of his or her wealth, the sovereign/government was now responsible to the people. Particularly in American democracy, grounded in the separation of powers and checks and balances, the auditing profession thrives. Indeed, the current Comptroller General of the United States and signatory on the 2003 governmental audit standards, David M. Walker, refers to his staff, inspectors general, and state and city auditors as members of the “accountability” profession. Consistent with that philosophy, in July 2004 Walker announced that Congress had acted to change the name of his agency from the General Accounting Office to the Government Accountability Office and underscored the GAO’s key functions of measuring the government’s performance and holding it accountable for results.

FINANCIAL AUDITS VERSUS PERFORMANCE AUDITS

Financial auditing is ubiquitous in government, where most entities must be audited on an annual basis. The requirement may arise from an expression in state statutes or local ordinances, from the requirements of outside grantors of revenue such as the national

government or foundations, or may arise from professional values of accountability to the public. Financial audits, in their most basic form, entail a verification of financial transactions by an outside party. Financial or “compliance” audits can help in answering several lines of inquiry, including: Are records credible and reliable? Has the government been spending its resources in an efficient manner, consistent with legislative guidance, and without unwarranted additional goals included? Performance auditing is less ubiquitous, but still a broad and growing professional activity, as many governments feel the need to submit their physical operations to closer scrutiny based on the values of efficiency and effectiveness. In the government arena, both financial/compliance auditing and performance auditing fall under standards and protocols specified by the U.S. GAO in the Yellow Book, a guide updated periodically and disseminated to national and subnational governments. While there are many commonalities between the two types of audit within the Yellow Book, the two also differ in significant respects. Both share independence as a core value, but they differ in the breadth of the audit and in other technical ways (see Table 1). The accounting profession and practices underpin financial auditing, while performance auditing evolved from the internal controls systems focus of accounting, the recognition of the contribution of the social sciences to assessing program effectiveness, and the increasing importance of holding public officials accountable for government program performance.

THE YELLOW BOOK STANDARDS

The GAO creates and disseminates generally accepted government auditing standards (GAGAS).^[4] While administered by a single national government agency, there has been a movement in recent decades to insert the common framework through legislative and bureaucratic action. The GAGAS—first released in 1972—became the norm for national auditing via the

Inspector General Act of 1978, as amended, 5 U.S.C. App. (2000) and the Chief Financial Officers Act of 1990. State and local governments that receive federal assistance were required to use GAGAS with the passage of the Single Audit Act of 1984, and Office of Management and Budget Circular A-133 added the requirement for higher education institutions and nonprofit agencies that receive federal funds. In this manner, GAGAS became the uniform standard for public audits. The standards have changed significantly from the financial auditing emphasis of the early 1970s to the far more expansive 2003 revisions, reflecting a growing emphasis on performance and accountability, heightened concern with the demonstrable independence of the auditing function, and acknowledgment of the complex needs of 21st century institutions.

The standards fall into the following areas:

1. *Professional proficiency.* The organization performing the audit has certain obligations to provide for proficiency and up-to-date knowledge among its staff members. Ongoing training is as important as entrance skills.
2. *Independence.* The auditor needs to be free of personal, external, and organizational impairments to provide a fair and comprehensive audit. The auditor must not have a current or recent financial or professional relationship with the auditee, or hold any prejudicial attitudes toward the audited entity. The auditor must be free of influence by other agencies or outside forces.
3. *Professional care.* The auditor must match the method to the job, using the best practices available and not performing either more or less than the audit agreement requires.
4. *External and internal quality control.* Internal control is exercised by adherence to the Yellow Book’s professional standards, and also requires an external analysis of compliance with procedures, norms, and standards at least every 3 years.

Table 1 The contrast between financial and performance auditing

Characteristic	Finance auditing	Performance auditing
Point of intervention	Postactivity	Pre- and postactivity
Primary targets	Individuals/accounts	Individuals/programs
Primary mechanism	Rules	Incentives
Role of sanctions	Negative	Positive
Role of management	Supervision/discipline	Goal setting/reinforcement
Role of oversight	Detection/enforcement	Evaluation/benchmarking
Complexity of strategy	Simple	More complex
Durability of effects	Short term	Intermediate

The standards apply to both financial/compliance and performance audits, but there is more flexibility to the procedures and techniques in the performance auditing realm. This is discussed here.

THE PLACE OF JUDGMENT IN THE AUDIT PROCESS: RISK ASSESSMENT AND TECHNIQUE

Fig. 1 illustrates a fairly typical performance audit process. The process proceeds through the following steps: risk assessment/annual planning, preaudit or preliminary survey, planning, field work, reporting, and follow-up. As mentioned previously, two important differences between financial and performance auditing are the breadth of potential and actual coverage of performance auditing activity, and the range of possible techniques for performance auditing, which involves more choice and discretion.

The risk assessment/annual audit planning process (first column) is discussed next. An individual performance audit will typically begin with a preliminary survey of the organizational landscape. In the “survey” phase, the audit team seeks to educate itself quickly as to what is occurring in the program, identify problem areas, and obtain reliable and relevant information for use in planning and performing detailed review work. Information obtained may include program mission, purpose, and objectives, governing laws and regulations, key management reports, reviews conducted by others, and data sources/key metrics for performance measurement. Frequently, the team will interview top officials and staff within and outside the program and may also test a sample of transactions in various management and service delivery systems.

Careful planning is vital to a successful performance audit. During this phase, the audit team leader/manager narrows and defines the scope of the audit and determines audit objectives. She or he writes a work

plan, develops timelines, and determines the mix of resources—expertise, experience, equipment, supervision—that will be needed and assigned to complete the work in accordance with GAGAS. Toward the end of this phase, the audit manager will typically meet leaders of the program to be audited and apprise them of the scope and duration of the review.

During the field work phase, the audit team gathers and analyzes data acquired through various techniques, including file and document review, testing transactions, observing staff in action, interviewing key players, assessing program outputs and outcomes, and much more, depending upon the scope and nature of the audit work plan. Even with adequate planning, the auditors often encounter unanticipated problems and opportunities in the course of the fieldwork. The natural temptation is to continually expand the scope of the audit, and it is especially important during this period that the audit managers exercise good judgment in deciding which new issues should be added to the scope of the audit and which should be cycled back into the annual audit planning process.

Toward the end of the field work and into the reporting phase, the audit team develops findings and recommendations, briefs the auditee on the findings, and reports the results to management and leadership of the audited entity, as well as other interested parties. Other interested parties may include a legislative body, an oversight agency, or the public more generally.

The follow-up phase, too often neglected by performance auditors as a result of limited resources, entails the monitoring of management’s corrective actions in response to the findings and recommendations. Ideally, the results of the follow-up work then become another element to be considered in the annual planning process.

As has been noted previously, the process outlined in Fig. 2 ideally occurs in a broader planning context. The scope of possibilities of auditing governmental performance is as large as the budget or the work force

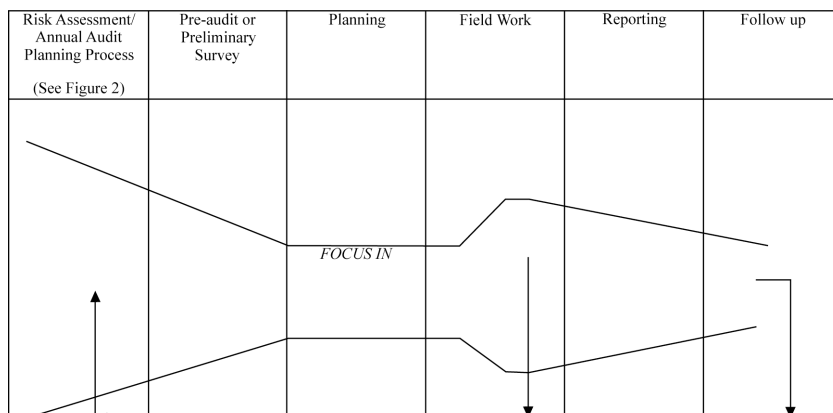


Fig. 1 Performance audit/review process.

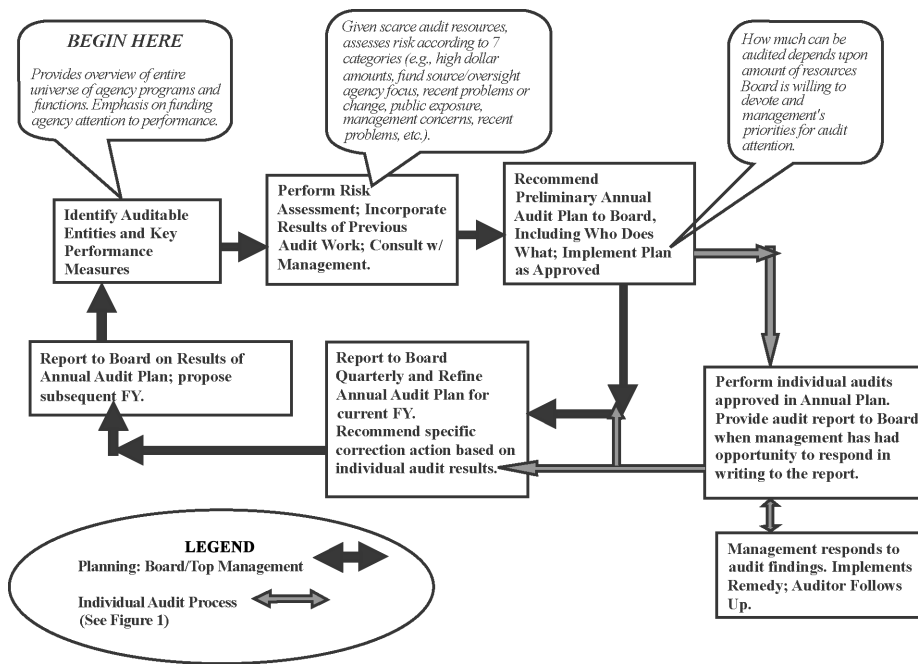


Fig. 2 Process for developing annual performance audit plan.

of the given unit; hence, an important focus of the performance audit function is the determination of what to audit. This relies on a triage function of sorts, based on various risk factors, or questions to be asked after the range of auditable entities have been identified: 1) How large are the dollar amounts of potential audits? Larger is probably a better subject for performance audit. 2) Who oversees the function and acts as the primary funding source? Those who appropriate their own tax revenues are probably the highest priority. 3) Has the program or function had recent problems or major changes? These events probably signal the need for more immediate audit attention. 4) Does the program involve a high degree of public contact? These may also merit more immediate audit work. 5) Has management expressed concerns about the program? Management concerns should be accorded weight in setting the audit agenda. Given limited resources, the auditors must exercise strategic judgment about the focus and scope of their planned audit work, while still preserving enough flexibility to respond to emerging issues.

In keeping with the standards for independence, the audit leadership should retain final discretion in determining the scope and focus of annual audit work, but should also engage management in the process and keep leaders fully apprised of the status and outcome of audit work.

Compared with professionals drawn from other disciplines, one sees relatively fewer traditional auditors (i.e., accountants) engaged in conducting performance audits. Professional guidance on planning and conducting performance audits may be obtained through

certification programs, training sessions, and written guidance from such organizations as the Institute of Internal Auditors and the National Association of Local Government Auditors.^[5] Performance auditing draws on a broad spectrum of professional expertise, including engineering, law, economic analysis, and the full range of social scientific inquiry. This analytical diversity is reflected in the range of possible techniques available to the performance audit team, including:

1. *Process analysis*—The study of the sequence of activities necessary to accomplish governmental work. The allocation of time to portions of the process, proper ordering and potential parallel coordination of activities, geographic or locational alignment between scheduled activities, and assignment of varied expertise and effort throughout the process may determine the overall efficiency and effectiveness of the complex service. Process analysis is a central component of total quality management improvement efforts.
2. *Quasiexperimentation*—Assessment of a program change or intervention over time, in comparison to a control group wherein the status quo without the program change is maintained. If the control and intervention groups are comparable at the outset, the quasiexperiment will display differences in outcomes that can only be driven by the specific different policy treatment between the two groups.
3. *Surveys and focus groups*—Written evaluation or open discussion regarding programmatic characteristics, gained from either direct participants

in service delivery, or clients who receive the service, or both. The use of these participant information gathering techniques will depend on whether the auditor needs answers to narrower questions, whether she is exploring factors related to service delivery, whether the participants require anonymity or not, and other characteristics of the situation.^[6,7]

While the performance audit field is open to a diversity of analytical methods, it is moving toward a common reliance on data related to programmatic performance. Four variable groups increasingly tend to be the analytic focus: Inputs—dollars and human resources for use with government programming; Process—steps within the delivery of governmental services, such as customer contacts, Web pages created, or documents produced; Outputs—the product of governmental effort, such as building inspections, public health nurse visits, or training sessions held; and Outcomes—the desired impacts of government (e.g., crime reduced, birthweight of babies increased, lives saved through safety programs, and the like.^[8,9]) Each type of measure can be the focus of a performance audit, but should ultimately be connected to positive programmatic outcomes. Another core measurement tool and audit procedure involves benchmarking. If a certain service has an accepted standard, or if a government can be compared to a favorable measure in a reputable jurisdiction, the government is working against a benchmark. Many services have been tied to benchmarks, but there is still a long way to go to arrive at accepted standards of measurement for the range of government services.^[10,11]

CONCLUSIONS

This description of the performance audit has listed the components common across compliance and performance variations, as well as the common framework dictated by the GAO's GAGAS or Yellow Book standards. But in addition, the differences within performance audits have been described, including topics to be analyzed and techniques used in the analysis. Ultimately, the performance auditing professional is as much an artist as she or he is a scientist, working from a broad mandate with an abundant toolbox of techniques and broad discretion in approach and focus. The performance auditor is best able to pursue her or his job when she or he "thinks like an auditor," exercising curiosity, independence, broad awareness, clarity, and is convincing in her or his presentation.^[12] With the

current emphasis nationwide on issues of performance and accountability, performance auditing will continue to be a growth industry.

REFERENCES

1. Sharkansky, I. *Audit*. In *Defining Public Administration*; Westwood Press: Boulder, CO, 2000.
2. Herbert, L. *Auditing the Performance of Management*; Wadsworth, Inc.: Belmont, CA, 1979.
3. Dittenhofer, M.A. History of internal auditing in government. Internal Auditing Summer, 1985.
4. General Accounting Office (subsequently changed to the Government Accountability Office). In *Government Auditing Standards*; General Accounting Office: Washington, DC, 2003. Accessible at <http://www.gao.gov>. For state-level equivalents, see, for example, the Louisiana Legislative Auditor (<http://www/lla.state.la.us/>), the Auditor of the Commonwealth of Massachusetts (<http://www/mass.gov/sao/auditor.htm>), or the Auditor General of Alberta, Canada (<http://www.oag.ab.ca/>).
5. For more information, see the websites of the following organizations: Institute of Internal Auditors (www.theiia.org) and the National Association of Local Government Auditors (www.nalga.org).
6. Ammons, D.N. *Tools for Decision Making: A Practical Guide for Local Government*; CQ Press: Washington, DC, 2002.
7. Wholey, J.S., Hatry, H.P., Newcomer, K.E., Eds.; *Handbook of Practical Program Evaluation*; Jossey-Bass: San Francisco, CA, 1994.
8. Hatry, H. *Performance Measurement: Getting Results*; Urban Institute: Washington, DC, 1999.
9. Raaum, R.B.; Morgan, S.L. *Performance Auditing: A Measurement Approach*; Institute for Internal Auditors: Altamonte Springs, FL, 2001.
10. Ammons, D.N. *Municipal Benchmarks: Assessing Local Government Performance and Establishing Community Standards*, 2nd Ed.; Sage Publications: Thousand Oaks, CA, 2001.
11. See the Municipal Performance Measurement Project, Ministry of Municipal Affairs and Housing, Province of Ontario, Canada, for a province-wide example of standard measurement and its development. Accessed at <http://www.mah.gov.on.ca>.
12. Riggs, R.R. How to think like an auditor. *Govt. Accountant's J.* **2000**, 49 (Winter), 10–15.

BIBLIOGRAPHY

1. Light, P.C. *Monitoring Government: Inspectors General and the Search for Accountability*; Brookings Institution: Washington, DC, 1993.

Performance Indicators in State Administration, The Use of

Peter J. Haas

Department of Political Science, San Jose State University,
San Jose, California, U.S.A.

INTRODUCTION

Performance measures are numeric indicators of activity integrated into government budgets and other planning documents to help ensure accountability and establish the link between resources and results, in theory. Many state governments have implemented performance measurement systems.^[1,2] The critical aspect of performance measurement is the selection, definition, and application of appropriate measures of government activities, such that the quantity, efficiency, and effectiveness of service delivery efforts may be accurately documented.

Creation of a performance management system entails much more than merely identifying performance measures. Moreover, under ideal circumstances, performance measurement efforts are integrated into a framework for improving government decisions as part of a *performance management system*. For example, the state of Virginia defines its performance management system as follows:

- **Strategic planning:** Systematic clarification and documentation of what an organization wishes to achieve and how to achieve it.
- **Performance measurement:** Systematic collection and reporting of information that track resources used, work produced, and intended results achieved.
- **Program evaluation:** Systematic collection and analysis of information to determine a program's performance and reasons for achieving the level of performance.
- **Performance budgeting:** Systematic incorporation of performance information (planning, performance measurement, and evaluation information) into the budgetary process (Fig. 1).

Thus, performance measurement is ideally not an end itself, but is linked to other management tools. The following steps are typically involved in the creation of a state strategic planning effort that includes performance measurement:^[2]

1. Identification of the populations served and outcomes sought by the agency
2. Identification of specific outcome indicators by which progress will be measured
3. Examination of the future environment and problems/barriers within which the agency and its programs operate (such as demographic, economic, and social factors and trends)
4. Identification of the latest available baseline values for each outcome indicator
5. Examination of alternatives and practical options for achieving outcomes, including the current service delivery approach
6. Analysis of each strategic option's costs, feasibility, and effect on the outcomes, including estimates of the out-year values for each outcome indicator and the costs included in the plan
7. A process for obtaining input from the legislature, agency customers, agency employees, and interest groups
8. Selection of one or two of the strategic options for review by key state officials in the executive and legislative branches

BENEFITS OF PERFORMANCE MEASUREMENT

Performance measures may have other uses beyond informing resource allocations decisions. Among the other benefits associated with performance measurement are improved managerial accountability, improved relations between administrators and elected officials, and improved relations with citizens.^[3] Liner et al. list the following possible reasons for governments to engage in performance measurement:^[2]

1. To respond to elected officials and the public's demands for accountability.
2. To help formulate and justify budget requests and policy choices.
3. To help in resource allocation decisions.
4. To raise questions as to why outcomes are not meeting expectations and to trigger in-depth examinations of why performance problems (or successes) exist.
5. To help motivate personnel to continuing program improvements.

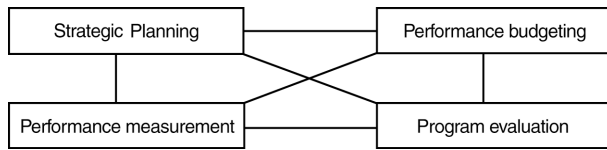


Fig. 1 Relationship between elements of a strategic planning system. *Source:* From <http://www.dpb.state.va.us/VAResult/HomePage/PMOverview.html>.

6. To formulate and monitor the performance of contractors and grantees (performance contracting).
7. To provide data for ad hoc, in-depth program evaluations.
8. To support strategic and other long-term planning efforts (by providing baseline information and subsequent tracking of progress toward long-term goals).
9. To help identify “best practices.”
10. To communicate better with the public and to build public trust.
11. Above all, to help provide better and more efficient services to the public.

CHARACTERISTICS OF PERFORMANCE MEASUREMENT SYSTEMS

Although most share many of the broad characteristics described earlier, state performance measurements systems vary considerably in practice. For example, how broad should the focus of the measurement effort be? Some states have developed statewide goals, benchmarks, or similar reports for the state as a whole, without a particular focus on the more specific roles of state government and state agencies. This approach—typified by Oregon’s Benchmarks, Minnesota’s Milestones, and Florida’s Governmental Accountability to the People Commission—may be more accurately categorized as state quality-of-life indicator systems rather than state-government-focused plans. Even among states that engage in a more applied version of performance measurement, the scope of the effort may differ considerably, ranging from plans that focus on a few agencies providing extensive, well-prepared strategies, outcomes, and performance measures, to less intensive reports from agencies that merely described current activities without much discussion of the future environment, outcomes, or out-year targets.^[2]

A second key characteristic of state performance measurement systems is the extent to which, if any, they are integrated into performance budgeting. A review of state performance measure systems found

that relatively few states actually incorporate performance measures into their budget process.^[2] Without a meaningful linkage to state budget process, performance measurement may become an empty gesture. One factor that apparently limits the extent to which states integrate performance measures in budgeting decisions is the lack of expertise and other resources required to do so. One observer noted that “using data on results to guide government management practices has proved to be successful in states and localities across the nation. Few jurisdictions have, as of yet, succeeded at using results data to shape budgets.”^[4]

A third characteristic is the locus of performance measurement analysis. In some states, agencies bear the primary responsibility for analyzing data, an arrangement that some find to be problematic: “Agency and program managers have had difficulty in systematically analyzing and using outcome data as part of the budget formulation process. Agency personnel often appear to consider outcome information something they are required to provide, rather than as something useful to them in preparing and justifying their budgets.”^[2] States with a strong legislative analysis arm may be more likely to make good use of performance measurement data.

CHALLENGES ASSOCIATED WITH IMPLEMENTING PERFORMANCE MEASUREMENT

A survey of local government managers identified issues that frequently arise with respect to the use of performance measure.^[3] Among the most commonly reported problems were the following:

1. maintaining the currency of measures
2. measuring the quality of programs and services
3. avoiding ambiguity and confusion with measures
4. compilation and distribution of measurement data in a timely manner
5. lack of capability to analyze collected performance data
6. maintaining support from lower-level employees for performance measurement efforts

The last problem is particularly critical, because an ongoing performance measurement system requires considerable effort from all levels of an organization. “Performance measurement systems are more effective in influencing behavior in desired ways when line managers and employees buy into the system and measures.”^[3]

UNINTENDED CONSEQUENCES

Performance measurement is not without difficulties, costs, and potential risks and pitfalls. Among those that have been identified are the following:^[5]

1. Selection of performance indicators with inadequate attention to who will use the data and how. Without a practical focus, indicators will not or cannot be used by anyone.
2. Inadequate preparation among legislators and executive branch managers for application of results-based government. Decision makers must understand how to use performance measurement data.
3. Inadequate linkages between information on funding and agency activities to the desired outcomes. Measuring performance may be relatively simple compared to establishing meaningful links between indicators and budget processes.

Beyond the sheer cost of collecting and analyzing performance data, perhaps the greatest potential risk associated with performance measurement systems is the potential for creating incentives for agencies that are based on instrumental indicators rather than genuine measures of performance. Instrumental indicators are indirect or inaccurate measures that are used in the absence of bona fide performance measures. Typically, instrumental indicators are linked to agency

inputs or efforts rather than outcomes or results of those inputs. For example, the performance of a drug treatment program might be measured in terms of the number of clients who receive treatment, rather than on the outcome of the treatment. If agencies act to maximize their performance as measured by instrumental indicators, they may not actually improve their actual performance in terms of desired outcomes.

Does performance measurement live up to its promise? It is now apparent that it is frequently used to measure the impact of government programs and services. Its potential, therefore, is now widely accepted. The extent to which performance measurement really improves state government is still largely unknown.

REFERENCES

1. Walters, J. Performance measurement: performance and pain. *Governing* June 1997, 10.
2. Liner, B.; Hatry, H.; Vicent, E.; Allen, R.; Dusenbery, P.; Bryant, S.; Snell, R. *Making Results-Based State Government Work*; The Urban Institute, 2001.
3. Poister, T.H.; Streib, G. Performance measurement in municipal government: assessing the state of the practice. *Public Adm. Rev.* July 1999, 59, 325.
4. Walters, J. Performance-driven government: using measures to manage. *Governing*. January 2000, 14.
5. Morley, E.; Bryant, S.P.; Hatry, H.P. *Comparative Performance Measurement*; The Urban Institute Press, 2001.

Performance Measurement

Patria de Lancer Julnes

Department of Political Science, Utah State University, Logan, Utah, U.S.A.

INTRODUCTION

Performance measurement refers to the production of information about an organization's performance (e.g., inputs, outputs, outcomes, efficiency) with regard to services or programs. Performance indicators should be derived from the organization's mission, goals, and objectives.

A key feature of performance measurement that differentiates it from other similar efforts is that it is ongoing. Measurement occurs regularly (e.g., weekly or monthly), and the resulting information is available on a real-time basis for monitoring results by management, elected officials, and citizens. The information can also serve as the basis for modifying plans to enhance performance, whether through the (re)allocation of resources or the development of goals and objectives. Also, it can help to provide justification for budgets and investment decisions, and keep organizations accountable and responsive when reported out. Thus, results of performance measurement are useful to organizations' internal and external stakeholders.^[1,2]

HISTORICAL DEVELOPMENT

The practice and concept of performance measurement are not new to the public sector. Measurement of program activities can be traced back to Taylor's scientific management approach in the early 1900s. Scientific management was so well regarded in the private sector that this approach to measuring inputs and outputs was quickly embraced in the public sector. By the 1930s, performance standards and work measurement had been introduced in several government agencies. Several professional organizations also developed grading systems to assess administrative performance. One of these organizations was the International City-County Management Association (ICMA). In 1938, under the authorship of Simon and Ridley, the organization published *Measuring Municipal Activity*. This publication suggested several types of information that local governments could use to monitor local services and to assess how well these services were being delivered. Included in the publication were measures of needs, costs, results, and efforts.

Other non-profit organizations have also been supporting performance measurement efforts in the public sector. One of the most notable is the Urban Institute, whose contribution began shortly after its founding in 1968, with the development of methods for government and human services to measure the performance of their programs.

Performance measurement has been an important component of numerous management reform efforts to improve performance. Early efforts, between the 1940s and 1970s, such as planning programming budgeting system, management by objective, and zero base budgeting, had measurement at their core. In the 1980s and early 1990s, total quality management, emphasized data collection and process analysis. Today, the measurement of performance on a regular basis is the cornerstone of the Managing for Results effort, a comprehensive approach to focusing an organization on its mission, goals, and objectives. Managing for Results links measures to program mission, sets performance targets, and reports regularly on the achievement of target levels of performance.^[3,4] Further, performance measurement is also the common denominator of the Chief Financial Officers Act, passed in 1990, and the Government Performance Results Act (GPRA), passed in 1993.

TYPES OF PERFORMANCE MEASURES

To understand how performance measurement information is used, it is helpful to distinguish several types of performance measures differing in terms of what is being measured. A good performance measurement system is said to be one that combines the various types of measures or indicators.

Fig. 1 helps to illustrate these different types in the context of an inoculation program conducted at a public health clinic and follows the program logic model. The indicators shown are not exhaustive in that many more could be added, but they provide a simple way to illustrate the various types of performance measures.

Input Measures

As illustrated in Fig. 1, input measures are often devoted to the accounting of resources. These may

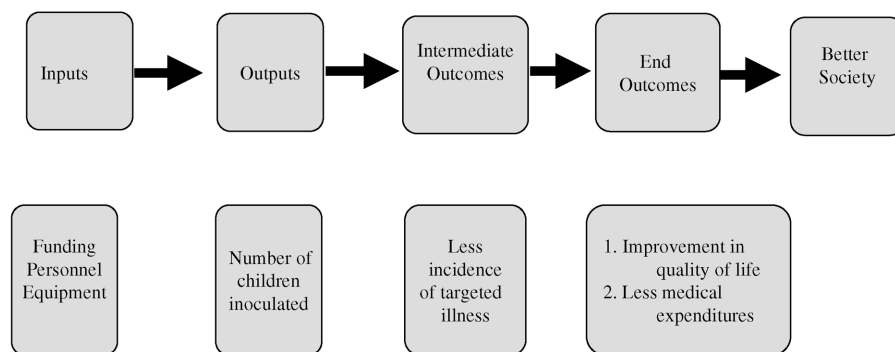


Fig. 1 Example of indicators: an inoculation program.

include dollar costs, staff and staff time, materials, and other resources. Input indicators tell us the amount of effort in delivering services and are used for accountability purposes. The example here includes the amount of funding, number of personnel who will provide the inoculations at the clinic, ancillary personnel, and equipment necessary for service delivery.

Workload or Activity Indicators

Workload or activity indicators deal with issues such as things that need to get done to provide the service (processes), incoming work, and work that has not been completed. They are not useful for public accountability per se, but they can help management track the workflow of their programs.^[5] In some settings, it is difficult to distinguish these measures from output measures. An indicator of workload might be the number of vaccination records filled.

Output Measures

Outputs are the final product or service delivered that ultimately, it is hoped, will lead to a desired outcome. Outputs are indicators of administrative or process effects. In our example, an output is the number of children who received the vaccines. Other outputs for a public health clinic include handing out brochures on the advantages of early inoculation and following up after tuberculosis testing.

Outcome Measures

Earlier attempts to measure performance were often limited to measures of inputs and outputs. However, the growing recognition that producing outputs does not always produce desirable results led to the emphasis on measuring outcomes. Outcomes are the consequences of outputs and are more complex. In our example, the inoculation program would only be

considered effective if it resulted in lower infection rates. Consider a flu vaccination program based on a different strain of the virus than the one that later reached epidemic proportions. In this case, the output did not result in an effective program. Yet, outcomes also include negative consequences. For example, if people become sick as a result of the flu vaccine, this is also an outcome.

In Fig. 1, the box labeled “better society” is included to remind us of the ultimate goal of public programs and policies. Although policy analysts rarely need to concern themselves with defining a better society, it is useful to reconsider even end outcomes from time to time. For example, the reduction of public medical expenses that result from the elimination of childhood illnesses might seem an obvious desired end outcome. However, some research suggests that a complete lack of exposure to illness may have negative consequences for the long-term development of a child’s immune system. Thus, outcomes are not an end in themselves, but a vehicle to reaching even greater values that lead to improvements in society as a whole. For that reason, it is useful to distinguish intermediate outcomes and end outcomes.^[5] Then, following with our example, a lower incidence of the targeted illness is an intermediate outcome; improvements in the quality of life of clients and reduced medical expenses are end outcomes that follow from the lower incidence of illnesses. The final goal is a better society.

Efficiency Measures

Also called productivity measures, efficiency measures indicate the ratio of output to input or outcome to input. When there is an increase in efficiency, it means that there are more gains for a given amount of effort. As is inherent in its definition, efficiency requires knowledge of the amount and type of inputs, the types and number of outputs, or the types and number of outcomes. Thus, in our example, an

efficiency measure can be calculated by dividing the number of children inoculated by the amount of funding (output-to-input ratio), or by dividing the reduction in incidence of infection by the funding (outcome-to-input ratio).

Explanatory Information

The Governmental Accounting Standards Board (GASB) suggested that government entities also include explanatory information when reporting performance. This information, qualitative or quantitative in nature, can help to illustrate the successes of the programs or services or provide information on why the program has not achieved expected results. Thus, it can help to interpret or put the obtained indicators in context. Together with input, output, outcome, and efficiency measures, explanatory information forms part of what the GASB calls service efforts and accomplishment indicators. Going back to the example in Fig. 1, an explanation for success could be that the program used a flexible schedule of services strategy to ensure that the program reached those children in households where parents or guardians worked during traditional business hours. Developing these explanations requires that the appropriate information is recorded and, within practical limits, the more indicators that one is measuring, the better the quality of the interpretations that one can make.

Complications of Measurement

Research shows that input and output indicators are more often developed and used than outcomes and efficiency indicators.^[6,7] Input and outputs are the more traditional types of performance indicators, but as noted above they do not say anything about the desired effect. A greater emphasis is now placed on the need to measure and report outcomes. However, measurement of outcomes is complex in that it is often difficult to distinguish the true consequences of a program from those caused by non-program factors. For example, in the health clinic example, it is difficult to determine how many people would have gotten a particular disease if there had been no vaccination program.

Conversely, it is also difficult to verify that illnesses appearing after the vaccination were caused by the vaccine. One of the major complaints of government employees involved in performance measurement is the difficulty in measuring outcomes because, they argue, it is often not possible to follow up with clients long enough to know whether the service or program had the desired effect. Thus, unlike the straightforward relationship of inputs to outputs, the relationship of inputs to outcomes is less clear and thus superficial

analysis runs the risk of overlooking the possibility that the observed relationship is in fact spurious. This has led to a debate as to whether this kind of causal analysis is appropriate from ongoing performance measurement or should be conducted in a full-blown program evaluation. This is elaborated in the next section.

Furthermore, obtaining a single measure of overall efficiency, even with ideal indicators of outcomes, can also prove to be difficult. This is particularly the case when dealing with multiple inputs and outputs or outcomes with different units of measurement. A suggested approach to overcome these difficulties is to use a technique such as data envelopment analysis, a linear programming technique.^[8] However, this solution is only appropriate if the organization has the technical capacity to analyze performance data.

CURRENT INITIATIVES IN SUPPORT OF PERFORMANCE MEASUREMENT

Many initiatives have contributed to the current emphasis on performance measurement. Two that warrant additional attention are GPRA and the American Society for Public Administration's (ASPA) Center for Accountability and Performance (CAP).

The Advent of GPRA

As noted above, since the early 1990s, performance measurement attained greater visibility at the federal level with the passage in 1993 of the GPRA. This act, the touchstone of then Vice President Gore's National Performance Review initiative, was designed to improve the effectiveness of federal programs and citizen satisfaction through the systematic measurement and reporting of performance.

Under GPRA, federal agencies are required to develop quantitative performance measures and targets, and to report these indicators to help monitor their performance.^[9,10] These federal requirements are having collateral effects, in that performance measures are increasingly being expected at all levels of government. Current research reports an increasing number of public organizations indicating that they have some type of performance measurement system.^[7,11]

CAP

In 1996, an ASPA task force recommended that CAP be created as a standing committee, or as a center to address the extensive emphasis and requirement for all levels of government to move to performance-based,

results-driven management. To that end, CAP's mission is to improve the practice of public service by helping ASPA members to acquire the knowledge, technical skills, and resources necessary to successfully Manage for Results.

In support of its mission, CAP provides numerous training and information resources. These include a performance measurement workbook, *Performance Measurement: Concepts and Techniques*, and case studies prepared by local, state, and federal government practitioners, describing their application of performance measurement and management, strategic planning, and Managing for Results.

Other important resources are the monthly CAP corner in ASPA's *PA Times*; a three-part video series, *Managing for Results: The Key to More Responsive Government*, which provides insights from national leaders on implementing results-oriented government; a multimedia CD-ROM training program, *Applying Performance Measurement*, cosponsored with the ICMA, the Urban Institute, and Public Technology, Inc.; a monthly Brown Bag Lunch on Performance Measurement and Evaluation, which is cosponsored with the U.S. General Accounting Office and George Washington University; and a web site (<http://aspanet.org/cap/index.html>) with extensive web links to performance measurement-related web sites and major best practices web sites.

CONTEMPORARY ISSUES

Utilization of Performance Measurement Information

Much research on performance measurement shows that in spite of its long lineage, development of systematic performance measurement systems is not as prevalent as one might expect. Furthermore, even organizations that have developed performance measures do not implement them in the sense of using the information in decision making.

With this lack of utilization, a key concern has been understanding the factors that affect it. Some identified lack of training, unavailability of data, resources, and other rational and technical issues, whereas others argued that political factors such as lack of support from elected officials and management inhibit utilization. Yet, others focused on the lack of an organizational culture that supports measurement and use. This existing controversy over such factors can be reconciled if we recognize that there are different stages of utilization.^[6,12]

First, there is the issue of adoption or development of performance measures. Organizations in this stage

of utilization have developed, what is called in the knowledge utilization literature, a capacity to act.^[13] This means that even though the knowledge has been produced, it can remain dormant. Research suggests that whether public agencies develop a performance measurement system is a function primarily of rational/technocratic factors, such as those mentioned earlier.

Second is the implementation stage or converting the knowledge into action, actually using the information. A complication at this stage is that, as the organization moves from adoption to implementation, more political risks and culture barriers emerge. Further, complicating assessment of implementation is that use can occur in various ways. Many who promoted performance measures envisioned instrumental use, in which the information guided a decision. However, as evidenced in evaluation theory,^[14] the actual use of performance measurement information can range from very subtle or for enlightenment to very concrete or instrumental.^[12] An example of enlightenment would be when performance measures are used for monitoring purposes, thus allowing for reflection about the program in question. An instrumental implementation would be the case for immediate resource allocation decisions, something that according to recent research does not happen often.

Distinguishing these two stages has useful implications for improving utilization of performance measures. It appears, however, that long-term success of performance measurement has an additional requirement—that the use of performance measures be sustained long enough to allow for an accommodation between the performance measurement system and the organization. This implies a need for a long-term commitment to continue to develop, implement, and refine performance measures.

Citizens' Participation

A greater emphasis is being placed today on involving citizens in performance measurement and performance improvement efforts. One of the arguments for this involvement is that it can help to (re)build public trust in government and lead to long-term sustainability of performance measurement.^[15,16] Furthermore, citizens' involvement in performance measurement can lead to a more impartial, objective, and useful measurement system, and can create greater accountability among policy makers.^[17] Involving citizens, however, requires overcoming practical concerns. These include the need to invest time and money, ensuring diversity so that the clients the programs serve are properly represented, and ensuring democratic participation.

Performance Measurement and Program Evaluation

It was noted earlier that performance measurement may be limited in its ability to develop outcomes. Related to that is the current debate questioning the value of performance measurement vis-à-vis program evaluation in promoting productivity improvement efforts.

One can begin to consider this question by highlighting the differences between performance measurement and program evaluation. The main difference between them is that the latter answers the “why” questions, and the former answers the “what” questions. Also, program evaluation tends to be a one-time (or a not-too-often) activity, whereas performance measurement is ongoing. Thus, they both have different purposes.

Performance measurement is best able to help with general oversight on a day-to-day basis, and program evaluation is particularly suited to providing a more indepth view on how to improve operations.^[3,18] When properly done, performance measurement can provide data for program evaluation, and program evaluation can provide information to help refine performance measurement. Accordingly, performance measurement and program evaluation can complement and strengthen each other and need not compete.

REFERENCES

1. Hatry, H.P.; Fountain, J.R., Jr.; Sullivan, J.M.; Kremer, L. *Service Efforts and Accomplishments Reporting: Its Time Has Come*; Governmental Accounting Standards Board (GASB): Connecticut, 1990.
2. Colwell, W.L.; Koletar, J.W. Performance measurement for criminal justice: the Federal Bureau of Investigation (FBI) experience. *Public Prod. Rev.* **1984**, *VIII* (3), 207–224.
3. Newcomer, K., Ed.; *Using Performance Measurement to Improve Programs*; NDE, Jossey-Bass: San Francisco, 1997; (75), 5–14.
4. Governmental Accounting Standards Board, Performance Measurement for Government. <http://accounting.rutgers.edu/raw/seagov/pmg/perfmeasures/index.html> (access August 2001).
5. Hatry, H. *Performance Measurement. Getting Results*; The Urban Institute: Washington, DC, 1999.
6. de Lancer Julnes, P.; Holzer, M. Promoting utilization of performance measures in public organizations: an empirical study of factors affecting adoption and implementation. *Public Adm. Rev.* **2001**, *61* (6), 693–708.
7. Governmental Accounting Standards Board (GASB). *Report on Survey of State and Local Government Use and Reporting of Performance Measure. Questionnaire Results*; GASB: Norwalk, CT, 1997.
8. de Lancer Julnes, P. Decision-making tools for public productivity improvement: a comparison of DEA to cost-benefit and regression analyses. *Public Budg. Account. Financ. Manage.* **2000**, *12* (4), 625–646.
9. Berman, E.M. *Productivity in Public and Nonprofit Organizations. Strategies and Techniques*; Sage: Thousand Oaks, 1998.
10. Kettl, D. *Reinventing Government? Appraising the National Performance Review*; Brookings Institution: Washington, DC, 1994.
11. Berman, E.; Wang, X. Performance measurement in U.S. counties: capacity for reform. *Public Adm. Rev.* **2000**, *60* (5), 409–420.
12. de Lancer Julnes, P. The utilization of performance measurement information: adopting, implementing and sustaining. In *Public Productivity Handbook*, 2nd Ed.; Holzer, M., Seok-Hwan, L., Eds.; Marcel Dekker, Inc.: New York, 2004; pp. 353–376.
13. Stehr, N. *Practical Knowledge. Applying the Social Sciences*; Sage: London, 1992.
14. Weiss, C.H. *Evaluation: Methods for Studying Programs and Policies*; Prentice Hall: Upper Saddle River, NJ, 1997.
15. Epstein, P.; Wray, L.; Marshall, M.; Grifel, S. *Engaging Citizens in Achieving Results that matter: A Model for Effective 21st Century Governance*; Symposium of the Center for Accountability and Performance: Washington, DC, Feb 11–12, 2000.
16. Weidner, M.; Noss-Reavely, M. *The Council on Human Investment: Performance Governance in Iowa*; ASPA's Center for Accountability and Performance: Washington, DC, 1996.
17. Callahan, K.; Holzer, M. Results-oriented government: citizen involvement in performance measurement. In *Performance and Quality Measurement in Government. Issues and Experiences*; Halachmi, A., Ed.; Chatelaine Press: Burke, VA, 1999; 51–64.
18. Mark, M.; Henry, G.T.; Julnes, G. *Evaluation an Integrated Framework for Understanding Guiding and Improving Policy and Programs*; Jossey-Bass: San Francisco, 2000.

Performance Measurement and Benchmarking in Local Government

David N. Ammons

*Institute of Government, University of North Carolina at Chapel Hill,
Chapel Hill, North Carolina, U.S.A.*

INTRODUCTION

For more than half a century, local governments have been encouraged to measure their performance.^[1] Today, leading professional associations in local government promote the practice.^[2,3] But what is performance measurement? What are the principal types or categories of measures? How does performance measurement contribute to local government administration? What distinguishes performance measurement from the related endeavor called *benchmarking*?

PERFORMANCE MEASUREMENT

Performance measurement is the art of gauging systematically the quantity, quality, efficiency, and effectiveness of goods and services produced by a department, program, or activity. Ideally, a set of performance measures not only will reveal *how much* work has been performed or *how many* units of service have been delivered, but will also report *how efficiently* the services were produced, their *quality*, and their *impact* on service recipients or on the community in general.

Performance measurement is extolled by its proponents—first, for its value as a method for ensuring the accountability of a local government and, second, for its usefulness as a tool supporting a host of managerial functions. How can a local government consider itself to be meeting its accountability obligations, these proponents ask, if the government's reporting focuses solely on financial stewardship and the avoidance of improprieties with public funds, but fails to provide meaningful evidence that commensurate services were secured with these resources?^[4] Although it is a crucial element of accountability, financial stewardship addresses only one part of the accountability question. Information about what the citizens receive in return for their tax dollars addresses the other part.

WHY MEASURE PERFORMANCE?

A sound performance measurement and reporting system enhances the accountability of a local government.

Frontline employees and their supervisors become more accountable for performing a fair day's work; departments and programs become more accountable for achieving their objectives; and the local government as a whole becomes more accountable for delivering a good return to the tax-paying public. Some local governments stop there, emphasizing the importance of performance measurement for accountability but using their system for little other than reporting on programs and services. Others go further—much further. They, too, use performance measures for various reporting purposes, but they also rely on their performance measurement and monitoring system to support a host of other management functions.^[5–8] These other functions include planning and budgeting, program evaluation, performance appraisal, reallocation of resources, directing operations, contract monitoring, the identification of deficiencies and opportunities for operational improvement, and the testing of strategies and processes designed to effect performance improvement.^[9] Categorizing the multiple purposes of performance measurement in a different fashion, performance measures may be used to evaluate, control, budget, motivate, promote, celebrate, learn, and improve.^[10]

As noted by Hatry, “Unless you are keeping score, it is difficult to know whether you are winning or losing. This applies to ball games, card games, and no less to government productivity.”^[11] When fully developed and applied, performance measurement fulfills this vital scorekeeping role for local governments, not only alerting local officials to deteriorating conditions or service delivery problems, but also informing them of general operating strengths and weaknesses and the degree to which various procedural modifications or program strategies are achieving their intended results.

TYPES OF PERFORMANCE MEASURES

Although attempts at categorizing performance measures occasionally have produced lists of 8 or 10 different types, many of these categories may easily be

combined to form a simpler set of four basic categories of performance measures in local government:

- Workload (output) measures.
- Efficiency measures.
- Outcome (effectiveness) measures.
- Productivity measures.^[9]

The first of these categories, workload measures (also called output measures), is well represented in city and county budgets and other performance-reporting documents. Examples of workload measures include “applications received,” “properties appraised,” “meters read,” and “tons of refuse collected.” Unfortunately, many of the local governments that report raw workload measures begin and end their performance measurement effort with this category. This is unfortunate because workload measures have the least managerial value of the four types of measures. Raw workload counts tracked over time reveal patterns of demand for a given service and reveal the scale (and perhaps even the “busy-ness”) of an operation, but they yield few managerially significant insights.

Efficiency measures relate the outputs of a program or activity to the amount of resources required to produce these outputs. Unit costs (e.g., cost per application processed) are efficiency measures, as are the many variations that link outputs and resources in other ways (e.g., applications processed per \$1000). Some variations replace dollars with another form of resource input, such as labor-hours or staff-hours, yielding measures such as applications processed per staff-hour.

Outcome (effectiveness) measures reveal the extent to which a program is achieving its objectives. Measures of service quality are often included in this category. Rates of success in limiting fire spread and recovering stolen property in the fire and police services, respectively, are examples of effectiveness measures.

Productivity measures gauge the efficiency and effectiveness of an activity by means of a single indicator, sometimes in the form of an index. Such measures may relate total costs to program successes (i.e., total costs divided by successful interventions rather than by all intervention attempts, regardless of whether successful or not) or may report production ratios that have been reduced to account for any defective work (a common practice in mechanic productivity calculations). Productivity measures have also been called “outcome-based efficiency ratios.”^[5]

Program officials and supervisors attempting to “steer” their operations—to borrow a term from the movement to “reinvent government”—can benefit from a good set of performance measures.^[8] A good set of fleet maintenance measures, for example, can help officials improve performance in the local government garage not only by keeping them apprised of the

condition of the fleet, the quality and efficiency of repair work, and the availability of vehicles, but also by tracking improved results or conditions that confirm the value of maintenance practices that are working well and by alerting operating supervisors and managers if performance begins to fall short of the intended mark.

It is important to remember that the various measurement categories are not of equal managerial value. Workload measures rest at the bottom of this scale. Managers who discover, for example, that the rate of on-schedule refuse collection service (effectiveness) has declined or that the cost for this service per household (efficiency) has soared will be prompted by this discovery to investigate further and to devise strategies to return the service to its previous condition or to an even more favorable performance level. In other words, these measures of efficiency and effectiveness will inspire managerial thinking.^[12] No such managerial thoughts are likely to be inspired by raw workload measures. Reporting the number of tons collected or households served is unlikely to prompt the manager to examine any aspect of the operation or disturb the status quo.

CRITERIA FOR A GOOD SET OF PERFORMANCE MEASURES

A good set of performance measures includes measures that are valid and reliable, understandable, timely, resistant to undesired behavior, comprehensive, non-redundant, sensitive to data collection costs, and focused on important facets of performance.^[5,7,13,14]

- *Valid*: High scores on a set of measures should be a reflection of true excellence on important qualities or dimensions of service.
- *Reliable*: Measures that are consistently accurate and exhibit little variation attributable to subjectivity or measurement by different raters are said to be reliable.
- *Understandable*: The meaning of each measure should be unmistakably clear.
- *Timely*: Measures that are fresh have greater value to operating managers and policy makers.
- *Resistant to undesired behavior*: An emphasis on performance results could tempt employees to take extreme or otherwise undesired steps to attain good marks. It is important to have well-designed measures that are unlikely to prompt such steps or that address performance from several dimensions simultaneously, thereby discouraging actions that would lead to good marks on only one dimension while undermining others.
- *Comprehensive*: A good set of measures addresses all the major facets of an organization’s performance.

- *Non-redundant*: Each measure should add a distinct, unduplicated element of performance information to the set.
- *Sensitive to data collection costs*: Most dimensions of service may be measured in several ways—some direct and some indirect or involving proxies. When measurement of a given function presents serious challenges, various options should be considered carefully, weighing the cost required to collect the measure against the potential value of such information.
- *Focused on important facets of performance*: A good set of measures will address the most important dimensions of a program or department's performance.

Not all sets of performance measures achieve these criteria. The best ones do or at least come close.

BENCHMARKING

Among public sector managers, the term *benchmarking* may be used to refer to any of three related management practices. Each practice features the use of a known or fixed point of reference to gauge current conditions or the status of performance, to inspire improvement, and to track progress toward a desired condition or performance level. In the most ambitious applications, benchmarking also entails the analysis of selected operating processes, to identify variations that account for superior performance among those deemed the “best in the business” for purposes of adapting “best practices.”

Benchmarking is a term borrowed from the vocabulary of land surveyors. Once surveyors pinpoint a spot of known location and elevation, they have a point of reference—that is, a benchmark—for other measurements. Similarly, managers can use performance benchmarks as points of reference to guide their efforts to improve program performance. Potential choices as performance benchmarks include relevant performance standards, statewide or national norms, and the performance targets or results achieved by respected counterparts.

Benchmarking takes three forms in the public sector:

- Corporate-style benchmarking.
- Use of benchmarks in “visioning” initiatives.
- Use of performance statistics as benchmarks.

Each form differs from the others. The common denominator is the use of points of reference called benchmarks in all three.

Corporate-Style Benchmarking

Derived from the practices of Xerox, Motorola, Digital Equipment Corporation, and other pioneers of benchmarking in the private sector, this form of benchmarking is more analytical than the others and narrower in that it focuses on a single process within an operation. This is the classic search for “best practices,” in which a government adopting this form of benchmarking would identify organizations regarded as outstanding performers of the selected process, analyze the process in place at these top performers to detect the reasons for their outstanding results, and adapt these “best practices” for use by the benchmarking government to improve its own performance. This form of benchmarking has been defined as follows:

Put quite simply, benchmarking is the art of finding out—in a completely straightforward and open way—how others go about organizing and implementing the same things you do or that you plan to do. The idea is not simply to compare your efficiency with others but rather to find out what exact process, procedures, or technological applications produced better results. And when you find something better, to use or copy it—or even improve upon it still further.^[15]

The essence of corporate-style benchmarking has been identified most simply as four steps: plan; collect; analyze; and improve.^[16] A slightly more detailed formulation designed in the late 1990s for implementation in the public sector identifies corporate-style benchmarking as a seven-step process: 1) decide what to benchmark; 2) study the processes in your own organization; 3) identify benchmarking partners; 4) gather information; 5) analyze; 6) implement for effect; and 7) monitor results, taking further action as needed.^[17,18]

Use of Benchmarks in Visioning Initiatives

From time to time, some communities undertake planning initiatives to establish a vision for their future. When they do so, they typically articulate desired conditions and sometimes quantify these conditions in the form of benchmarks. Organizers of such efforts usually enlist broad community support, including local government, and measure their success in terms of progress toward the benchmarks.

This form of benchmarking has garnered more media attention than the other public sector forms, although most of the attention has been directed toward state-level visioning initiatives such as the Oregon Benchmarks.^[19,20] Community undertakings of a similar nature, such as the “Life in Jacksonville (Florida)” project, typically focus on local socioeconomic

conditions and quality of life rather than the mechanics of local government service delivery.

Use of Performance Statistics as Benchmarks

Local government officials who compare their own organization's performance statistics with national standards, norms, or the performance statistics of other organizations providing similar services are practicing yet another form of benchmarking. Benchmarks here are the standards, norms, or performance levels shown to be achievable by other programs or service producers.

This is probably the most common form of benchmarking in the public sector, even if it is less publicized than the version associated with visioning initiatives. Although comparison of performance statistics involves less depth of analysis than corporate-style benchmarking and is less likely to lead directly to the identification of "best practices," this form nevertheless enjoys the benefit of greater breadth as a diagnostic tool. Comparisons are rarely limited to a single process or a single component of a given department, as are corporate-style benchmarking projects, and can often reveal operating strengths and weaknesses more broadly, sometimes as a prelude to more detailed analysis.

CONCLUSION

Performance measurement is an important tool not only for fulfilling a local government's obligations for accountability, but also for guiding its efforts to achieve successful programs and services that are efficient and effective. Benchmarking is an advanced application of performance measurement that takes three forms in local government, each designed in its own way to lead to service improvements or to enhanced community conditions.

REFERENCES

1. Ridley, C.E.; Simon, H.A. *Measuring Municipal Activities: A Survey of Suggested Criteria for Appraising Administration*; International City Managers' Association: Chicago, 1943.
2. Government Finance Officers Association. *Recommended Practices for State and Local Governments*; GFOA: Chicago, 2000.
3. National Advisory Council on State and Local Budgeting. *Recommended Budget Practices: A Framework for Improved State and Local Government Budgeting*; GFOA: Chicago, 1998.

4. Governmental Accounting Standards Board. *Preliminary Views of the Governmental Accounting Standards Board on Concepts Related to Service Efforts and Accomplishments Reporting*; GASB: Norwalk, CT, 1992.
5. Hatry, H.P. *Performance Measurement: Getting Results*; Urban Institute Press: Washington, DC, 1999; 60 pp.
6. Poister, T.H.; Streib, G. Performance measurement in municipal government: assessing the state of the practice. *Public Admin. Rev.* **1999**, *59*, 325–335.
7. Broom, C.; Jackson, M.; Vogelsang Coombs, V.; Harris, J. *Performance Measurement: Concepts and Techniques*; American Society for Public Administration: Washington, DC, 1998.
8. Osborne, D.; Gaebler, T. *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*; Addison-Wesley Publishing Company: Reading, MA, 1992; 156 pp.
9. Ammons, D.N. *Municipal Benchmarks: Assessing Local Performance and Establishing Community Standards*; Sage Publications: Thousand Oaks, CA, 2001; 11–13.
10. Behn, R.D. Why measure performance? Different purposes require different measures. *Public Admin. Rev.* **2003**, *63*, 586–606.
11. Hatry, H.P. The status of productivity measurement in the public sector. *Public Admin. Rev.* **1978**, *38*, 28.
12. Ammons, D.N. Performance measurement and managerial thinking. *Public Perform. Manage. Rev.* **2002**, *25*, 344–347.
13. Hatry, H.P.; Blair, L.H.; Fisk, D.M.; Greiner, J.M.; Hall, J.R., Jr.; Schaenman, P.S. *How Effective Are Your Community Services? Procedures for Measuring Their Quality*; Urban Institute and International City/County Management Association: Washington, DC, 1992; 2–3.
14. Bens, C.K. *Strategies for Implementing Performance Measurement*; Series: ICMA Management Information Service Report; International City/County Management Association: Washington, DC, November 1986; Vol. 18, 1–14.
15. Harris, B. Best practices emerge from the synergy of technology, processes, and people. *Emerg. Technol. Suppl. Gov. Technol.* **1995**, *8*, 16.
16. Kinni, T.B. Measuring up: benchmarking can be critical, but it doesn't have to be expensive. *Ind. Week* **2005**, *243* (22), 27–28.
17. Southern Growth Policies Board & Southern Consortium of University Public Service Organizations. *Benchmarking Best Practices, Module 2 of Results-Oriented Government*; Southern Growth Policies Board Research: Triangle Park, NC, 1997.
18. Keehley, P.; Medlin, S.; MacBride, S.; Longmire, L. *Benchmarking for Best Practices in the Public Sector*; Jossey-Bass: San Francisco, CA, 1997.
19. Walters, J. The benchmarking craze. *Governing* **1994**, *7* (7), 33–37.
20. Oregon Progress Board. *Achieving the Oregon Shines Vision: The 2005 Benchmark Performance Report*; Oregon Progress Board: Salem, OR, 2005.

Performance Measurement: Citizen-Driven

Kathe Callahan

*School of Public Affairs and Administration, Rutgers University,
Newark, New Jersey, U.S.A.*

INTRODUCTION

Performance measurement in the public sector is the systematic and continuous assessment of how well services and, in some instances, products are being delivered. A good performance measurement system can increase the knowledge about public sector service delivery and, therefore, enhance the decision-making process of elected officials and public managers. It can also strengthen accountability systems in that the public has an increased understanding of what they are getting, or not getting, for their tax dollar. Performance measurement results in information that can be used to inform decision making and inspire managerial thinking. Performance indicators typically measure the quantity, quality, efficiency, productivity, and outcomes of public services.^[1]

Citizen-driven performance measurement involves the use of measures, or indicators, that are socially relevant to citizens. By socially relevant, we mean measures that are understandable to the average citizen and inform them about the quality of life in their community and the quality of service delivered by a government agency. Socially relevant performance measures provide citizens with comprehensible data that enable them to hold government officials accountable for performance and enable them to answer the question, “Am I getting value for my tax dollar?” So, in addition to collecting data on the tonnage of refuse collected in a quarter, a public works department might also collect data on street cleanliness. The cleanliness rating of a street means more to the average citizen than the overall amount of garbage collected. Citizens, elected officials, the media, and other interested stakeholders can use relevant data, such as street cleanliness ratings, to hold public administrators accountable for their performance. Data such as this can also be used as a benchmark to improve service over time and foster competition among service providers. Socially relevant performance measures equip citizens with the information necessary to ensure accountability and to ensure that governments and public agencies do what they are supposed to do, ultimately improving the quality of people’s lives.

Ideally, citizen-driven performance measurement entails citizens being involved in the establishment of

appropriate performance indicators. If citizens are not directly involved, public administrators, at the very least, should consider the value that the performance measurement system, or specific performance indicators, has for citizens and should willingly share performance data with the public in a timely and understandable way.

ACCOUNTABILITY AND PERFORMANCE

Until recently, government accountability was largely a matter of financial accounting, hence its name. When public funds were appropriated, the key accountability questions focused on how much money was spent and on what items—personnel, supplies, travel, or communications. Today, the concept of governmental accountability has taken on a much broader meaning. According to Romzek, Dubnick, and Shafritz, accountability is a relationship in which an individual or agency is held to answer for performance that involves some delegation of authority to act.^[2] Accountability mechanisms are the means established for determining whether the delegated tasks have been performed in a satisfactory manner.

Thus, governments, in an effort to demonstrate accountability, must be able to measure what they do. Specifically, they need to show their constituents:

1. What they are getting for their tax dollars, in terms of services—a police presence, public hospitals, roads, airports, libraries, public education, and water supplies.
2. How efficiently and effectively their tax dollars are spent—cost per mile of roads built to certain standards and cost per pupil of students graduating from high school.
3. How such expenditures benefit their lives or the lives of those they care about—eradication or suppression of diseases such as polio and acquired immune deficiency syndrome (AIDS), high sense of security in a neighborhood, and safe and reliable water supplies.

This type of accountability holds government responsible not only for its accounts but also for the

results of its actions—the impact of each service on each citizen.

To emphasize to constituents what they are getting for their tax dollars, governments need to be able to measure and report what they are accomplishing. Measures that are unreported, suppressed, or poorly reported are virtually useless in improving performance. Well-reported (i.e., easy to understand and widely distributed) performance measures emphasize how well a government is doing at meeting its constituents' needs and at living up to its commitment to provide quality services for the taxes people pay. Performance measures can help elected officials, public managers, and citizens understand the impact various expenditures have on the quality of their lives and the lives of those they care about. Performance measures also enable elected officials, public managers, and citizens to see the outcomes, or results, of specific government services and to place a relative dollar value on each service provided. Measures of progress have the added benefit of breaking through negative stereotypes to build lasting support for public services.

Measurement of performance has always been implicit in questions of outputs and outcomes: Is crime up? Are the streets cleaner? Is the air quality better? How well are our children doing in school? In short, is the program producing as promised? What are the results? Performance measures provide an opportunity to answer such questions with hard data, rather than the more common situation-subjective statements based on vague assessments of efficacy: “the neighborhood seems safer,” “litter here is worse than ever,” or “the schools in this town are pretty good.”

It is helpful to think of performance measurement as a process, or system, of measures and procedures, whereby organizations assess how well they are doing compared with previous performance, compared with other organizations (benchmarking), compared with national standards, and in terms of how well they are achieving their stated goals and objectives. A well-designed performance measurement system should clearly articulate service goals and objectives, define service outputs and outcomes, and specify the expected quality levels for these outputs and outcomes.

To measure something—IQ, weight, speed, attitude, blood pressure, rushing yards per game, and on base percentage—you need to quantify what it is you want to measure by using a defined set of rules. Obviously, some things are more difficult to measure because they cannot be observed directly or there is no agreed-upon standard on which to base the measurement. We know that to run the mile in under 4 min is an exceptional performance, and we know that a blood pressure reading of 160/100 is something to be concerned about. We know this because these measures have an accepted standard on which to compare other

measures. Public sector services are often more difficult to measure because, for mission-driven organizations, there is, more often than not, no agreed-upon standard of performance and no bottom line. Not driven by profits, or stock value, public managers operate in a world of multiple, often-competing, and often-vague expectations of what good performance looks like. Yes, it is more difficult to measure the performance of the Department of Homeland Security than it is to measure the performance of Coca-Cola or Nike, but it can and should be done.

Performance measures can be quantitative (average response time) and qualitative (how safe people feel in their neighborhood). Managers can develop performance measurement systems incorporating various performance indicators, such as those defined by the Government Accounting Standards Board (GASB).^[3]

- *Input indicators:* These measure the resources expended on a program, such as the amount of money spent or the total number of employee hours needed to deliver a service.
- *Output indicators:* These report the quantity of products or units of service provided to a service population. They also include “workload” measures that reflect the amount of effort expended to produce a product or provide a service. Examples of output indicators include number of meals delivered, miles of road paved, or number of students passing the high school proficiency test.
- *Outcome indicators:* These measures report the results of programs and services. Outcome indicators possess quantitative and qualitative aspects. Examples of outcome indicators include the number of individuals employed 6 mo after participation in a job training program or percentage of residents who feel safe in their neighborhood.
- *Efficiency and cost-effectiveness indicators:* These measures focus on how a goal is achieved, rather than what was achieved. Specifically, efficiency indicators refer to the ratio of the level of service (tons of refuse is collected, number of meals delivered, and number of flu shots administered) to the cost, in dollars and labor, of providing the services. They measure the cost per unit of an output or outcome. Examples of efficiency indicators include cost per meal delivered, cost per ton of garbage collected, and cost per flu shot administered.
- *Productivity indicators:* These measures, according to Ammons, combine the dimensions of efficiency and effectiveness in a single indicator.^[4] For example, the number of meals delivered per hour measures the efficiency, and the number of meals delivered on time (and warm) measures the effectiveness. The unit cost (labor-hour) per on-time delivery measures productivity. To determine the

productivity indicator, the cost of late deliveries, as well as the on-time deliveries, is included in the numerator of such a calculation, but only on-time deliveries are included in the denominator, thereby measuring efficiency and effectiveness.

Designing a good performance measurement system may be challenging to managers not accustomed to measuring or setting performance targets. However, this effort should not be dismissed as too complex or too difficult. Developing a system involves an understanding of what the program is trying to accomplish, who the main users or customers are, and a basic knowledge of the current level of services. A good performance measurement system should include the following seven-step system developed by the National Center for Public Productivity, a research center affiliated with the School of Public Affairs and Administration at Rutgers University Campus at Newark:

1. *Identify the program to be measured:* To start with, programs to be measured must be clearly defined. Programs are groupings of routine activities aimed at providing support for specific public services. Groupings of individual activities make up a program. For example, the following activities—street resurfacing, street patching, seal coating, and curb repair—constitute a program that is traditionally called street maintenance. Usually, programs are defined by governments and are generally listed on an organizational chart contained in the operating budget. Programs relate directly to the organizational structure and the managerial areas of responsibility. Choosing which programs to measure is a matter of judgment. On the one hand, programs should not be too few, so that only a small portion of services are covered or the information collected is insufficient. On the other hand, too much reporting can be excessively costly or overwhelming and impractical. Generally, performance measurement systems work best when they concentrate on collecting limited, but essential, information about basic programs that need the most managerial oversight and where accountability reporting is most important.
2. *State the purpose and identify the desired outcome:* Typically, a government department or agency initiates a strategic planning process to clarify its mission, goals, and objectives. Through this process, the agency can identify the outcomes, or results, it wants to achieve through its programs. A manager can only measure the performance of a program if its purpose is clearly stated. Preparing a well-articulated statement of purpose is an important first step. Ideally, a clear mission statement is the starting point. If that is not available, a thorough program description is a good place to begin. For example, GASB offers the following statement of purpose for a public transportation system: The basic purpose is to provide safe, dependable, convenient, and comfortable transportation services at minimum cost to the citizens, including special clients groups such as the handicapped and the elderly.^[5]
3. *Select measures or indicators:* A good system uses a few selected indicators to measure outcomes and performance. Most government programs that have established performance measurement systems incorporate the indicators described earlier in this chapter: input, output, outcome, efficiency, and productivity. For example, in a maternal health clinic, the input measures might be the number of labor-hours, the operating budget, the number of clinics, or the cost of supplies. Output measures could include total population served, number of pregnant women served, or number of prenatal visits provided. Outcome measures might include number of full-term deliveries, number of premature births, or number of low-birth-weight babies. Efficiency measures could include labor-hours per woman served, dollars spent per 1000 infants, or dollars saved as a result of healthy babies, and productivity indicators could include measures such as the cost per full-term delivery or the cost per 100 deliveries.
4. *Set standards for performance and outcomes (performance targets):* Under this step, public managers should specify the conditions under which program goals and objectives are met. Managers need to determine what service effectiveness and quality means for a particular program and explicitly state how they are going to determine whether the stated terms of effectiveness and quality have been met. This involves comparing actual program outcomes or results with some agreed-upon standards, including previous performance (the percentage of residents who feel safe in their neighborhoods this year compared with last year), performance of similar organizations (the percentage of residents who feel safe in their neighborhood compared with the percentage in a neighboring community), performance of the best organizations (the percentage of residents who feel safe in their neighborhoods compared with the percentage of residents who feel safe in their neighborhoods in the recognized “safest communities” in the country), and preset targets

(next year, 85% of all residents will feel safe in their neighborhoods, and in 3 yr, 95% will feel safe). Performance targets should be realistic. They should reflect an attainable standard. Unrealistic performance targets, both unattainably high and insultingly low, demoralize staff members and contribute to dysfunctional behavior.

5. *Monitor results:* Each performance target should be monitored on a continuous basis. Monitoring provides the manager with the results needed to decide whether the performance targets have been met. Systematic and periodic monitoring gives the manager an opportunity to keep tabs on the operation of the program and take corrective action when necessary. Usually, monitoring will vary, depending on the program and target accomplishments. For the most important programs and services, monthly data collection and reporting systems that indicate results will be necessary.
6. *Performance reporting:* A good performance measurement system reports program results on a regular basis. The report focuses on what was accomplished and what it costed the public. Reports should be brief, should convey information graphically, and should include minimal explanatory information. Information should be presented in such a way as to facilitate comparisons over time, comparisons with similar organizations, and comparison with best programs nationwide and against preset targets.
7. *Use outcome performance information (analysis and action):* Information from an effective performance measurement system is regularly used in program planning to re-evaluate goals and objectives, and to adjust priorities. Another use is in managing for results where outcome information is used to promote continuous improvement in program operations and results. A well-developed performance measurement system will enable managers to spot weaknesses and challenges to program delivery, as well as program strengths and opportunities for improvement.
8. *Performance management:* It is useful to distinguish between performance measurement and performance management. Performance management is broader than performance measurement. Performance management incorporates the results of performance measurement into general management practices. The National Performance Review defines performance management as the use of performance measurement information to help set agreed-upon performance goals, allocate and

prioritize resources, inform managers to either confirm or change current policy or program directions to meet those goals, and report on the success of meeting those goals.

It is essential that performance measurement be considered an inherent and indispensable part of the management process. Measurement for measurement's sake alone is insufficient. Managers can have stacks of data at their fingertips, but unless they use that data to improve policy and procedures it is virtually useless.

GETTING CITIZENS INVOLVED

Citizen involvement in performance measurement can help public managers stay focused on what really matters in their communities. Citizen involvement can increase the impact of performance measurement by encouraging managers to look beyond traditional output measures (that often have little meaning to citizens) and, instead, focus on quality-of-life issues and community goals. When you think about it, why do managers care about government performance? They want to improve the quality of service delivery and, ultimately, the quality of life in the communities they manage.

The overall goal of involving citizens in performance measurement is to build lasting processes that involve citizens in assessing municipal performance, so that government policies and services reflect the needs of the community. Citizens and public managers, together, can establish performance measures that are meaningful to both citizens and managers. So, for example, instead of just calculating the tons of refuse collected, public managers might also ask citizens to rate the cleanliness of their streets. Instead of counting the number of squad cars deployed to specific neighborhoods, public managers could ask citizens how safe they feel in their own neighborhoods or in city parks.

The relevance of performance measures increase when managers incorporate citizens' perceptions. A powerful management tool results when public managers combine or compare traditional output measures with outcome measures that reflect citizens' perceptions. For example, city managers might learn that an increased police presence in a residential area has no correlation to a citizen's feeling of safety in that neighborhood. Yet, increased police presence in public areas, such as city parks, directly affects a citizen's feeling of safety. Meaningful measures that the average citizen can understand provide citizens with the opportunity to assess government performance and influence how government services can be made more responsive to community needs and priorities.

Although citizen participation can ultimately improve the level and quality of municipal service provision by making services more responsive to the needs of citizens, it is often difficult to achieve. Citizen involvement typically results in effective policy and meaningful public sector programs;^[6–8] yet, active citizenship is often perceived as burdensome, costly, and time consuming.^[9,10] Performance measurement systems may be easier to design and implement when citizens are excluded and when managers are the ones determining what will be measured and how it will be measured. Although easier to design and implement, such performance measurement systems fall short in terms of measuring what matters most in a community. Citizen involvement increases the social relevance of indicators by combining facts—hard data—with value—how citizens feel. With citizens involved in making decisions, acceptance of the decision may increase, enhancing the likelihood of successful implementation and reducing the amount of vocal opposition. As involvement increases, citizen understanding of governmental operations may increase and criticisms of governmental agencies may lessen, improving the plight of the often-criticized public administrator.

Although challenging, strides are being made. In the United States, in recent years, we have seen the opportunities for citizen involvement in performance measurement increase. The National Neighborhood Indicators Project, an initiative of the Urban Institute in Washington, DC, has been successful in what they refer to as democratizing data to meaningfully involve citizens in the measurement of government performance and to use neighborhood-level data in local policymaking and community building.^[11] The Boston Indicators Project, a partner in the National Indicators Project and one that is recognized as one of the best local level initiatives in the country, relies on broad measures of performance that communicate value. They ensure that public services are responsive to the needs of the citizens, not the convenience of the service providers. They involve the public in making data meaningful to the broadest segment of the population, stimulate public discourse around government performance, and track progress on shared goals and objectives to hold government accountable. The goal of the project is to engage citizens, community-based organizations, the media, the business community, and government in meaningful dialogue to develop a better understanding of the opportunities and challenges confronting the city. They stimulate informed public discourse by effectively communicating data with a broad segment of the population and together, with the community, they track the progress on shared goals and milestones.^[12]

There are various approaches to including citizens in performance measurement efforts, and it is up

to public administrators to select the appropriate approach for their community based on the unique characteristics of that community. Meaningfully involving citizens is not a one-size-fits-all approach. Public administrators need to find the right approach for their jurisdiction, and they need to build trust with citizens. A half-hearted approach will only add to the level of frustration and cynicism that many citizens have toward government. In finding the right approach, it may help to revisit the seven-step performance measurement system discussed earlier in this entry, but this time incorporate citizen involvement in each step. Reflect on your community and think about the most appropriate ways to involve citizens in each step. It may be appropriate to establish a citizen advisory committee, or an ad hoc committee of citizens, to achieve some of the objectives. In other instances, a citizen survey might be required. The revised seven-step process might look something like this:

1. *Identify the programs to be measured:* To start with, programs to be measured must be clearly defined by citizens and public administrators. Choosing what programs to measure is a matter of judgment, and citizens should have an equal voice in determining what programs are to be measured. Programs that are identified by citizens should be recognized as a community priority.
2. *State the purpose and identify the desired outcomes:* A government department or agency, in collaboration with the citizens it serves, should initiate a strategic planning process to clarify its mission, goals, and objectives. Through this process, the agency and citizens identify the outcomes, or results, they want to achieve through agency programs.
3. *Select measures or indicators:* Citizens and managers should determine the most appropriate measures for each program, rather than relying exclusively on the data collected by individual programs. The most responsive program measurement systems use citizen surveys, or some other appropriate mechanism, to assess citizen satisfaction and perceptions.
4. *Set standard performance and outcomes (targets for accomplishment):* Under this step, citizens and public managers should specify the conditions under which program goals and objectives are met. Citizens and managers need to determine what service effectiveness and quality means for a particular program and explicitly state how they are going to determine whether the stated terms of effectiveness and quality have been met.

5. *Monitor results*: Each accomplishments target should be monitored on a continuous basis by the management and staff delivering the service and the citizens receiving it. Monitoring provides the citizens and managers with the results needed to decide whether the performance targets have been met.
6. *Performance reporting*: Reports are shared with citizens, elected officials, the media, and other government watchdogs. Performance data should be presented in a way that is meaningful to citizens, so that they can understand what is happening in their neighborhood, as well as in the community as a whole.
7. *Use outcome and performance information (analyses and action)*: A well-developed performance measurement system will enable citizens and managers to spot weakness and challenges to program delivery, as well as program strengths and opportunities for improvement.

CONCLUSION

Improving the outcomes and results of public sector programs, and the performance of government in general, can be realized when performance measurement systems are more closely aligned with accountability mechanisms and when citizens are meaningfully involved in both the concepts. Appropriate, relevant, and objective measures of performance can be developed when citizens are meaningfully involved in defining performance targets and in framing and shaping performance indicators. The appropriate, relevant, and objective measures that result from this process can then be used to hold individuals and organizations accountable for results that actually matter. Citizens can assume a more significant position in their relationship with government, when they are meaningfully involved in dialogue and deliberation surrounding the establishment of goals and objectives for public sector programs and public sector service delivery, when they are asked what good government looks like to them so that they can help establish appropriate performance targets, and when they are recognized as an integral part of legitimate accountability systems. Accountability systems are more effective when they are based on trust and recognize the primacy and importance of accountability to the public.

Accountability for performance is strengthened when multiple stakeholders agree on what good performance looks like. When citizens and service providers collectively agree on the standards and outcomes that reflect effective governance, relevant accountability mechanisms can be established. And when performance information is widely shared with the public, citizens are able to hold government accountable for results.

REFERENCES

1. Hatry, H.; Kearney, R.; Berman, E.; Eds. Performance measurement principals and techniques. In *Public Sector Performance: Management, Motivation and Measurement*; Westview Press: Boulder, CO, 1999; 304–328.
2. Romzek, B.; Dubnick, M.; Shafritz, J.; Eds. Accountability. In *International Encyclopedia of Public Policy and Administration*; Westview Press: Boulder, CO, 1998; 6–11.
3. *Government Accounting Standards Board (GASB) Service Efforts and Accomplishment Measures*; GASB: Norwalk, CT, 1997.
4. Ammons, D. *Municipal Benchmarks: Assessing Local Performance and Establishing Community Standards*; Sage Publications: Thousand Oaks, CA, 1996.
5. Hatry, H.; Fountain, J.R., Jr.; Sullivan, J.; Kremer, L. *Service Efforts and Accomplishments Reporting: Its Time Has Come*; Library of Congress Catalog Card Number: 90-80879; Government Accounting Standards Board (GASB): Norwalk, CT, 1990.
6. Stivers, C. The public agency as a polis: active citizenship in the administrative state. *Adm. Soc.* **1990**, 22, 86–105.
7. King, C.S.; Stivers, C. *Government Is Us*; Sage Publications: Thousand Oaks, CA, 1998.
8. Schacter, H. *Reinventing Government or Reinventing Ourselves: The Role of Citizen Owners in Making a Better Government*; SUNY Press: Albany, NY, 1997.
9. Timney, M.M. Overcoming NIMBY: using citizen participation effectively. A Paper Presented at the 57th National Conference of the American Society for Public Administration, Atlanta, GA, 1996.
10. Thomas, J.C. *Public Participation in Public Decisions*; Jossey-Bass Publishers: San Francisco, CA, 1995.
11. The Urban Institute, Neighborhood Indicators Project. <http://www2.urban.org/nnip/> (accessed November 9, 2006).
12. Callahan, K. *Elements of Effective Governance: Measurement, Accountability, and Participation*; Taylor & Francis: New York, 2006.

Performance Reporting

William Earle Klay

*Askew School of Public Administration and Policy, Florida State University,
Tallahassee, Florida, U.S.A.*

Sam M. McCall

City Auditor, City of Tallahassee, Tallahassee, Florida, U.S.A.

INTRODUCTION

Performance reporting is the conveyance of information, about activities and accomplishments and their related costs, by persons who are responsible for achieving something for those to whom they are responsible. Performance reporting is a form of feedback that is essential in all organizations—for-profits, nonprofits, and governments. Performance reporting is indispensable in the fulfillment of the stewardship and accountability obligations of organizational leaders to those outside the organizations who have entrusted their money and other resources to the leaders. Reporting performance in governments and nonprofit organizations, though, is uniquely different from reporting performance in for-profit businesses.

DIFFERENCES BETWEEN FINANCIAL AND PERFORMANCE REPORTING

In businesses, the most critical element of performance is to be profitable. Profit occurs when revenues, accomplished through the sale of goods or services, exceed the costs incurred. A business' financial statements, when properly prepared, should show whether it has produced profits. The retention and compensation of executives is likely to be determined by performance as revealed through financial statements. Other aspects of a business' performance, such as the costs or benefits that it might create in the broader society, are likely to go unreported. Being "externalities," those costs or benefits are not identifiable within the context of a firm's financial accounting and reporting systems.

In government and nonprofit organizations, the transaction-based information that reveals their financial performance cannot reveal their accomplishments relative to efforts. If a business firm has been profitable, its balance sheet should show an increase in its assets relative to its liabilities. But if a nonprofit organization or a government agency shows that its assets have increased relative to its liabilities, its stronger

financial position could indicate that performance has been poor—e.g., that the nonprofit has not tried as hard as it should have to provide services or that a government has charged higher taxes or fees than it needed to do. The information that is vital to measure their performance is that which reveals their effects upon the recipients of their services.

In short, the financial statements of government and nonprofit organizations are inherently insufficient and incapable of reporting vital aspects of their accomplishments. To report performance for these organizations, special efforts must be put forth to identify these organizations' outputs and outcomes. Outputs are the activities done by an organization (such as students taught, gallons of clean water pumped, or miles of roadways repaved) and outcomes are the societal effects of the activities (such as jobs well performed by graduates, waterborne diseases averted, or the ease with which users transit the roadways). Even though conscientious governments and nonprofit organizations have shown that it is possible to measure and report useful information about performance, most governments and nonprofit organizations are deficient and many have not developed rudimentary, much less adequate, performance reporting systems.

STEWARDSHIP AND ACCOUNTABILITY

Performance reporting by the officials of governments and nonprofit organizations is essential to the fulfillment of their stewardship and accountability obligations. The concept of stewardship has existed for at least two millennia in western culture. The Bible in the book of Matthew describes stewardship as requiring active and faithful management by servants of the money entrusted to them by a master. The essence of the concept remains basically unchanged today. Stewardship obliges public officials to strive to deliver good value to their "masters," their fellow citizens who pay taxes and fees. The leaders of charitable nonprofit organizations are similarly obligated to deliver value to clients on behalf of donors.

Accountability is a relationship between those entrusted to lead and manage an entity and those who entrust the leaders to manage on their behalf. The concept of accountability requires that results be communicated. Without communication of results, the efficacy of organizational leadership cannot be ascertained. For government officials, this relationship is well summarized by the Governmental Accounting Standards Board (GASB) in its Concepts Statement Number 1.^[1]

Accountability requires governments to answer to the citizenry—to justify the raising of public resources and the purposes for which they are used. Governmental accountability is based on the belief that the citizenry has a “right to know,” a right to receive openly declared facts that may lead to public debate by the citizens and their elected representatives.

PERFORMANCE REPORTING HAS LONG HISTORY

The need to hold operating officials accountable for performance is not new. It is unlikely that the pyramids could have been built so well without reports being submitted on the performance of surveyors, stonemasons, suppliers, and so on. Performance reporting was practiced in President George Washington’s first administration, as is evidenced by Treasury Secretary Alexander Hamilton’s orders to the commanding officers of the new revenue cutters that they keep a regular journal of activities and submit these to the Treasury monthly.^[2]

A century later, the activities of a philanthropically supported organization called the New York Bureau of Municipal Research deeply influenced the emergence of professional public administration in America. The Bureau’s founders maintained that reporting about performance was an obligation of public servants in the interest of fostering democratic government. The best way to assure that governments will continually seek to become more efficient, they argued, is to inform citizen-voters about the performance of government relative to its costs.^[3] The influence of the Bureau was subsequently felt in several efforts to improve the performance of the federal government. Most notable among these was the first Hoover Commission in 1947. Seeking to hold administrators accountable for performance, the commission called upon Congress to adopt a performance budget approach for the federal government.^[4] Budgeting for performance remains a major theme of reformers to the present. Any serious attempt to implement performance budgeting must include performance reporting because the former is impossible without the latter.

CONTEMPORARY DEVELOPMENTS IN PERFORMANCE REPORTING

In the early 21st century, performance reporting has become a focal point of reform in both budgeting and financial reporting. In 1998, the National Advisory Council on State and Local Budgeting issued a set of recommended practices to guide budgeting in state and local governments. In its recommended practice titled “Monitor, Measure, and Evaluate Performance,” the council underscored the essentiality of performance measurement and reporting to budgeting.^[5]

Performance measures, including efficiency and effectiveness measures, should be presented in basic budget materials, including the operating budget document, and should be available to stakeholders. Performance measures should be reported using actual data, where possible. At least some of these measures should document progress toward achievement of previously developed goals and objectives. More formal reviews and documentation of those reviews should be carried out as part of the overall planning, decision-making, and budget process.

Performance reporting is now mandated by law in the federal government. The Government Performance and Results Act of 1993 required every major federal agency to set goals, measure performance, and report its accomplishments. The law mandated that every agency prepare a Performance and Accountability Report containing program performance information.

One of the obstacles to doing performance reporting in the past was the absence of professional recognition for doing it well. The Association of Government Accountants (AGA) now has two parallel programs to provide professional recognition for good performance reporting. Each recognition program uses a peer review process. For the federal government, the AGA established in 1996, in conjunction with the Chief Financial Officers Council and the OMB, the Certificate of Excellence in Accountability Reporting program (CEAR). Its purpose is to foster excellence and improve financial and program accountability by streamlining reporting and improving the effectiveness of reports.

Through this program, the AGA advises federal agencies on how to prepare integrated and user-friendly performance reports that show what an agency has accomplished with taxpayer dollars and the challenges that remain. Participating agencies submit their reports to an AGA review team for an in-depth evaluation using established guidelines. If the reports measure up to the AGA standards, the Certificate of Excellence in Accountability Reporting is awarded.

Participating federal agencies receive training by the AGA on how to prepare the report and then receive a comprehensive evaluation of their performance reporting along with detailed recommendations for improvement. In 1998, the first year of the program, three federal agencies received the CEAR Certificate. By 2003, that number had increased to 10 federal agencies, including the departments of Education and Labor, the General Accountability Office, and the Social Security Administration.

In 2003, the AGA inaugurated a program for state and local governments called the Service Efforts and Accomplishments (SEA) certificate program. The SEA program encourages those governments to effectively communicate performance to the public by preparing and issuing high-quality performance reports. Among the 20 organizations that volunteered to participate in the first year were the Oregon State University; Missouri Department of Transportation; Austin, TX; Charlotte, NC; Phoenix and Maricopa County, AZ; Portland, OR; New York City, NY; and Prince William County, VA. For generations, reformers have admonished governments to do performance reporting. The admonishments have often fallen on deaf ears. Perhaps the creation of professional recognition programs for performance reporting will encourage its future development.

GASB's SUGGESTED CRITERIA

Since its creation in 1984, the GASB has been encouraging research and experimentation in performance measurement and reporting. In 1997, GASB received a grant from the Alfred P. Sloan Foundation to support its research. Many observers expect that the board might someday initiate its due process procedures that could lead to an authoritative pronouncement that general-purpose external financial reporting for state and local governments should include performance reporting. When, and whether, this will occur is uncertain. Governmental Accounting Standards Board's research, though, has produced an important set of guidelines for doing performance measurement and reporting.

In a special report, GASB identified 16 suggested criteria to guide state and local government officials in preparing external reports on performance.^[6] The purpose of the criteria is to help officials to produce useful performance reports, ones that will enable readers to ascertain the extent to which an organization has accomplished its mission, goals, and objectives and to better understand the context in which this has been done. The GASB criteria emphasize the following:

- Identify major goals and objectives.
- Design performance reports to clearly communicate to users.

- Focus on key relevant measures to assess results.
- Clarify costs and resources used to attain results.
- Present citizen/customer perceptions of performance.
- Compare performance to past periods, targets, or other governments.
- Use consistent measures to allow comparisons over time.
- Help users understand factors affecting performance.
- Help readers to assess the credibility of reported information.
- Report performance regularly to facilitate decision-making.

ALTERNATIVE WAYS TO REPORT PERFORMANCE

It is possible to make a performance report an integral part of a government's Comprehensive Annual Financial Report (CAFR). Doing that, however, would cause much consternation over the scope of the organization's external financial audit. It is generally advisable, therefore, to clearly separate the reporting of an organization's operational performance from the financial statements that report the organization's financial performance. The primary reason for clearly demarcating the two is to avoid confusion as to whether an organization's financial auditors, in addition to their auditing activities that are related to the financial statements, are obligated to also audit the reports of operational performance. The accountants who are trained to audit an organization's finances are seldom trained to review the program evaluation activities that undergird the reporting of operational performance. Their expertise, though, is relevant to reviewing the reported costs of an organization's activities.

It is also possible to report performance within the context of annual budgets. Reporting on performance within the context of budget requests is essential to doing performance budgeting. Many governments, however, prefer to report performance in ways that are independent from both budgets and CAFR's. Seven counties and 11 cities volunteered to be charter participants in the Association of Government Accountants' Certificate of Excellence in Service Efforts and Accomplishments Reporting program. Six of the seven counties and 10 of the 11 cities chose to use a format for reporting performance that was separate from either their budget or CAFR documents. The other two governments reported their performance information within their budgets.^[7] The AGA certificate program is quite flexible, structured to accommodate whichever form of performance report

the government considers to be its primary report. The variation in the reporting methods used by these governments suggests that leeway should be provided to enable governments to tailor performance reports to their specific situations.

ENSURING CREDIBILITY OF PERFORMANCE REPORTING

Governmental Accounting Standards Board researchers conducted a series of focus group programs with citizens in several locales. The participants indicated that they want performance information to be independently verified. In its suggested criteria for performance reporting, GASB emphasizes the importance of assessing the credibility of the reported information. The Government Finance Officers Association has stressed the need to verify performance information and has indicated that internal auditing procedures should be sufficient.

Standards for auditing performance reports are enumerated by the U.S. Government Accountability Office in its “Yellow Book,” formally titled *Government Auditing Standards*. The internal audit staffs of state governments and larger local governments should know how to use the Yellow Book standards to review the credibility of performance reports. Small governments, though, often lack significant internal audit capacity. For them, other means need to be explored. Independent citizen review boards and the AGA’s certificate review process can be used to provide an informed opinion about the credibility of the information reported about performance.

Whatever the means used to assure credibility of performance reports might be, one thing about performance reporting should always be remembered. Ultimately, the most important users of performance reports are the citizens who hold officials accountable at the ballot box. Providing credible information about performance to citizens strengthens democracy.

REFERENCES

1. Government Accounting Standards Board (GASB). Concepts Statement No. 1, *Objectives of Financial Reporting*; GASB, 1987; paragraph 56.
2. U.S. Coast Guard. Alexander Hamilton’s Letter of Instructions to the Commanding Officers of the Revenue Cutters, June 4, 1791. Accessed at <http://www.uscg.mil/hq/g-cp/history/FAQS/hamiltonletter.html>.
3. Schacter, H.L. Democracy, scientific management and urban reform: the case for the Bureau of Municipal Research and the 1912 New York City School inquiry. *J. Manage. History* **1995**, 1 (2), 52–64.
4. U.S. Commission. In *The Hoover Commission Report on Organization of the Executive Branch of the Government*; McGraw Hill: New York, 1949. (A single volume compilation of the commission’s recommendations.)
5. National Advisory Council on State and Local Budgeting (NACSLB). In *Recommended Budget Practices: A Framework for Improved State and Local Government Budgeting*; 1998. Accessed at <http://www.gfoa.org/services/nacslb/>.
6. GASB. In *Reporting Performance Information: Suggested Criteria for Effective Communication*; Special report; GASB: Norwalk, CT, August 2003.
7. Association of Government Accountants. *SEA Program*. Accessed at <http://www.agacgfm.org/>.

Performance-Based Budgeting

Judith R. Brown

*Planning, Research and Evaluation Division, Governor's Office of Planning and Budget,
Atlanta, Georgia, U.S.A.*

INTRODUCTION

What benefits do citizens realize from their investment in government? We pay taxes to fund government services; but how can we be sure the services have their desired impact? Over the past century, expectations for government budgeting, management, and accountability have shifted from controlling resources to measuring the public benefit derived from using the resources. Although performance-based budgeting promises to be a rational, systematic way to link funding to program effectiveness and efficiency, a greater promise may lie in its potential for improving government policy making and management.

DEVELOPMENT OF PERFORMANCE-BASED BUDGETING

Traditionally, funding for public agencies has been based on previous years' expenditures and activities. While public policy and program impact are discussed, the emphasis of the budget process is control of inputs and activities. However, the concept of linking performance to funding is not new. One of the earliest calls for performance measurement was a recommendation of the Taft Commission on Efficiency and Economy's in 1912:

In order that he [the administrator] may think intelligently about the subject of his responsibility he must have before him regularly statements, which will reflect results in terms of quality and quantity, he must be able to measure quality and quantity of results by units of cost and units of efficiency.^[1]

Since then, numerous strategies to link program funding to performance have been implemented and, for the most part, discarded. Three recent strategies were precursors of performance-based budgeting (PBB): Planning Programming Budgeting Systems (1965), Management by Objectives (1973), and Zero-Based Budgeting (1977).

While most governments that had adopted these strategies reverted to incremental budgeting, a few, mainly local governments, continued the effort to link

performance to program funding. As early as 1973, Sunnyvale, California, began adopting comprehensive performance planning, measurement, and budgeting.^[2] Gradually, other state, county, and municipal governments followed Sunnyvale's lead. The federal government joined the trend with the enactment of Chief Financial Officers Act of 1990^[3] followed by the Government Performance and Results Act of 1993.^[4] Both Acts mandated that federal agencies collect and report performance measures for use in the budget process. In 1994, the Government Accounting Standards Board published Concepts Statement 2, which promotes the use of performance measures in state and local financial reports.^[5] By 1999, the federal government, myriads of county and city governments, and 49 of 50 states had initiated some form of PBB.^[4,6-8]

COMPONENTS OF PERFORMANCE-BASED BUDGETING

Performance-based budgeting (PBB) links the results of program activities to budgeting and policy making. It moves the budget process away from "who gets how much" to "what benefit is derived for how much." Generally, a PBB will include six primary components: program; program purpose; goals; performance indicators; performance targets; and explanatory data. [Table 1](#) provides definitions and examples of each component including five types of performance indicators.

BENEFITS AND CHALLENGES

Performance-based budgeting (PBB) has been in existence long enough to identify key benefits and challenges to its continued use and contribution to government management and policy making. Some of these benefits and challenges are discussed in the following paragraphs.

Defining Purpose and Goals

To some extent, the first tasks in developing a PBB can be among the most challenging and useful. Articulating

Table 1 Components of performance-based budgets

PROGRAM: A systematic set of activities undertaken to accomplish a valued goal

Example: Teen Pregnancy Prevention Program

PROGRAM PURPOSE: Why the program exists; who or what it is to impact

Example: Reduce the teen pregnancy rate through education and services

GOAL: A broad statement of what will happen if the program is effective

Example: Education program participants will not become pregnant before age 20

PERFORMANCE INDICATOR: Short-term, quantified measure of progress toward a goal

INPUT MEASURE: Resources applied to the program

Example: Number of staff allocated to the education program

PROCESS MEASURE: Units of activities and services provided

Example: Number of workshops offered

OUTPUT MEASURE: Incremental change resulting from program activities

Example: Number of teens aged 14–17 completing the program

EFFICIENCY MEASURE: Ratio of input to output; Cost of unit of service

Example: Percentage of program participants completing the program

OUTCOME: What specifically will happen if program is successful

Example: Pregnancy rate for participants 3 years after program completion compared to similar teens not participating in the program

PERFORMANCE TARGET: Unit of improvement or difference

Example: Program participants: three fewer pregnancies per 100 than non-participants.

EXPLANATORY INFORMATION: Clarifying Information; data limitations; disclosures

Example: The characteristics of 15,000 program participants were matched with a comparison group of 2500 non-participants

a program's purpose and its primary goals can require considerable effort on the part of program management. Program staff must focus on what program activities are to accomplish rather than the activities. This can be difficult, as it is not uncommon for agency personnel to become so involved with routine process that they begin to think the activities *are* the program's purpose. This effort can be complicated by ambiguous program mandates and evolving expectations. The development of outcome-oriented

goals forces program managers to create a specific picture of what program success looks like. This gives precise definition to often vague mandates. While the process can be trying, the results can be useful to the policy makers (who may have conflicting views of the program's purpose); the public (who can develop significant misconceptions and unreasonable expectations of the program); and program management (who may realize some of their efforts have been misdirected to activities that do not support their program's purpose).

Developing Performance Indicators

Performance indicators provide program staff, policy makers, and the public with an overview of how well government is addressing public concerns. Because they are short-term measurements, the indicators not only disclose program effectiveness and efficiency but also allow program managers to monitor internal performance. Performance indicators are the linchpins of PBB. In keeping with their importance to PBB, they also are the most problematic.

Although emphasizing the need to measure program results, most PBBs report few outcomes, relying instead on input, process, and output indicators. This is because outcomes are difficult to measure. Sometimes, program staff work so hard improving process that they forget what the processes are to accomplish. While secure in conceptual goals, others are unable or unwilling to translate those goals into measurable indicators. All too often, programs assert they have no responsibility for their outcomes. In their concern over the impact of external factors on effectiveness, they abdicate any responsibility for program results. In other instances, there are legal, logistical, or practical issues that impede the collection of outcome data. Occasionally, no one has found a good way to measure a specific impact. And, often, programs have never collected outcome data, and to modify their information systems would be prohibitively expensive. With the option of reporting other types of performance indicators, many programs avoid reporting on outcomes and assume that if they are effective and efficient in providing services, there is no need to test whether these services achieve desired results.

Identifying Performance Targets

Initially, PBB was promoted as a way to reward programs that performed well and to penalize those that did not. While a few governmental units still adhere to this premise, most have discovered it is simplistic

and counterproductive. PBB had been thought to provide definitive answers on program efficiency and effectiveness. Many have found that, instead, PBB provides more questions. What does it mean when a program has not met its performance target? Is the program strategy ineffective? Is management inefficient? Did something external to the program adversely impact results? Is funding insufficient to achieve success? Without additional research to answer these questions, policy makers lack enough information to take action. Perhaps the most beneficial way to view performance indicators are as tools to diagnose strengths and weaknesses in governmental effectiveness that require follow up to determine the causes.

Although many users take this diagnostic approach, some programs fear that failure to meet a performance target will result in punitive action. Programs must set performance targets several years in advance. To compensate for unforeseen deterrents to performance, some programs understate performance targets to avoid sanctions if they cannot be met. Thus, the targets cease to be inspirational benchmarks and instead become symbols of future mediocrity.

Reporting Useful Data

Issues relating to the adequacy of data are inherent to any type of record keeping and reporting. Many information systems are depositories for easily collected, but not necessarily useful, data. Other times, the data are meaningful, but due to resource constraints or the methods in which the data are collected, there are likely to be undetected errors, omissions, and delays. While some users, for example, the states of Texas and Louisiana, have established systems for independently verifying PBB data, most rely on programs to monitor the integrity of their information.^[9]

Data must also be presented in context. To report process data as though they were outcome data gives the PBB user a false picture of what is measured. Labeling performance indicators by type helps prevent this problem. Sometimes, however, it is necessary to use one type of measure as a “proxy” for another. Other times, flaws in data collection overstate or understate performance, and some data is too expensive to collect. While all of these data concerns are significant, this does not mean that governments should stop collecting data on performance. It means, however, that data limitations should be identified and fully disclosed. Furthermore, because program performance may vary significantly from year to year, multiple years of performance data should be shown to allow the identification of trends.

Effective reporting also requires that the users of the PBB be aware of the reliability and implications of the data reported therein. This requires explanatory data that discloses data limitations and, when necessary, spells out what the reader should surmise from the information presented. Because factors not within the program’s control can cause actual performance to vary significantly from year to year, it is important to present several years’ of information.

Presenting all the PBB data on purpose, goals, performance, and disclosure has proven to be a daunting task. An official in Arizona, one of the pioneers of performance reporting, commented, “we collect so much information at the agency, program and sub-program level that the state is overwhelmed We are data rich, but insight poor.”^[8] This is not an uncommon experience. To address this situation, most PBBs are now more concise, including only key programs and indicators. The State of Georgia has elected to report only outcome data, reasoning that it does not matter if programs are efficient or well run if they are not successful in addressing the problems they were created to solve.^[10]

CONCLUSION

Performance-based budgeting has been advanced as a way to link performance to funding. However, it is unclear the degree to which federal, state, and local governments have been able to do so. The dream of accurately determining the cost of a unit of result is far in the future. And, upon reflection, the concept of rewarding “good performance” and penalizing “bad performance” makes little sense. However, many policy makers and public administrators are finding PBB more useful than imagined. Performance based budgeting is now being reframed as “Managing for Results.” Utilizing program purposes and goals to articulate public policy and using performance data to diagnose problems, identify successes, and improve government services has great promise. Performance-based budgeting has not been proven to be a “silver bullet” that drives the funding of public programs. However, it may be of even greater value in articulating public policy, informing the budget process, improving public services, and making government more accountable to the citizens it serves.

REFERENCES

1. *Report of the Taft Commission on Economy and Efficiency: The Need for a National Budget*; Congress: 2nd Session., House Document No. 854, 1912, Serial Vol. 6300.

2. *National Performance Review: Mission Driven, Results-Oriented Budgeting*; U.S. Government Printing Office: Washington, DC, 20 September, 1993; 21.
3. Public Law 101-576.
4. Public Law 10362.
5. Government Accounting Standards Board. *Concept Statement No. 2: Service Efforts and Accomplishments Reporting*; Government Accounting Standards Board: Stanford, CT, April, 1994.
6. Harkness, P.A. Publisher's desk. Gov. Mag. February **2002**, 15 (5), 4.
7. Melkers, J.; Willoughby, K.G. *Policy Brief: Performance Based Budget Requirements in State Governments*; Andrew Young School of Policy Studies, Georgia State University: Atlanta, GA, 1998; 1 pp.
8. Barrett, K.; Greene, R.; Mariani, M. Grading the states 2001: a management report card. Gov. Mag. **February 2001**, 14 (6), pp. 40, 41, 66.
9. Liner, B.; Hatry, H.P.; Vinson, E.; Allen, R.; Dusenbury, P.; Bryant, S.; Snell, R. Making results-based state government work. Urban Inst. **April 2001**, 12-13.
10. Governor's Office of Planning and Budget, State of Georgia. *The Governor's Budget Report: Fiscal Year 2003*; Office of Planning and Budget, State of Georgia: Atlanta, GA, 37.

Policy Change

James L. True

Department of Political Science, Lamar University, Beaumont, Texas, U.S.A.

INTRODUCTION

Policy change refers to a significant departure from past policy, which may be defined as whatever decision-making organizations chose to declare and implement. Policy changes have enduring qualities, by which they can be distinguished from routine decision-making. Examples could be drawn from business, religious, or other organizations, but public policy and public administration refer to what governments say and do over time, and the examples of policy changes that follow are drawn from U.S. national government policymaking.

DIFFERENT PERSPECTIVES ON GOVERNMENT POLICY CHANGE

What constitutes significance in government choices may depend upon whose ox is being gored. An alteration in government policy that has direct personal effect may be seen as significant, while a similar alteration that affects someone else may not. But in a larger sense, what one sees as significant in new government decisions depends upon one's overall perspective on government policymaking and on the societal consequences of various policies over time.

Government policy changes have historically been viewed from at least four different theoretical perspectives on collective decision-making. They are the perspectives of comprehensive rationality, incrementalism, organizational (or cybernetic) theory, and cognitive decision theory. Each of these theories is briefly reviewed below, along with a more recent approach called punctuated equilibrium theory.

Comprehensive rationality (sometimes called positive political theory or formal theory) is a major framework for much policy research, but it has not been very helpful in distinguishing between routine government decisions and ones significant enough to be seen as policy changes. The contribution of this framework to studying policy change is to emphasize the purposefulness in individual and collective decision-making. Choices are analyzed based on individual preferences and intentions. Although it is clearly possible for policymakers to want one thing and to achieve something else, this quality of purposefulness pervades the study of policymaking.

Comprehensive rationality usually sees government as a single entity that seeks to maximize utility and control risk. Government adaptively changes public policy in response either to new demands for public goods and services or, more generally, to any important change in the external environment. (One variant of this perspective analyzes policy change with a principal-agent model, wherein the government is the agent of the public or, alternatively, where the bureaucracy is the agent of congress and the president.) But comprehensive rationality often draws no clear distinction between routine decisions (essentially policy continuation) and changes significant enough to be called policy changes.

Incrementalism essentially focuses on policy continuation, but important policy change can occur in the form of long-term program growth or contraction. Incrementalism sees government policy ameliorating whatever problem is presented for decision, first by accepting or ignoring much of what has gone before and second by making marginal adjustments to the policy under consideration. From the incrementalist perspective, most policy or issue areas (the internal environment) will largely be accepted as is, with little or no policy change. Some policies will be considered and incrementally revised, and a series of marginal increments (or decrements), if repeated enough times, can result in significant change.

Cybernetic theory distinguishes between routine and non-routine changes in policy. This theory sees government as an adaptive machine—a system that simplifies its decision- and policy-making tasks by ignoring much of the external environment. The machine survives by dealing with a circumscribed set of inputs and by invoking simplified decision rules or standard operating procedures. From this perspective, most decisions are routinely handled without a major change. On occasion, however, an input outside of expected channels will invoke a second level of responses that can result in substantial changes.

One popular form of cybernetic or organizational theory is institutionalism. Institutionalism (and later neoinstitutionalism) stressed the importance of internal organizational arrangements and rules in producing public policies. Yet institutionalism does not fully explain why the same government institutions sometimes react to a particular input with little or no policy change and, at another time, they react to a similar input with a major policy change.

Cognitive theory is something of a special case insofar as policy changes are concerned. Cognitive theory arises from cognitive psychology and generally sees government decision-makers as adaptively or maladaptively changing policy only when external changes in the environment capture their attention. Changes in the objective circumstances can be ignored for some time, when they run counter to what decision-makers have come to expect. From this perspective, it is not just changes in the external environment that generate policy change but the decision-maker's recognition and acceptance of those changes and his or her internal responses to them. Thus, cognitive theorists are concerned with both the external environments and the perceptions and cognitive mechanisms of the decision-maker.

Punctuated equilibrium theory attempts to integrate the incremental, organizational, and cognitive perspectives into a view that sees government as an interactive system of at least two different levels of government decision-making. The interactions between these two internal levels and between the system and the external environment produce both periods of stasis when individual policies do not change and periods when individual policies do change, either incrementally at one level of government or radically at the other level. U.S. government institutions are seen as conservative in the sense that it usually takes a series of collective agreements before a policy can be changed. As a control against tyranny, the framers of the constitutions sought a separation of independent institutions, each branch of which required some cooperation from at least one other branch to make new policy. Yet, separate branches in a federal system also provide multiple venues from which policy changes can be initiated.

From the punctuated equilibrium perspective, the growth of the national government in the twentieth century augmented constitutional sharing of powers between branches and between national and state governments by creating subsystems, "iron triangles," and issue networks, in order to handle the greatly increased tasks assigned to these governments. Policymaking within subsystems features a politics of adjustment, with incremental change resulting from bargaining among interests and marginal adjustments to changed circumstances. The growth of modern society and the expansion of technology further complicated policymaking by creating a variety of new issues and new ways of looking at old issues. Separate branches, multiple levels, and overlapping jurisdictions create a complex interactive system of government. Once a subsystem has garnered policymaking authority in an issue area, public mobilizations are often required to overcome the inertia of vested interests and existing arrangements. From this perspective then, issue definition is critical for attracting the attention of

policymakers outside of a policy monopoly or for mobilizing larger public involvement.

Punctuated equilibrium theorists expect policymaking to produce two sorts of changes: 1) very small changes when policymaking is in the hands of a policy monopoly or stalemated as two fractions of a subsystem vie for control and 2) very large changes after the issue is on the active agenda of the top levels of the national government and that level chooses to rebalance the problems, solutions, and political pressures upon it. The first sort of change is viewed as equilibrium and the second sort as policy punctuations. The first sort would be defined here as policy continuation and the second sort as policy change.

PORTRAYING LASTING AND PURPOSEFUL CHANGES

Although measuring policy changes over time presents several difficulties, longitudinal records of government spending can be used, for they offer the advantages of an easily understood measure (dollars) and an authoritative historical record (the annual budgets of the U.S. government). Some policies and changes to them have little impact on national spending, but most policies require resources to implement them, and the historical record of that resource use can provide a picture of policy changes over time. These examples graph the historical spending records for two representative national purposes: 1) support for national parks and other recreational resources and 2) national funding of medical insurance for the aged. The policy process and the spending for the first purpose are categorized as discretionary, while those for the second purpose are categorized as mandatory. The data for these figures are taken from the *FY 2007, Budget of the United States Government*, "Historical Tables." They trace U.S. Treasury payments for "recreational resources" (budget subfunction 303) and "Medicare" (budget function 570). The effects of inflation have been removed from the series by converting the annual figures to constant FY 2005 dollars, through the use of the chained gross domestic product deflator in the historical tables.

GOVERNMENT SUPPORT OF RECREATIONAL RESOURCES

National government spending for recreational resources is discretionary. The budget requests for the National Park Service and other pertinent programs in this subfunction are scrutinized by the Department of the Interior, are reviewed by the Office of Management and Budget, are presented as part of the President's annual budget, and are reviewed by Congress,

which annually appropriates specific levels of funding for them. Policy changes in this discretionary area include decisions concerning the nature and size of park lands and the proper level of funding for each year.

Fig. 1 indicates that annual national government spending for “recreational resources” grew incrementally through FY 1977. That is to say, annual executive and Congressional reviews of the budget proposals for recreational services resulted in successive years of incremental growth. In FY 1978, there was a policy change from President Carter’s ambitious land heritage initiative, and there was another change in FY 1982 when the Reagan administration chose not to support that initiative at the same level. Later, Congress refused to make the decreases proposed in the executive budget submission for FY 1990. Instead, Congress passed the Emergency Wetlands Resources Act (its own policy change in this area), after which outlays moved to a new level. In FY 1999, Congress agreed with Clinton administration on the use of government oil and gas revenues from the outer continental shelf to fund a Land and Water Conservation Fund that financed the acquisition of new park lands, green spaces, and other initiatives. The budgets of the Bush administration feature park maintenance and improvements rather than expansions, and policy changes again appear to be incremental in this area, although remaining at a relatively high level. Policy changes occurred in this small discretionary area, either as a result of incremental increases reiterated over a number of years or as a result of program initiatives and reassessments that had purposeful and lasting effects.

GOVERNMENT SUPPORT FOR MEDICAL INSURANCE FOR THE AGED

National government spending for Medicare is categorized as mandatory. Medicare is an entitlement to a nation-wide health insurance program for people aged 65 and over and for certain people with disabilities.

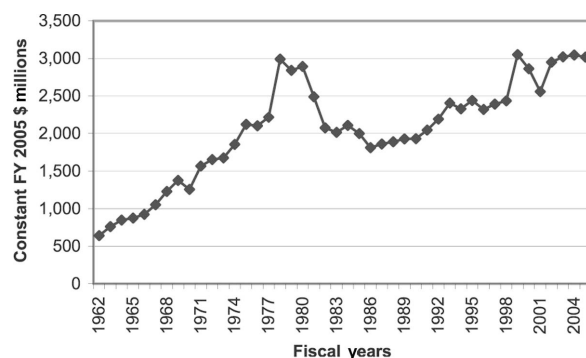


Fig. 1 Outlays for recreational resources. (From Tables 3-2 and 10-1 of the FY 2007 *Budget* “Historical Tables.”)

Medicare entitlements are established by public law and generally are funded without additional Congressional deliberations. A policy change for this entitlement requires a change in program law as to who is covered or what benefits will be funded. Medicare consists of three parts: hospital insurance (Part A); supplementary medical insurance (Part B); and voluntary prescription drug coverage (Part D). The total spending for Medicare consists of the sum of the annual cash outgoes of both the Health Insurance and Supplemental Medical Insurance trust funds. Voluntary prescription drug coverage did not begin until January 1, 2006.

The initiation, expansion, contraction, and latest expansions of Medicare illustrate the purposeful and lasting qualities of policy changes. Medicare was created by the Social Security Amendments of 1965 and expanded in 1972 to include certain people with disabilities. Fig. 2 portrays the long-term growth in Medicare spending, driven largely by increases in the covered population and by expansions in medical technology, treatments, and costs. The purpose of the 1965 amendments was to provide medical insurance support for those aged 65 and over. The growth of the eligible population and the growth of Medicare coverage within that population occurred without a change in that original purpose. This expansion of government spending for Medicare can best be analyzed as a continuation of the policy changes of 1965 and 1972, although the steady expansion in Medicare spending illustrates how repeated incremental growth can result over time in significant program expansion.

Government did change Medicare policy in the Balanced Budget Act of 1997 (BBA of 1997 in P.L. 105-33). BBA of 1997 sought to decrease Medicare spending both to improve the financial viability of its trust funds and as part of other reductions aimed at balancing the federal budget by 2002. Medicare had some special urgency in that the 1996 report of the trustees of the Federal Health Insurance Trust Fund had projected that, without changes, the fund would be in deficit in 2001. The biggest budget savings in BBA of 1997 came from reducing Medicare payments

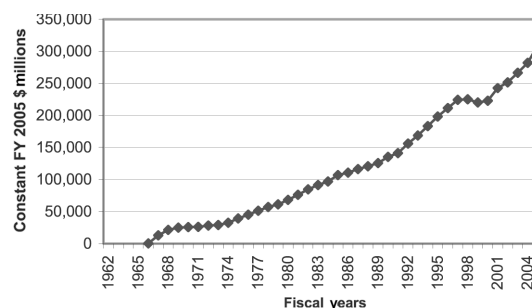


Fig. 2 Annual cash outgo for Medicare. (From Tables 10-1 and 13-1 of the FY 2007 *Budget* “Historical Tables.”)

to doctors, hospitals, nursing homes, home health agencies, and managed care plans. When the payment reductions proved to have larger impact than expected (and under pressure from a variety of interest groups), Congress rescinded some of the reductions with the Balanced Budget Refinement Act of 1999 (BBRA of 1999 in P.L. 106-113).

The historical cash outflows in Fig. 2 indicate that Medicare spending leveled off after the 1997 Act and resumed its increases after the 1999 act. Incremental Medicare expansions were underway even as policy changes on payment reductions were enacted and rescinded. Separate provisions of the two budget acts and the Benefits Improvement and Patient Protection Act of 2001 (incorporated in P.L. 106-554 enacted in December 2000) expanded beneficiary access to preventive-care services. In addition, the Medicare Modernization Act of December 2003 (P.L. 108-173) would likely lead to new annual increases. However, they are not part of this historical spending record, because Part D prescription drug coverage did not begin until January 1, 2006.

Policy changes of a lasting and purposeful nature occurred and may occur again even in the policies and spending for this very large mandatory entitlement program.

POLICY CONTINUATION AND POLICY CHANGE

Theoretical perspectives on policy- and decision-making provide bases for characterizing changes to policy and for distinguishing between the routine nature of policy continuation and the lasting and purposeful nature of policy change. Comprehensive rationality invites a focus on preferences and purposes and looks to the external environment as the impetus for change. Incrementalism emphasizes the importance of an organization's internal environment and notes that government usually responds to external pressures for change with marginal accommodations. Modest increments (or decrements) must be continued over a long period of time if one is to recognize them as significant. Cybernetic theory also focuses on internal processes and the organization's structured relationship with its environment. It also notes the routine nature of most government decision-making and policy, but it adds the possibility of a second level of larger government responses when the circumstances are right. Cognitive

theory notes that decision-makers are not neutral to external circumstances. They accept some external inputs and ignore others. Analyzing policy changes will require some understanding of both external and internal environments, with special emphasis on the attentiveness and attention of the decision-maker. Punctuated equilibrium theory agrees on the importance of policymaker's attention and adds to it the importance of public mobilizations and multilevel government institutions. The result is its view of government as a complex interactive system. Objective changes in the external environment do not automatically generate appropriate government responses. They must be attended to and interpreted before they can energize government policy changes, and there is no guarantee that they will not be ignored. Policymaking at the subsystem level of the national government involves processes and politics that produce policy continuation or at most incremental changes. Policymaking at the top levels of the national government involves processes and politics that can produce significant departures from past policy—policy changes that are likely to have lasting and purposeful effects on the environments.

BIBLIOGRAPHY

1. Anderson, J. *Public Policymaking*, 6th Ed.; Houghton Mifflin, 2006.
2. Baumgartner, F.; Jones, B. *Agendas and Instability in American Policy*; University of Chicago Press, 1993.
3. CQ Press. In *Congressional Quarterly Almanac*; CQ Press [annual].
4. Dye, T. *Understanding Public Policy*, 10th Ed.; Prentice-Hall, 2002.
5. Jones, B. *Reconceiving Decision-Making in Democratic Politics: Attention, Choice, and Public Policy*; University of Chicago Press, 1994.
6. Kingdon, J. *Agendas, Alternatives, and Public Policies*; Harper Collins, 1995.
7. Riker, W.; Ordeshook, P. *An Introduction to Positive Political Theory*; Prentice-Hall, 1973.
8. Sabatier, P., Ed.; *Theories of the Policy Process*; Westview Press, 1999.
9. Schepsle, K.; Bonchek, M. *Analyzing Politics: Rationality, Behavior, and Institutions*; W. W. Norton, 1997.
10. Steinbruner, J. *The Cybernetic Theory of Decision*; Princeton University Press, 1974.
11. United States, Office of Management and Budget. *The Budget of the United States Government*; Government Printing Office [annual].

Policy Design

Anne L. Schneider

College of Public Programs, Arizona State University, Tempe, Arizona, U.S.A.

Helen Ingram

School of Social Ecology, University of California, Irvine, California, U.S.A.

INTRODUCTION

Policy design refers to the content and substance of public policy—the tools, instruments, and other features or “architecture” of policy. Just as a city has a design, so also do public policies—whether constitutions, statutes, governmental pronouncements, agency guidelines, or practices of implementers. Moreover, just as a city’s design can be characterized along many different dimensions, so too the description and analysis of policy design must encompass multiple dimensions.

Governments at all levels have many choices of tools or instruments with which they carry out their responsibilities. These responsibilities all fall within the broad framework of what we call public policy: allocating goods or distributing costs, regulating, extracting resources, persuading, controlling, listening to, and leading the population. The choices governments make in the details of policy designs, however, have long-term effects on politics, democracy, justice, and citizenship.

APPROACHES TO THE ANALYSIS OF POLICY DESIGN

The study of policy design traces to the very beginnings of the study of public policy and administration in the sense that the content and substance of public policy has always been central to these fields of study. The complexity of policy content, however, resulted in the early attempts to study policy design focusing mainly on case studies, without a common framework of description or analysis. More recent work attempts to identify systematic ways of describing policy designs that will enable researchers to compare designs across different policy arenas, and to compare the same kinds of policies across different countries or time periods.

The importance of policy instruments was highlighted by Dahl and Lindblom as early as 1953 when they noted that “this process [the invention and innovation in social techniques] is perhaps the great political revolution of our time.”^[1] They attributed the

capacity of government to avoid having to make stark dichotomous choices between socialism on the one hand and capitalism on the other to the proliferation of techniques that blurred the lines between these seemingly absolute dichotomies.

Most of the initial efforts attempted to describe policy content in broad strokes. Dahl and Lindblom^[1] posited five dimensions they believed could capture differences in designs: private to public; compulsory to informative; direct to indirect; compulsory to voluntary membership; and prescriptive to autonomous within the hierarchy. Ripley’s work^[2] cast the history of the United States in terms of three broad types of policy designs: subsidies (such as the Louisiana Purchase and the homestead act); regulation (of railroads, business, labor); and the most recent, manipulation, which included efforts to reduce the gulf between the indulged and the deprived. Bardach^[3] suggested four types: prescription, enabling, incentives, and deterrence.

Lowi^[4] is widely acknowledged as being the first scholar to propose a way of characterizing policy content that sparked theoretical and empirical development. His work emphasized that the study of policy design should not focus exclusively on why different kinds of designs emerge, but in fact, should recognize that policy content is an important cause of politics itself. The possibilities offered by policy, Lowi contends, create the arena in which politics is played out, with different types of policy-producing variations of pluralist or elitist politics.

Lowi’s typology features two dimensions: the extent of coercion (i.e., whether benefits or burdens are being distributed); and whether the policy is directed at specific groups or whether it consists of rules that affect the more general environment. This fourfold typology produces four types of policies: distributive, regulatory, redistributive, and constituent. His theory suggested that distributive policy, which is the classic pork barrel in that benefits are being distributed to specific groups, will produce a form of particularistic politics with agencies or bureaus, the members of congressional committees and narrow clientele groups benefitting through log-rolling and bargaining. Regulatory policy, which imposes costs that impact many

groups, will produce classic pluralist competition with various interest groups vying to gain regulations that benefit them or inflict costs on their competitors. The politics that results is often partisan, with contests between different coalitions of interest groups played out in congressional floor debates that are highly contentious. Redistributive policy, he believed, would create a form of elitist politics in which the President and close advisors, often including leaders in either business, labor, or other broad sectors of society, would formulate a policy and build support for it. This support often was built on the basis of appeals to class or economic divisions. The policies that were formulated would simply be ratified in the congress.

James Q. Wilson^[5] also developed a typology for policy that offers an explanation of how policy content produces different patterns of politics. The two dimensions in his typology are whether benefits or costs are being distributed and whether the recipients are concentrated or dispersed. Majoritarian politics, he says, will ensue when benefits and costs are both widely distributed over a large number of people. The rationale here is that all will have about the same incentive to pay attention to politics. Interest group politics will emerge when benefits are concentrated on only a few, and costs are distributed among only a few (such as regulations that favor workers at the expense of business or vice versa). His reasoning is that the motivation of both sides will be about the same and both will mobilize to protect their interests. Clientist politics is when elected leaders distribute favors to small and concentrated groups (who then are highly motivated to protect such favors), whereas costs (through taxes) are broadly dispersed leading to no particular motivation to pay attention. Welfare policy is the main example of this type, and Wilson argues that government will always expand its clientist policies because they offer such high payoffs for elected leaders. Entrepreneurial politics occur when small portions of the population pay the costs to benefit large numbers of people. Taxation policies, for example, that impose especially heavy burdens on the rich and distribute the funds widely such as to education or social security, are examples. Wilson argues that government also will expand its entrepreneurial politics far more than is warranted because so many are motivated to continue gaining benefits at the expense of so few.

There are some difficulties with both of these typologies. Lowi's framework has sparked considerable research, but empirical studies are difficult because almost all policies contain elements of more than one of the types. There are almost no pure types. Wilson's framework has this same problem, but in addition, has over time been shown simply to be wrong in many respects. Welfare policy, which he believed would continually expand, has not expanded but retracted.

Federal policy has reduced (not increased) the taxation rate paid by the richest segments of society.

While Lowi and Wilson were interested in the broad relationship of policy content to types of politics, other researchers focused on fleshing out notions of policy elements and tools of government. In 1985, Stephen Linder and Guy Peters characterized policy design analysis as middle-level theory that was a mix of three dimensions: a theory of causality, evaluation, and instruments.^[6] David Boborow and John Dryzek^[7] concentrate on the process through which designs are constituted and legitimated. Their aim is to foster a broad dialogue among different design perspectives on issues such as instrumental and communicative reason so that the design process is open and democratic. Another group of researchers addresses in great detail the advantages and disadvantages of the adoption of certain specific policy instruments or tools. A volume edited by Lester Salmon^[8] addresses a variety of policy tools including grants, tax expenditures, loan guarantees, contracting, public information, and other instruments. The concern of many authors in the tools field is that widespread policy designs embodying contemporary notions of decentralization and devolution may have unexpected and perhaps negative effects.

The understanding of policy design and development of theories of design have been advanced recently by scholars who factor into the analysis the power and social constructions of policy participants, particularly target populations. The basic argument is that variation in designs can be at least partly explained by differences in the power and social construction of target populations and other interested parties in the policy space.

TARGET POPULATIONS, SOCIAL CONSTRUCTIONS, AND THE ELEMENTS OF DESIGN

The symbolic aspects of public policy have long been recognized in the study of politics, but their incorporation into studies of public policy generally and policy design specifically has been more recent. Building on the work of Murray Edelman,^[9] several policy scholars have emphasized the central role played by policy images, the way issues are framed, and the social construction of target populations, events, or the "items" of public policy.^[10-18] What these approaches have in common is the incorporation of meaning, interpretation, symbolism, and social constructions into the description of policy design and into theories of both the causes and consequences of variation in policy design.

Schneider and Ingram^[15,16] posited that policy designs—including constitutions, statutes, agency

guidelines, and everyday practices of implementers—almost always contain a set of common elements including:

- Goals or problems to be solved.
- Target populations (those impacted directly or indirectly by the policy).
- Allocations of benefits or burdens (both material and symbolic).
- Tools (the devices used to insure the behavior needed by the policy).
- Rules (who is to do what, when, with what resources, to whom, with what constraints).
- Implementation structures (agencies, contractors, street-level case workers).
- Rationales (the logic, causal arguments, rhetorical devices, data).
- Underlying assumptions (implicit or explicit, agreed upon or contested).

Each of these has multiple dimension. For example, tools can be conceptualized in terms of their underlying behavioral assumptions producing several “types” of tools: positive incentives, negative incentives, force, information, capacity building, authority (pronouncements without penalties); persuasion; and learning.

Building from a framework of design elements, Schneider and Ingram developed a two-dimensional typology of target populations consisting of the political power resources of target groups and other interested parties (ranging from high to low) and the social constructions of target populations and other interested parties (ranging from positive to negative) (Fig. 1). The policy space created by examining political power and social constructions produces four “types” of target populations (although these are best thought of as a continuous rather than categorical characterization). *Advantaged* populations are those that have considerable political power resources and are positively constructed, usually as “deserving,” or “good.” Examples include business, the middle class, family farmers, and senior citizens. *Contenders* have political power resources to use but are negatively constructed, often as “greedy” or “undeserving” in some way. Examples of these groups include gun owners, the rich, and CEOs of large corporations. *Deviants* is the label given to those who have no legitimate forms of political power and are negatively constructed as dangerous and violent. Examples include criminals, terrorists, flag burners, and gangs. (Terrorists, of course, have the power to generate fear and take the lives of civilians; but by definition, the power of

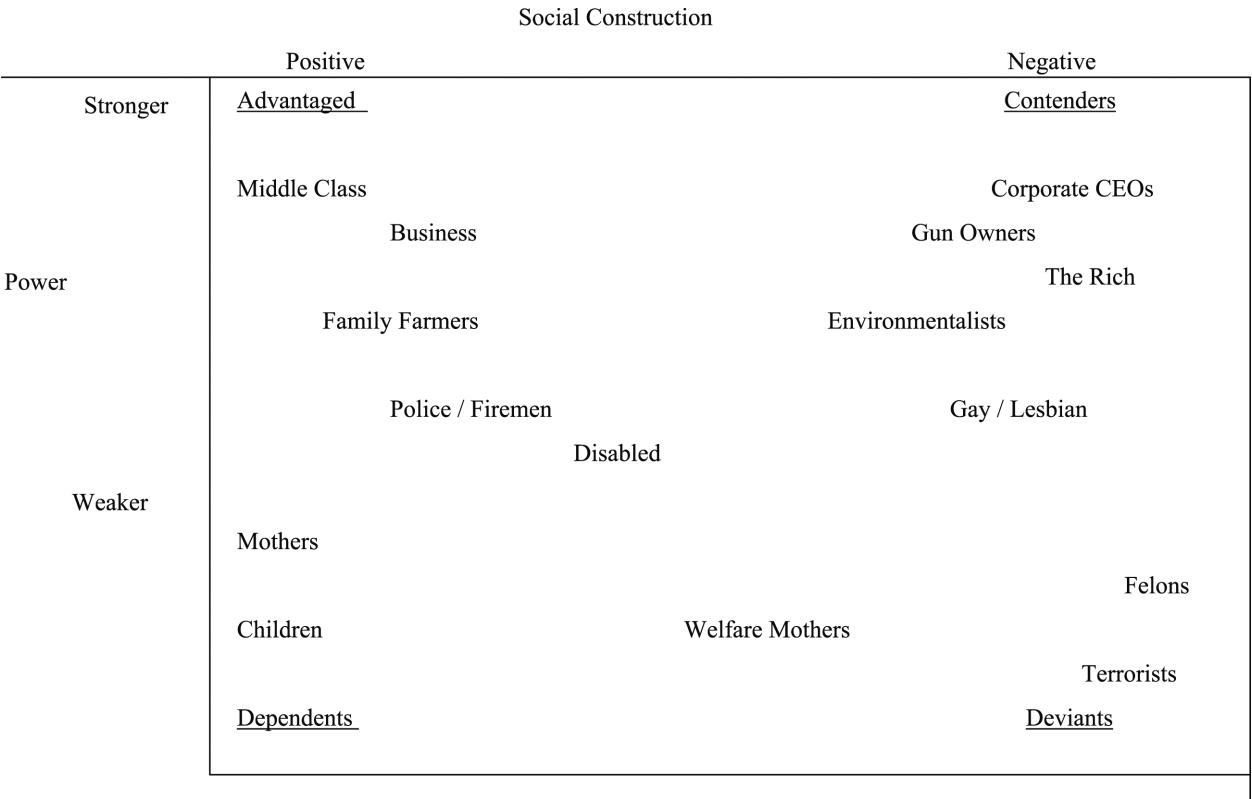


Fig. 1 Power and social constructions of target populations.

terrorists is not legitimate political power in the usual sense of that term.) *Dependents* are persons or groups with little or no political power but who have generally positive constructions, such as “good, deserving, people,” although “helplessness” is also a common characterization of these groups who include children, mothers, the poor, homeless, and so on.

The theory of target populations and policy design that we propose is that each of these segments of the policy space will often be characterized by common types of design elements. Under conditions of pluralist or hyperpluralist political institutions, much can be learned about the characteristics of rationales, implementation structures, rules, tools, goals, and allocation patterns if one focuses on which type of target population the policy (or segments of it) is directed toward. For example, benefits are more likely to be oversubscribed to advantaged populations and undersubscribed to deviants, whereas punishments or costly regulations are more likely to be undersubscribed to advantaged and oversubscribed to deviants. The tools for delivering benefits to the advantaged typically will focus on capacity building (such as subsidies, entitlements, free information, and outreach programs), whereas the tools for delivering burdensome regulations to this same group are more likely to focus on self-regulation, learning, positive inducements, standards and charges, and only rarely, sanctions. Sanctions or regulations are likely to be imposed only when needed to protect one group of advantaged from another group of equally advantaged people. Tools for delivering benefits to dependents are expected to be income-tested subsidies in which the clients must establish their own eligibility, whereas tools for delivering benefits to advantaged groups are expected to be broad-based and not require means testing (social security, e.g., or farm subsidies). Rationales are expected to differ. When benefits are delivered to advantaged groups, rationales usually will be justified on the grounds of important national interests or an efficient means to the goal, whereas benefits to dependents are more likely to be justified on justice grounds such as equal opportunity, need, or fairness. Benefits to deviants generally do not occur unless required by courts on the grounds of fundamental rights or human decency.

This aspect of the theory has produced a significant amount of empirical research that generally supports the contention that social constructions of target populations are centrally implicated in the types of designs that emerge (or whether policy will be adopted). Donovan,^[19–21] e.g., characterized target populations of persons with HIV/AIDS in terms of how they contracted the infection and found that allocations for HIV/AIDS research disproportionately allocated funds to those contracting the disease who were

“advantaged” (medical workers, health professionals, people who got it through transfusions) compared to those who were not so “innocent” (i.e., drug users, homosexuals). Analysis of the generosity of government subsidies varies quite directly with the social construction and power of the target populations when one compares corporate welfare, e.g., with aid to those who are poor.

Other studies have shown that the manipulation of the social construction of target populations is a central aspect of discursive strategies used in legislative policy making. Jensen^[22] shows how some types of veterans but not others came to be eligible for veterans pensions. Sidney^[23] examines the ways in which a Black middle class was separated from “urban rioters” in debates about eliminating discriminatory practices in housing so that benefits could be provided to the former and denied to the latter. Studies of immigration policy document that underlying racist motivations have produced negative constructions as “undeserving” for many minority groups leading to highly restrictive immigration policy.^[24,25] In a similar way, federal drug policy provides a much heavier penalty for crack cocaine than for powder cocaine, arguably because the former is used mainly by inner-city Black males and the latter by White middle-class professionals or athletes. The elimination of welfare as an entitlement owes much to the ability of moral entrepreneurs who have depicted welfare recipients as “welfare queens” or as unmarried, immoral teenagers who have children so that they can collect a welfare check.

Following the ideas contained in the Lowi and Wilson theories, we also have proposed that policy designs produce politics. Social constructions become embedded in public policy, producing messages to target populations that are interpreted and internalized with significant implications for political participation, attitudes toward the state, and citizenship. Mettler^[26] documents that the GI bill was a crucial factor in the political participation of veterans. Soss’s^[27,28] empirical research on welfare compares the impacts of two very different policy designs: social security disability policy and Aid to Families with Dependent Children. His analysis presents strong evidence that differences in these designs have produced critical differences in trust of the state, efficacy, inclination to vote one’s interest in elections, and overall participation patterns.

Theories of policy change also have begun to incorporate the importance of image, symbolism, and social constructions of policy participants in understanding whether policy change will occur and whether it will be incremental or sudden. Baumgartner and Jones,^[13] e.g., in their analysis of punctuated equilibrium theories take into account the policy images and the social constructions of target populations.

CONCLUSION

The central contention of scholars who focus on policy design is that the content and substance of public policy is as important to study as the processes that produce policy. Furthermore, the policy design scholars emphasize that process cannot be well understood without a careful analysis of the product that has been produced. The power, social construction, and ideas that drive policy designs become embedded in the design elements. The goals, problem definitions, rules, tools, rationales, and implementation structures all reflect the dynamics of the policy production process. Policy designs also have significant impacts on target populations, others interested in the policy arena, and the general public. Differences in policy designs have been linked to important differences in attitudes about the state, conceptions of citizenship, and levels of political participation. The study of policy design offers rich potential for understanding how American democracy works, who benefits, and who suffers.

REFERENCES

- Dahl, R.A.; Lindblom, C.E. *Politics, Economics, and Welfare*; Harper and Brothers: New York, 1953; 8.
- Ripley, R.B. *Policy Analysis in Political Science*; Nelson Hall: Chicago, 1985.
- Bardach, E. *The Implementation Game*; The MIT Press: Cambridge, MA, 1979.
- Lowi, T. Four systems of policy, politics, and choice. *Public Adm. Rev.* **1972**, *11*, 298–310.
- Wilson, J.Q. *American Government: Institutions and Policies*; D.C. Heath: Lexington, MA, 1979.
- Linder, S.; Peters, G. From social theory to policy design. *J. Public Policy* **1985**, *4* (3), 237–259.
- Bobrow, D.; Dryzek, J. *Policy Analysis by Design*; University of Pittsburgh Press: Pittsburgh, 1987.
- Salamon, L. *The Tools of Government: A Guide to the New Governance*; Oxford University Press: Oxford, 2002.
- Edelman, M.J. *The Symbolic Uses of Politics*; University of Illinois Press: Urbana, 1964.
- Stone, D. *Policy Paradox: The Art of Political Decision Making*; W.W. Norton: New York, 1997.
- Kingdon, J. *Agendas, Alternatives, and Public Policies*; Little, Brown: Boston, 1984.
- Baumgartner, F.R.; Jones, B.D. *Agendas and Instability in American Politics*; University of Chicago Press: Chicago, IL, 1993.
- Schram, S. Putting a black face on welfare: the good and the bad. In *Deserving and Entitled. Social Constructions and Public Policy*; Schneider, A.L., Ingram, H., Eds.; State University of New York Press: Albany, 2003.
- Schneider, A.L., Ingram, H., Eds.; *Deserving and Entitled. Social Constructions and Public Policy*; State University of New York Press: Albany, 2003; and the introductory chapter, Public Policy and the Social Construction of Deservedness.
- Schneider, A.L.; Ingram, H. *Policy Design for Democracy*; University Press of Kansas, 1997.
- Schneider, A.L.; Ingram, H. The social construction of target populations. *Am. Polit. Sci. Rev.* **1993**, *87/2*, 334–347.
- Ingram, H.; Schneider, A.L. Improving implementation by framing smarter statutes. *J. Public Policy* **1990**, *10* (1), 67–88.
- Ingram, H.; Schneider, A.L. Target populations and policy design. *Adm. Soc.* **1991**, *23* (3), 333–356, 15.
- Ingram, H.; Smith, S.R. *Public Policy for Democracy*; Brookings Institution Press: Washington, DC, 1993.
- Ingram, H.; Schneider, A.L. Constructing citizenship: the subtle messages of policy design. In *Public Policy for Democracy (Red)*; Ingram, H.M., Smith, S.R., Eds.; Brookings: Washington, DC, 1993.
- Donovan, M.C. *Taking Aim: Target Populations and the Wars on AIDS and Drugs*; Georgetown University Press: Washington, DC, 2001.
- Jensen, L.S. *Patriots, Settlers, and the Origins of American Social Policy*; Cambridge University Press: Cambridge, England, 2003.
- Sidney, M.S. Contested images of race and place: the politics of housing discrimination. Schneider, A.L., Ingram, H., Eds.; 2003.
- DiAlto, S. From “problem minority to model minority: the changing social construction of Japanese Americans. Schneider, A.L., Ingram, H., Eds.; 2003.
- Newton, L. It is not a question of being anti-immigration: categories of deservedness in immigration policy making. Schneider, A.L., Ingram, H., Eds.; 2003.
- Mettler, S. Bringing the state back in to civic engagement: policy feedback effects of the G. I. Bill for World War II veterans. *Am. Polit. Sci. Rev.* **2002**, *96* (2), 351–366.
- Soss, J. Making clients and citizens: welfare policy as a source of status, belief, and action. Schneider, A.L., Ingram, H., Eds.; 2003.
- Soss, J. *Unwanted Claims: Politics, Participation and the U.S. Welfare System*; University of Michigan Press: Ann Arbor, 2002.

Policy Diffusion

Karen Mossberger

Department of Political Science, Kent State University, Kent, Ohio, U.S.A.

INTRODUCTION

“Diffusion is the process by which an innovation is communicated through certain channels over time among the members of a social system,” according to Everett M. Rogers, one of the pioneers in diffusion research.^[1] Rogers defined the diffusion process as a form of social change, based on communication, ideas, and information.^[1] Diffusion research has been undertaken in a number of social sciences and professional disciplines, such as sociology, geography, communications, economics, organizational studies, business, and education. *Policy diffusion* involves the spread of policy ideas across political jurisdictions and may occur within a single country or across national borders.

DIFFUSION AMONG THE AMERICAN STATES

Within the United States, the primary focus has been policy diffusion among the states rather than local governments. Jack Walker's^[2] seminal article, “The Diffusion of Innovations Among the American States,” made several important arguments that largely set the research agenda for the following decades. He contended that there was an identifiable league of more innovative states, that there were regional patterns of emulation, that state and professional organizations provided a network to communicate policy ideas and information, and that policy diffusion represented a form of “bounded rationality” that reduced the costs of adopting new policies. The hypotheses set forth by Walker were both amplified and challenged in subsequent studies. Among the topics that have been explored are innovativeness as a characteristic of states, alternative determinants of diffusion, policy variation, sources of diffusion, and processes.

Timing of Adoption

Consistent with diffusion research in other disciplines, policy diffusion studies have often tried to explain the difference between pioneers and laggards, or between early and late adopters of an innovation. This approach defines “innovativeness” as a trait of adopters. Examining 88 policies adopted from 1870 to 1965, Walker^[2] created an innovation index, ranking

the states according to their propensity to be early adopters across a range of policies. He found that early-adopting states tended to be larger, wealthier, and more urbanized. These findings were soon challenged by Virginia Gray,^[3] who argued that patterns of policy innovation varied across different types of policies. The speed of diffusion and the trajectory of adoptions differed among the three policy areas she examined—civil rights, education, and welfare policy. Gray's work served to direct subsequent research toward the characteristics of *policies* and their interaction with state characteristics.

Later research supported some of Walker's contentions regarding innovative states. Robert L. Savage^[4] updated Walker's innovation scores, comparing 181 policies. He discovered that some states did tend to innovate across policy areas, but that changes occurred in the relative positions of states in the latter twentieth century, with only about one-half of the states retaining their former rankings. Survey research and interviews confirmed that states do in fact consider some states to be policy leaders.^[5–7] Grupp and Richards^[6] examined numerous policy areas and found that there was often consensus on leading states, but that these could differ by issue area.

Policy Determinants

Later studies tended to focus on the factors that influence adoption of a single policy. This policy determinants approach involves the examination of variables that Mooney and Lee^[8] described as resources, demands, and constraints. These factors can be categorized as external factors and internal characteristics.^[9–11] External factors might be the influence of federal grants, policy networks, and neighboring states. Internal state characteristics often included in these studies are social and demographic variables (population size, income, poverty rates, urbanization, education level, religion), political variables (public opinion, party competition, political culture, interest group activity), and measures of state resources (legislative professionalization, state budgets, or state wealth). The methodology for studying policy diffusion has become more sophisticated in recent years. Event history analysis, first used in diffusion research by Berry and Berry in 1990,^[9] has become the method of choice. This has

allowed researchers to explore the influence of factors that change in any given year: for example, regardless of whether there was an election that year, whether a neighboring state had passed similar legislation in the prior year, and so forth.

One limitation of these detailed studies is generalizing beyond the specific policy involved (see Refs.^[12,13] for a similar, more general critique of diffusion studies). A common approach to theory building has been to distinguish patterns within policy types. Hwang and Gray^[14] categorized state policies as either redistributive or developmental and conclude that political factors are more important in the adoption of redistributive policy, whereas economic factors are more important for developmental policy. Research on morality policies, which involve the legitimation of values or first principles,^[15,16] reveal distinct patterns that share little with either redistributive or developmental policies. Factors important in the diffusion of morality policies are demographic characteristics, such as the religious composition of the state population and political variables, such as interest group activity, partisanship, and public opinion.^[15] Administrative reforms such as e-government may depend more on factors such as legislative professionalization and professional networks. Because administrative reforms often lack political visibility, political variables such as public opinion, party competition, and interest group activity may not be relevant.^[17] Although some diffusion research draws comparisons across policy areas,^[18] no general theory of policy diffusion exists.

Policy Variation—Reinvention, Learning, and Policy Labels

Studies of innovativeness and the determinants of diffusion have described innovation as early adoption rather than the development or adaptation of an idea. Timing of adoption fails to capture important dimensions of innovation. Later adopters may engage in significant fine-tuning or may adopt more extensive policies, exhibiting arguably more innovative behavior than early adopters who merely copy a policy.^[19,20] Some evidence suggests that later adopters do tend to embrace more comprehensive or more permissive policies.^[8,19] Clark^[21] found that both “leaders” and “laggards” tended to have more extensive programs than middle adopters, except in the case of complex policies or redistributive policies, where laggards did not have broad programs. Hays^[19] also discovered that the tendency toward more comprehensive policies did not hold for those that were controversial and argued that ideology trumped learning in these cases.

Reinvention and comprehensiveness are often used as a means of measuring the amount of policy learning

that occurs through diffusion.^[8,19] As later adopters observe the actions of other states, they are able to elaborate upon them, especially because such policies are seen as less risky than in the earlier stages of diffusion. This represents a more sophisticated view of diffusion and begins to chip away at the normatively significant question of policy learning. However, other research is needed as well, to understand whether social learning occurs. We do not know, for example, whether adopters have observed the experience of other states and have drawn policy-relevant conclusions from these in fashioning their own programs. Programs may become more comprehensive in a sort of bandwagon effect, even if evidence emerges that a policy is not really effective. The rapid diffusion of tax incentives and competition among states to increase their packages of incentives defied the research of the time that such programs were not effective.^[22]

There is often wide variation in state policies that has little to do with increasing scope over time (see also Ref.^[1]). In her five-state comparative study of the adoption of the enterprise zone program, Mossberger^[22] found that the idea had diffused as a policy label, or a vague category of policies for distressed areas, encompassing wide variation. These policies had little in common other than the general idea of targeted incentives and the symbolic content the policy conveyed. Still, even within this context, “informed decision making” was possible, as states were well aware of what others had done and used this information to design programs to fit their own needs.^[22]

Sources of Diffusion—Regional Emulation, Networks, and Entrepreneurs

Walker's^[2] research traced geographic patterns of diffusion, revealing what he called regional systems of competition and emulation. This tendency toward regional emulation has been confirmed in numerous subsequent studies (among the more recent ones are Berry and Berry,^[9] Berry and Berry,^[23] Mintrom and Vergari,^[11] Mintrom,^[24] Mooney and Lee,^[8] Hays and Glick^[10]). Walker^[2] hypothesized that policy makers took cues from the most innovative states, their neighbors, and national state and professional organizations.

Walker's proposition that state and professional organizations provided a national diffusion network has been corroborated by later studies.^[22,25–28] Policy networks are often more extensive than these state and professional organizations because they may involve interest groups or issue advocates participating in national networks with contacts in the states.^[11,24,29] Mintrom's^[24] study of school choice examined the special role of policy entrepreneurs within such national networks.

Federal regulations, grants, and dissemination of information often stimulate the diffusion of policies across the states. Eyestone^[30] characterized this as “point-source” diffusion, emanating from the federal government to the individual states. Gray^[3] and Welch and Thompson^[31] noted more rapid diffusion of federally supported innovations. More recently, the networks in intergovernmental programs have been described as “polydiffusion” rather than point-source diffusion.^[22,28] Interviews with state officials and research in two different policy areas indicated that federal promotion of a policy idea is accompanied by the diffusion of information by state and professional organizations, and also interstate contacts. States report multiple sources of information in an interactive network.

There are some examples of policies that diffuse from the states to the federal government, in what is often called vertical or state-federal diffusion, but there has been less research in this area than in interstate diffusion. The examples of state experimentation with New Deal policies such as unemployment insurance and welfare policies are well known.^[32,33] Studies of more recent federal legislation document the influence of state programs on the Clean Air Act of 1990,^[34] welfare reform and other social policies,^[35] and the Empowerment Zones.^[36] Although it does occur, learning from state programs is often complicated by differences between federal and state policy needs, and the ambiguity of state experience.^[35,36]

Information Use

As the preceding review of the American literature shows, relatively little attention has been devoted to information use in the diffusion process—what information circulates in networks, how information is used, and how decisions are made. Two types of studies directly relate to what Rogers has called the innovation decision, or the use of information in diffusion. Comparative case studies of agenda setting have used comparative case studies to find out how and why issues arrive on the agenda in some states, but not others.^[27,37,38] Mossberger's^[22] study of the enterprise zones compared decision-making processes in five states. Two of the states fit the model of “bounded rationality” employed by Walker in his arguments, and by Richard Rose (Ref.^[39], see below). Three other states more closely resembled the policy streams model^[40] or the garbage-can model of decision making.^[41]

CROSS-NATIONAL DIFFUSION AND POLICY TRANSFER

Research on cross-national diffusion addresses geographic and temporal patterns of diffusion, the

determinants of innovation, and causes and sources of diffusion. In addition, however, literature on information use (cross-national policy transfer or lesson drawing) has flourished. Importing a policy across national borders is often more problematic than borrowing from a neighbor in the same country. Differences in institutions, political culture, resources, policy history, and other factors call for more attention to the way in which emulation occurs.

Timing of Adoption and Determinants

Early cross-national diffusion research also concentrated on searching for general models to describe and predict patterns of diffusion, much like the American concern with innovativeness as a trait or the policy determinants research. Collier and Messick^[42] examined the timing of adoption of social security policy. They found that the social security policies did not begin in the most economically advanced countries, but among nations with a medium level of development. A certain threshold of development appeared to be necessary to support such policy. It later traveled “up” the hierarchy of economically developed nations, and then switched to a pattern of spatial diffusion among neighboring countries. The Collier and Messick^[42] study suggested that multiple forces contribute to innovation—resources, political demands and constraints, and emulation.

Causes of Diffusion

Both “push” and “pull” factors are apparent in cross-national processes of diffusion.^[43] Voluntary adoption, based on the attractiveness of policies addressed to similar problems, is the most-studied method of policy diffusion. However, policies also spread through imposition or force. Colonialism and military occupation are the most draconian methods of policy transmission. Other “push” factors include supranational organizations and international treaties that promote policy diffusion through regulation and penalties, as well as through information exchange.^[44] Economically and politically powerful countries are also able to export their policies to other countries.^[43,45] Other factors facilitating emulation are geographic or ideological proximity,^[45,46] informal policy communities based on professional ties or expertise,^[46–48] or policy entrepreneurs.^[44,49]

Information Use as Lesson Drawing

Richard Rose^[39,46] offers a model of voluntary policy transfer as bounded rationality that he refers to as lesson drawing. Policy adopters may emulate the example

of others in whole or in part. For Rose, the impetus for cross-national policy transfer is a gap between policy performance and expectations. This performance gap stimulates a search for possible solutions. Lesson drawing uses the experience of others to bound the search for possible solutions, and to reduce the uncertainty and risk involved in a completely untested policy. Policy makers engage in a process of prospective policy evaluation, in which they estimate whether and how a policy that currently works in another country will work under the conditions of their own country in the near future. For lesson drawing, adopters do not need to have a comprehensive knowledge of the original policy and comparative conditions in both countries. Instead, they must only develop a sketchy conceptual blueprint of what will be required for success, given a comparison between the two policy settings and knowledge of key elements of the policy.^[46]

Diffusion as Policy Transfer

This model of lesson drawing is prescriptive and empirically based. Voluntary policy transfer, however, does not always resemble the model of lesson drawing. As Bennett^[50] showed, learning and emulation are not always the goals of those who use policy evidence from abroad. Foreign experience may be introduced merely to persuade policy makers to put an idea on the agenda. A program may also be adopted without examining how well the idea worked and what adaptations are needed. Dolowitz and Marsh^[44] argued that policy failure occurs because transfer is uninformed, inappropriate, or incomplete. Ideas can also be imported as a quick fix, a crisis response, or a strategy to quell political discontent.^[43,50,51]

Some scholars have suggested that cross-national policy transfer may resemble the opportunistic decision making of the policy streams or garbage-can models.^[49,52] Wolman^[49] used the term “policy transfer” rather than learning or lesson drawing precisely because he tested the rationality of the process, using the example of British emulation of the American Urban Development Action Grant. The British were unaware of the failures of the American policy, and were more concerned with adopting a general idea rather than a specific policy design. Robertson and Waltman^[52] saw the policy streams model as a general phenomenon in policy borrowing, because of uncertainty about policy outcomes and the potential for political conflict. Mossberger and Wolman^[53] surveyed 17 cases of cross-national policy transfer, finding that problems were common for these and other reasons. However, they also discovered some ways in which policy makers coped with the challenges of cross-national transfer. They recommended the use of “mixed

scanning” and various heuristics with the goal of achieving bounded, rather than comprehensive, rationality.

CONCLUSION

The study of policy diffusion has both limitations and promise. Research is often fragmented, with little generalizability across policy areas. Understanding of information use has come under greater scrutiny only since the 1990s, first on the cross-national level, and to a lesser extent, within the United States. Process-oriented studies of information use suffer from their own limitations, as they are often based on a single case, or perhaps a handful of cases. Since the beginning of the 1990s, more sophisticated methods have been incorporated into quantitative studies, allowing for a wider range of factors to be included in models. Network research has begun to examine the *influence* of networks and policy entrepreneurs, rather than just tracing the participants in networks. More studies address issues such as reinvention, learning, or rationality, and more still needs to be understood about whether and how diffusion improves public policy. It is only in this way that we will get closer to comprehending policy diffusion as social change.

REFERENCES

1. Rogers, E.M. *Diffusion of Innovations*, 4th Ed.; The Free Press: New York, 1995; 5–6, 174–180 on reinvention.
2. Walker, J.L. The diffusion of innovations among the American states. *Am. Polit. Sci. Rev.* **1969**, *67*, 880–899.
3. Gray, V. Innovation in the states: a diffusion study. *Am. Polit. Sci. Rev.* **1973**, *57*, 1174–1185.
4. Savage, R.L. Policy innovativeness as a trait of American states. *J. Polit.* **1978**, *40*, 212–219.
5. Walker, J.L. Innovation in state politics. In *Politics in the American States*, 2nd Ed.; Jacob, H., Vines, K.N., Eds.; Little, Brown, and Co.: Boston, 1971; 354–387.
6. Grupp, F.W.; Richards, A.R. Variations in elite perceptions of American states as referents for public policy making. *Am. Polit. Sci. Rev.* **1975**, *69*, 850–858.
7. Freeman, P.K. Interstate communication among state legislators regarding energy policy innovations. *Publius: J. Fed.* **1985**, *15*, 99–111, (Fall).
8. Mooney, C.Z.; Lee, M.H. Legislating morality in the American states: the case of pre-roe abortion regulation reform. *Am. J. Polit. Sci.* **1995**, *39* (3), 599–627.
9. Berry, F.S.; Berry, W.D. State lottery adoptions as policy innovations: an event history analysis. *Am. Polit. Sci. Rev.* **1990**, *84*, 395–415.
10. Hays, S.P.; Glick, H.R. The role of agenda setting in policy innovation: an event history analysis of living-will laws. *Am. Polit. Q.* **1997**, *25* (4), 497–516.
11. Mintrom, M.; Vergari, S. Policy networks and innovation diffusion: the case of state education reforms. *J. Polit.* **1998**, *60* (1), 126–148.

12. Downs, G.W.; Mohr, L.B. Conceptual issues in the study of innovation. *Adm. Sci. Q.* **1976**, *21*, 700–714.
13. Downs, G.W.; Mohr, L.B. Toward a theory of innovation. *Adm. Soc.* **1979**, *10*, 379–408.
14. Hwang, S.; Gray, V. External limits and internal determinants of state public policy. *West. Polit. Q.* **1991**, *44*, 277–299.
15. McFarlane, D.R.; Meier, K.J. *The Politics of Fertility Control*; Chatham House Publishers: New York, 2001.
16. Mooney, C.Z. The public clash of private values. In *The Politics of Morality Policy: The Public Clash of Private Values*; Mooney, C.Z., Ed.; Chatham House: New York, 2001; 3–20.
17. McNeal, R.; Tolbert, C.; Mossberger, K.; Dottwerweich, L. Innovating in digital government in the American states. *Social Science Quarterly*, **2003**, *84*, 52–70.
18. Nice, D. *Policy Innovation in State Government*; Iowa State University Press: Ames, IA, 1994.
19. Hays, S.P. Influences on reinvention during the diffusion of innovations. *Polit. Res. Q.* **1996**, *49* (3), 631–650.
20. Goggin, M.; Bowman, A.O'M.; Lester, J.D.; O'Toole, L.J., Jr. *Implementation Theory and Practice: Toward a Third Generation*; Harper Collins, 1990.
21. Clark, J. Policy diffusion and program scope: research directions. *Publius: J. Fed.* **1985**, *15* (4), 61–70.
22. Mossberger, K. *The Politics of Ideas and the Spread of Enterprise Zones*; Georgetown University Press: Washington, DC, 2000; pp. 68–69, 131; Chapter 5.
23. Berry, F.S.; Berry, W.D. Tax innovation in the states: capitalizing on political opportunity. *Am. J. Polit. Sci.* **1992**, *36* (3), 715–742.
24. Mintrom, M. *Policy Entrepreneurs and School Choice*; Georgetown University Press: Washington, DC, 1999; 194 pp.
25. Baumgartner, F.R.; Jones, B.D. *Agendas and Instability American Politics*; University of Chicago Press: Chicago, 1993; 231 pp.
26. Beaumont, E. Enterprise zones and federalism. In *Enterprise Zones: New Directions in Economic Development*; Green, R.E., Ed.; Sage Publications: Newbury Park, CA, 1991; 41–57.
27. Nelson, B.J. *Making an Issue of Child Abuse: Political Agenda Setting for Social Problems*; University of Chicago Press: Chicago, 1984.
28. Mossberger, K.; Hale, K. "Polydiffusion" in inter-governmental programs: information diffusion in the school-to-work network. *Am. Rev. Public Adm.* **2002**, *32*, 398–422.
29. Kirst, M.W.; Meister, G.; Rowley, S.R. Policy issue networks: their influence on state policy making. *Policy Stud. J.* **1984**, *13* (2), 247–263.
30. Eyestone, R. Confusion, diffusion and innovation. *Am. Polit. Sci. Rev.* **1977**, *71*, 441–447.
31. Welch, S.; Thompson, K. The impact of federal incentives on state policy innovation. *Am. J. Polit. Sci.* **1980**, *24*, 715–729.
32. Skocpol, T. *Social Policy in the United States: Future Possibilities in Historical Perspective*; Princeton University Press: Princeton, NJ, 1995.
33. Waltman, J.L. *Copying Other Nations' Policies: Two American Case Studies*; Schenkman Publishing Co., Inc.: Cambridge, MA, 1980.
34. Kamieniecki, S.; Ferrall, M. Intergovernmental relations and clean-air policy in Southern California. *Publius: J. Fed.* **1991**, *21*, 143–155.
35. Boeckelman, K. The influence of states on federal policy adoptions. *Policy Stud. J.* **1992**, *20*, 365–375.
36. Mossberger, K. State-federal diffusion and policy learning: from enterprise zones to empowerment zones. *Publius: J. Fed.* **1999**, *29* (3).
37. Glick, H.R. *The Right to Die: Policy Innovation and Its Consequences*; Columbia University Press: New York, 1993.
38. Jacob, H. *Silent Revolution: The Transformation of Divorce Law in the United States*; University of Chicago Press: Chicago, 1988.
39. Rose, R. *Lesson-Drawing in Public Policy: A Guide to Learning Across Time and Space*; Chatham House: Chatham, NJ, 1993.
40. Kingdon, J.W. *Agendas, Alternatives, and Public Policies*, 2nd Ed.; Harper Collins, 1993.
41. Cohen, M.D.; March, J.G.; Olsen, J.P. A garbage can model of organizational choice. *Adm. Sci. Q.* **1972**, *17*, 1–25.
42. Collier, D.; Messick, R. Prerequisites vs. diffusion: testing alternative explanations of social security adoption. *Am. Polit. Sci. Rev.* **1975**, *69* (4), 1299–1315.
43. Majone, G. Cross-national sources of regulatory policy-making in Europe and the United States. *J. Public Policy* **1991**, *11* (1), 79–106.
44. Dolowitz, D.; Marsh, D. Learning from abroad: the role of policy transfer in contemporary policy-making. *Governance* **2000**, *13* (1), 5–24.
45. Hoberg, G. Sleeping with an elephant: the American influence on Canadian environmental regulation. *J. Public Policy* **1991**, *11* (1), 107–132.
46. Rose, R. What is lesson-drawing? *J. Public Policy* **1991**, *11* (1), 3–30.
47. Haas, P.M. Do regimes matter? Epistemic communities and Mediterranean pollution control. *Int. Organ.* **1989**, *43* (3), 377–404.
48. Stone, D. Non-governmental policy transfer: the strategies of independent policy institutes. *Governance* **2000**, *13* (1), 45–62.
49. Wolman, H. Understanding cross nation policy transfers: the case of Britain and the U.S. *Governance* **1992**, *5* (1), 27–45.
50. Bennett, C.J. How states utilize foreign evidence. *J. Public Policy* **1991**, *11* (1), 31–53.
51. Robertson, D.B. Political conflict and lesson-drawing. *J. Public Policy* **1991**, *11* (1), 55–78.
52. Robertson, D.B.; Waltman, J.L. The politics of policy borrowing. In *Something Borrowed, Something Learned? The Transatlantic Market in Education and Training Reform*; Finegold, D., McFarland, L., Richardson, W., Eds.; The Brookings Institution: Washington, DC, 1993; 27 pp.
53. Mossberger, K.; Wolman, H. Policy transfer as a form of prospective policy evaluation: challenges and recommendations. *Public Admin. Rev.* **2003**, *63* (4), 428 pp.

Policy Implementation

Ann O'M. Bowman

*Department of Government and International Studies, University of South Carolina,
Columbia, South Carolina, U.S.A.*

INTRODUCTION

Implementation occurs in the middle of the policy process. It results from the stages that precede it, policy formulation and adoption, and it affects the subsequent stages: evaluation and redesign. The verb *to implement* means in its most basic sense, to carry out, to fulfill, or to accomplish. When applied to public policy, implementation is the process of putting into effect or carrying out an authoritative decision of government. This decision is most often enacted by a legislative body (e.g., the Congress, a state legislature, a city council), but it also can be a directive of the executive branch (e.g., a President's Executive Order, an agency's regulation) or a ruling by the judiciary. Implementation puts the objectives of policy adopters into action in an effort to accomplish desired results.

Implementation makes a policy happen, it gives a policy life. Policy formulation and adoption are necessary precursors, but implementation adds sufficiency. That may seem easy enough but several decades of research on policy implementation have demonstrated the complexity of putting policies into action. Few policies are self-implementing, most require concerted, coordinated action by actors other than those who adopted the policy. Implementation may stall, it may be diverted or subverted by implementors with conflicting interests, it may attenuate due to insufficient resources, it may encounter any number of pitfalls. A flawed implementation process makes the accomplishment of policy objectives all the more difficult. Since the early 1970s, researchers have sought answers to a fundamental question: What makes implementation successful?

DISCUSSION

"Policy implementation is what develops between the establishment of an apparent intention on the part of government to do something, or to stop doing something, and the ultimate impact in the world of action."^[1] In the simplest cases, implementation is handled by a single agency at the governmental level where the policy was adopted. When implementation is multiagency, and when it is intergovernmental, the

potential for slippage increases. Because the U.S. federal system is decentralized in design and, by the late 20th century, devolutionary in preference, most significant domestic policies are intergovernmental. Take, for example, environmental protection policies. Federal legislation such as the Clean Water Act requires the U.S. Environmental Protection Agency (EPA) to take certain actions but the bulk of the Act's implementation is the responsibility of state agencies. The Clean Water Act may have been enacted by the U.S. Congress and it may instruct the EPA to promulgate requisite rules, but it is Alabama's Department of Environmental Management and Vermont's Agency of Natural Resources, working with regional EPA offices, which are implementing the law. As it is in so many policy areas, implementation is increasingly a multiagency and multilevel process.

First-Generation Studies

Intergovernmental policy implementation is a complex process, a fact that emerged from a study of the "new towns in-town" program of the Johnson administration.^[2] This was confirmed in 1973 with the publication of Jeffrey Pressman and Aaron Wildavsky's book, *Implementation*.^[3] The subtitle said it all: *How Great Expectations in Washington Are Dashed in Oakland; Or, Why It's Amazing that Federal Programs Work at All*. Why did a well-intentioned, innovative employment policy enacted by Congress encounter so many obstacles during its implementation in the city of Oakland, California? The program provided funds for public works projects and business loans that would in turn produce jobs for unemployed minorities in Oakland. The level of enthusiasm for the program was high, agreement on the policy goals was substantial. But the program did not deliver on its promise. After three years, more than \$1 million in business loans were made, but few jobs were created and major public works projects had not commenced. The Oakland EDA case reinforced the new towns finding: intergovernmental implementation is not easy. In Oakland, it was the "technical details," but most especially, the complexity of joint action required in the implementation process, that doomed the program.

At more than 30 decision points, multiple participants with diverse, often conflictual, perspectives collided.

Research in the Pressman and Wildavsky mold analyzed how a single authoritative decision was carried out—or, as more frequently, not carried out. One valuable result of these case studies was the recognition of the complex nature of implementation. But, although these accounts provided contextually rich, detailed information about a specific program, it was difficult to generalize from them. Implementation succeeded (or was delayed, or was subverted) for a host of seemingly idiosyncratic reasons.

Second-Generation Research

Even as these so-called “first-generation” studies continued, another type of implementation research was underway. This work sought to develop analytical frameworks that identified factors important to implementation. In other words, this “second-generation” research focused on the determinants of implementation success, defined primarily as the achievement of policy objectives. Conceptual models of the implementation process were developed and tested in several different substantive areas. Two approaches predominated: top-down and bottom-up.

Top-down approach

Top-down models of implementation focus on the authoritative decision (e.g., the new statute) and the formal actors (decision makers) in an almost hierarchical fashion. The questions flowing from this model include: What did the law specify? What decisions were made? Did compliance (with the directives of the statute) occur? Were policy goals met?

Top-down researchers found that successful implementation was more likely if the following conditions existed:

1. Clear and consistent policy objectives embodied in the statute (or other legal directive).
2. A sound theory underpinning the statute, one that specifies the factors affecting policy objectives and gives implementors sufficient authority.
3. The placement of the implementation function in a sympathetic agency with adequate structure, supportive decision rules, sufficient funding, and adequate access to supporters.
4. Implementors with substantial managerial and political skill, and commitment to statutory goals.
5. Active support from organized constituency groups and key elected officials throughout the

implementation process, with the courts being neutral or supportive.

6. An absence of conflicting public policies and socioeconomic conditions that would weaken causal theory or political support.^[4]

Other things being equal, a policy that possesses these conditions is more likely to experience successful implementation than a policy lacking in them. Although seldom are all six conditions met, even in suboptimal situations, several steps can be taken to enhance the likelihood of successful implementation. There is no “routine” or “natural” path to implementation, four implementation scenarios are plausible.^[4] With the effective implementation scenario, as the label implies, implementation is smooth and swift. This is likely to occur in programs that seek only moderate changes in the status quo. In the gradual erosion scenario, implementation begins in a promising manner but slows precipitously as forces opposing the policy gain strength. The presence of numerous veto points wears down supporters and emboldens opponents. The link between the authorizing statute and the implementation process attenuates. The cumulative implementation scenario posits steady, slow progress toward implementation. Adjustments are made throughout the implementation process such that compliance from target groups is secured. In the rejuvenation scenario, initial start-up slows but then later resumes. Implementation is saved from the gradual erosion fate by changes in the statute (e.g., legislative clarification), in contextual factors (e.g., economic growth), and in agency characteristics (e.g., new leadership) that promote renewal of the implementation effort.

Bottom-up approach

Bottom-up approaches see implementation through a different set of theoretical and analytical lenses. Starting from the opposite end, bottom-up models focus on the existence of a public problem and the actors connected to it. To understand implementation, bottom-uppers look to the target population and the implementing environment.^[5] Street-level bureaucrats are queried as to their goals for the program or service and relevant strategies, activities, and contacts. This information provides the basis for the next step: constructing a broader network that is multilocal and multilevel. In effect, implementation is mapped from the bottom-up. Rather than unfolding from the top, implementation builds from the bottom or “micro-level.” To achieve successful implementation, local actors adapt policies and programs to fit local conditions. This helps explain how the same policy can produce very different implementation results from one setting to another. Implementation is determined

by the bargaining (explicit or implicit) in the policy subsystem: members of an implementing organization and their clients. Alliances are formed, coalitions are created, and the implementation process is managed. When a new policy or program fits the local context, implementation proceeds. When it does not, implementors seek to delay it or to otherwise thwart the objectives of enactors.

Assessing the two approaches

Top-down approaches are criticized for overemphasizing the role of central decision makers while bottom-up models are said to overstate the amount of autonomy that local level implementers possess.^[6] Regardless, second-generation research yielded several important findings about the implementation process. Chief among them was the recognition that implementation varies over time, across policies, and from one setting to another. Furthermore, it matters at what point in history implementation takes place and over what period of time.

Third-Generation Research

Third-generation research synthesizes the best features of both top-down and bottom-up approaches, in an effort to determine which variables, and in what setting, affect implementation. For example, policy-makers are encouraged to engage in both “forward mapping” and “backward mapping.”^[7] In forward mapping, policy objectives, implementation guidelines, and evaluative criteria are stated clearly and explicitly. But backward mapping shifts the focus to the local level and to target groups. The question becomes: “What has to happen at the local level for this policy to be successful?” Once answered, the follow-up question is: “What has to happen at the next level for this policy to be successful?” This procedure is repeated all the way back to the point of enactment. By engaging in this hybrid top-down/bottom-up exercise, policy-makers have a clearer understanding of the implementation process and therefore will select a better set of policy instruments. The result, the logic goes, is a policy that achieves its objectives.

Another way to conceive of implementation is as a multilevel subsystem full of messages, messengers, channels, and targets.^[8] Implementation is a dynamic process as messages are designed, sent, received, interpreted, and responded to. The content and form of the policy messages, and the standing of federal-level communicators among the implementers, act to induce or constrain implementation. At the subnational level, inducements and constraints are generated by organized interests, elected and appointed

government officials, and the focal state agency. These implementation-related interactions occur in a state setting that is structured by economic capacity, political conditions, and a set of situational variables such as the salience of the problem to a state, the amount of media attention given the problem, and a state’s organizational capacity.

States decide how they will implement a federal directive. Four patterns or styles are possible. The first implementation style, compliance, reflects a state context in which the new federal policy (and /or program) fits, i.e., the goals of the policy are in sync with state goals, the funding is sufficient, and implementors have the required skills. Moreover, it may demonstrate the value of incentives built into the policy that encourage rapid, unmodified implementation. In the second implementation style, strategic delay, a state delays implementation in an effort to make the new policy conform to the state context. This assumes that the new policy contains ample flexibility and discretion for these adjustments to be made. The state’s delay is strategic, thus it anticipates several rounds of negotiation and bargaining between federal-level actors and the state-level implementors. The third and fourth styles of implementation reflect a new policy at odds with a state context. In these instances, a state may simply drag its feet (the “delay” style) awaiting a judicial ruling, a change in the law, or a greater incentive to comply. Contrarily, a “defiance” state may opt to chart its own path, ignoring the dictates of the new policy. State policymakers may consider defiance preferable to the federal policy, but it may lead to the withdrawal of federal funding or the intrusion of federal enforcement officials.

Just as implementation is about communication, it is also about cooperation. However, implementation is often a matter of achieving cooperation among actors who are not necessarily inclined to cooperate.^[9] These “reluctant partners” who may hold different values and may have conflicting interests have to be induced to cooperate. But, unlike the traditional top-down argument, it is not the national-level actors who hold all the cards. Instead, authority is diffuse and subnational actors possess resources that they can marshal strategically. Implementation regimes form to guide the interactions of the various participants as they jockey for position. These regimes are composed of rules, norms, and procedures that shape the way participants interact.

Another influence on implementation is the working relationship that evolves between federal and state personnel.^[10] Two key dimensions of the working relationship are trust and involvement. In instances where trust and involvement are high, personnel at different governmental levels work together and achieve synergies, thereby facilitating implementation.

Where one or both of the characteristics are low, the working relationships are suboptimal. To move these working relationships into the facilitating variety, implementors are encouraged to increase personal contact, use multiple communication channels, share information, promote interagency learning, and provide timely communication and feedback. To increase trust among the actors, open dialog is encouraged while back-door and end-run communication tactics are discouraged, ends rather than means are to be kept in sight, meaningful input by all participants is welcomed, and a re-centering of the relationship on shared visions, whether policy goals, problem dimensions, or available solutions, is to be sought.

The growing use of networks to deliver programs and services underscores the lessons of the third generation. Implementation is often indirect with programs managed through complex networks of non-governmental providers such as non-profit organizations.^[1] Backward and forward mapping, effective communication, cooperative efforts, and increased trust and involvement are essential if implementation is to be successful.

CONCLUSION

The first generation of research on policy implementation produced fairly discouraging conclusions. Seldom, it appeared, were new governmental policies implemented as expected. Second-generation research sought to improve the situation by identifying solutions. However, the work was limited by an either/or definition of the problem: either top-down or bottom-up. This meant that the recommendations themselves were limited. The more recent third-generation research synthesizes the top-down and bottom-up approaches to produce more plausible explanations for why some policies are implemented successfully and others are not. It is not simply a matter of crafting a perfect piece of legislation for adoption. Nor is it enough to create a workable organizational structure

or to rally a cadre of supporters to the cause. Instead, successful implementation involves converting the potential energy of diverse participants into concerted, positive action. Policymakers are instructed to engage in forward and backward mapping. Implementation managers are encouraged to improve communication and foster cooperation, and to increase the levels of trust and involvement among implementers. On balance, as the use of networks grows, perhaps the most likely implementation pattern is one of strategic delay.

REFERENCES

1. O'Toole, L.J., Jr. Research on policy implementation: assessment and prospects. *J. Public Adm. Res. Theory* **2000**, 10 (2), 263–288.
2. Derthick, M. *New Towns In-Town*; Urban Institute: Washington, DC, 1972.
3. Pressman, J.L.; Wildavsky, A.B. *Implementation*; University of California Press: Berkeley, CA, 1973.
4. Mazmanian, D.A.; Sabatier, P.A. *Implementation and Public Policy*; Scott, Foresman: Glenview, IL, 1983; pp. 41, 42, 278–282.
5. Hjern, B. Implementation research—the link gone missing. *J. Public Policy* **1982**, 2 (3), 301–308.
6. Matland, R.E. Synthesizing the implementation literature: the ambiguity–conflict model of policy implementation. *J. Public Adm. Res. Theory* **1995**, 5 (2), 145–174.
7. Elmore, R.F. Forward and backward mapping: reversible logic in the analysis of public policy. In *Policy Implementation in Federal and Unitary Systems*; Hanf, K., Toonen, T., Eds.; Martinus Nijhoff: Dordrecht, 1985; 33–70.
8. Goggin, M.L.; Bowman, A.O'M.; Lester, J.P.; O'Toole, L.J., Jr. *Implementation Theory and Practice*; Scott, Foresman/Little, Brown: Glenview, IL, 1990.
9. Stoker, R.P. *Reluctant Partners: Implementing Federal Policy*; University of Pittsburgh Press: Pittsburgh, PA, 1991.
10. Scheberle, D. *Federalism and Environmental Policy: Trust and the Politics of Implementation*; Georgetown University Press: Washington, DC, 1997.

Policy Networks

Jack W. Meek

*College of Business and Public Management, University of La Verne,
La Verne, California, U.S.A.*

INTRODUCTION

Policy networks describe formal and informal social relationships among interested parties that form agreements to achieve individual and common goals in public arenas. Policy networks are viewed differently by political scientists interested in policy formation, by public administration scholars interested in policy implementation, and by sociologists interested in influence relationships among network partners. The influences of policy networks, and their great variety, provide evidence of social interaction that influence social behavior and influence public value. Some view these forms of associations as working alongside or even replacing formal government service and forming what has come to be described as governance. From another perspective, some forms of policy networks can be considered “dark” and there is a necessity to understand how these operate in addition to those policy networks that show promise.

POLICY FORMATION

Policy choices are expressions of social value traditionally thought to be developed by legislators who are representing the citizenry. In traditional theory, these choices are thought to be implemented by neutral public servants working in government bureaucracies. Political scientists have long observed that policy choices and governmental bureaucratic implementation are influenced by policy networks that are formed around the relationships and by participants closely involved in the policy choice. Those who decide (legislature), those who implement (bureaucracy), and those most affected (individuals, neighborhoods, industry or service) by policy form partnerships of interest that influence both the content and implementation of policy. These partnerships, referred to variously as subgovernments,^[1] as iron triangles or as cozy triangles are characterized by the cultivation of continuing relationships that engage in mutual exchanges that favor the needs of clientele groups agencies and representatives. Hugh Hecl^[2] broadened the frame of reference of these narrowly defined triangles to include a web of policy participants, with a variety of interests, that is

characterized as an “issue network.” Self-interest and exchange held issue networks together, but this form of policy influence was considered broader than the subgovernment form of relationship.

One meaning of policy networks is that policy formation is subverted or at least influenced by the central participants of the issue and that citizenry are kept far from playing any form of influential role in the process. Indeed, of concern about policy networks are the mutual benefits generated from political-administrative connections. No longer is there a neutral bureaucracy implementing public policy; it is now understood that bureaucrats are policy advocates. Networks replaced neutrality with self-serving outcomes of elected officials, selected clientele, and bureaucracy. On the other hand, policy networks can also be viewed as functional necessities of the modern nation-state where relationships of the most central players usefully orchestrate public will in competition with other interests in the state.

POLICY ADMINISTRATION

Public administration scholars have observed that public programs are increasingly produced and managed through various collectives.^[3] As networks, these collectivities include both governmental and non-governmental partners. For Laurence O'Toole, these “networks are structures of interdependence involving multiple organizations or parts thereof, where one unit is not merely the formal subordinate of the other in some larger hierarchical arrangement.”^[4] These networks are policy networks that are characterized by a “relatively stable relations between (different) governmental and (semi-) private organizations in which processes of policy making take place.”^[5] Thus policy networks warrant much attention, not only from the policy or political demand or “input” side, but also from the policy “output” or implementation side. And there are different positions on how to view policy network outputs and their impact on public governance.

One position, as evidenced in the work of Hugh Miller, is that an emergent characteristic of policy network (implementation) is administrative discretion

that approaches “political activism” or even political professionalism by civil servants.^[6] Such an outcome challenges the assumptions made by “progressive” public administration built around the principles of hierarchical control, scientific management, and neutral competence. The charge is that mutually supportive relationships spillover into the output/implementation side of policy where public administrators, motivated by a sense of agency expansion and self-promotion, become political administrators. The key issue is that administrators are hidden from direct public accountability and manipulate service design and delivery based upon a variety of reasons. It may very well be that administrators have captured and manipulated service implementation, which could serve many interests.

Another position among public administration theorists is to view the formation of networks as valuable tools in the implementation of public policy where administrative leaders face resource scarce environments.^[7] The size and nature of policy implementation networks range widely from linkages between two or more organizations to a formal coalition of organizations formally designed to carry out a broad mission. Networks, viewed through this frame of reference, are forms of collaborations designed to effectively carry out policy directives. The privatization movement effectively constructs a web of networks managed by government engagement that oversees policy administration.

NETWORK MANAGEMENT

The unique feature of policy networks is that they can be considered a third form of social relationships, the other two relationships are categorized as hierarchy and market.^[8] Whereas hierarchy and markets are usefully examined through the lens of superior-subordinate and rational self-interest, respectively, networks are much more elusive to understand. Network behavior is best matched with work on indeterminate behavior: participants come and go at will, there is no determinate pattern of leadership, power or hierarchy or leadership. Networks emerge and dissipate.

For public administration, accountability becomes a deep concern in policy administration networks. One study on the local management of economic development found that networks of multiple actors do represent an important form of governance. However, the management of networks adds to the tasks of public management and implies that both the capacity and policy making realm of cities becomes more decentralized and complicated.^[9] Indeed, if governance means a pronounced reliance on exchange and a de-emphasis of the role and strength of government, then serious challenges to equity and accountability may undermine the ability and legitimacy of government to govern.

In addition, the concern for policy networks is how roles are conceived as well as their concomitant expectations of rights, responsibilities, obligations, and duties that guide behavior.^[10]

The absence of hierarchy, the accountability vacuum, and the lack of role identity within networks have pushed public administration scholars to seek useful frames of reference that would guide public managers in determining what constitutes management cornerstones of policy networks. Robert Agranoff and Michael McGuire begin with examining the skills of the administrator and differentiating those required in traditional hierarchical public administration management (POSCORB) and those required of network management.^[11] The class of behaviors they found relevant in networks includes the following: activation, framing, mobilizing, and synthesizing. Activation refers to the identification and development of stakeholders that might be part of the network. Framing is the ability to establish rules of operation for the network. Mobilizing is the ability to coordinate and accomplish the goals and objectives of the network. Synthesizing refers to providing an environment when conditions are created to enhance network productivity. The combined effect of these skills is nearly “seamless in their applicability” and may result in outcomes that emerge and evolve rather than those that are targeted and achieved. Current research in management behaviors in networks is focusing on propositions based on a strategic-contingency logic, that program performance will vary based upon the problem context of the network.^[12,13] Drawing upon the examination of 12 network processes, Robert Agranoff has established 10 basic lessons in the management of networks (sharing burdens, sharing expertise, being creative, orchestrating agendas are examples), most of which reveal the fundamental need to be flexible within the fluid nature of network functioning.^[14] These may have been valuable skills in hierarchies, but it is in networks that these skills are preminent.

NETWORK PERFORMANCE

How well do these networks perform? Because policy networks involve many participants with multiple agendas, evaluating network performance becomes problematic, but necessary. One study^[15] examined publicly funded health, human service, and public welfare organizations from three levels (community, network, and organizational/participant) each of which has unique effectiveness criteria that often may be in conflict. The authors argue that while “service-delivery networks must be built and maintained at the organizational and network levels, overall network effectiveness will ultimately be judged by community-level stakeholders.”

To be successful, policy administration in networks calls upon different skills than those linked to successful administration in hierarchies. The very management of networks suggests the need for very different orientations and strategies of management.^[5] Reliance on collaboration from participants with different interests and who are accountable to different “publics” offers constraints in determining and achieving common goals. Based on a “theory of network management,” Meier and O’Toole found that managers who incorporated a networking management style (those who have greater interaction with environmental actors who are not direct line subordinates or superiors) had a significant impact on program performance and were able to influence program results.^[16] While this is perhaps a limited measure of managerial networking, it is a beginning in the establishment of what is effective in managing policy network.

GOVERNANCE

With the growth and evolving complexity of urban areas, policy networks have appeared and approach policy problems and solutions side by side with governmental bureaucracies. These networks are forms of governance. These networks are seen to be more responsive to the citizenry, in terms of participation and accountability, than the traditional governmental structures although these networks often include governmental structures. In fact, there is evidence that networks are forms of governance that have very enlightening features.^[17] Such features are more responsive to social problems that go beyond institutional boundaries set up by specific geographic jurisdictions that are represented by elected officials from that region. Administrative conjunction, the horizontal association of various networked public actors and the resultant administrative behavior of the network,^[18] best describes the administrative city-state in urban arenas today. Region-like collaboratives, such as councils of governments, regional transportation, and air quality boards, represent emergent forms of governance that engage in a web of system interaction that, as a whole, act as an urban regime of linked participants of mutual and competing interests. Indeed, R.A. Rhodes may be quite accurate when he states that these networks are “inter-organizational linkages (that) are the defining characteristic of service delivery.”^[19] How these networks are held accountable, and to whom, are powerful questions within this emerging dynamic of networked public administration. O’Toole reviews the role of bureaucracy and the emergence of networks in relation to the central democratic political norms of responsibility, responsiveness, and enhancement of political deliberation,

civility, and trust. He finds that networked public administration possesses many challenges, yet the lessons of such a status provide “both complications and opportunities to facilitate parts of the democratic ideal.”^[20] According to Erik-Hans Klijn and Joop F.M. Koppenjan, a central role should be played by government in the management of policy networks. It is their argument, that the government’s special resources and unique position can be used to represent the common interests of the public. This vital position means that government needs to play significant roles as network managers that arrange and facilitate interactive processes within communities.^[21] The authors defend this network approach to governance based upon the various theoretical foundations of the network approach and the emerging ability of that approach to explain program success and failure based on network determinants, including power, network management, and network structure. It would seem that much of the judgment about the impact of networks, as to how they influence the implementation of public policy, will rest on the nature of each network and how each network is managed.

CONCLUSION

Because policy networks are viewed differently by various authors and disciplines, it will be difficult in the near term to develop a coherent theory of networks. However, the work of Laurence O’Toole provides valuable parameters from which we can begin to address the development of network theory.^[4] O’Toole recommends that we explore the causal, conceptual, and descriptive dimensions of networks in order to improve our understanding of their role in public administration in a democratic society. LaPorte^[22] also offers three approaches to examining and understanding networks: 1) from within the network (the net rider) or those organizational units that traditionally function within the network system; 2) from above (the net thrower) or those who have a view of the entire network, such as a policy maker; and 3) from the side (the net puller) or entities that view a network not as an integral network participant but one that has cross-cutting interests or influence, such as a professional association or those interested in some form of reform. Each of these perspectives allows for different insights, different research traditions, and addresses different interests of the same network phenomenon. There have also been several advances toward a typology of networks framed mostly from the context of a local government perspective.^[7,9,15] Still others have felt that network theory for public administration and public policy is not likely to be advance on its own without considering networks as expressions of collective

action.^[23] Thus we have several valuable options from which to view policy networks. And, while most of the research on policy networks is focused on how networks and collaboration are quite positive, it is useful to note that networks can also have a dark side.^[24] Here the work of Jorg Raab and H. Brinton Milward is instructive: there are “dark networks” such as drug cartels, terrorist networks, arms smuggling that require response. Surprisingly or not, these dark networks share the same characteristics as “positive” or legal networks. The exception is that these networks rely on secrecy and the use of physical force. The key contribution of this finding is that network analysis is useful not only in examining positive networks, but “dark” networks as well. As we examine further the wonderful dimensions of policy networks as forms of governance, we can expect many different research strategies to inform us about how these entities are formed, how they behave, and what impact they will have on our lives.

REFERENCES

1. Cater, D. *Power in Washington*; Random House: New York, 1965.
2. Hecl, H. Issue networks and the executive establishment. In *The New American Political System*; King, A., Ed.; American Enterprise Institute for Public Policy Research: Washington, DC, 1978; 87–124.
3. Kettl, D.F. *Sharing Power: Public Governance and Private Markets*; Brookings Institution: Washington, DC, 1993.
4. O'Toole, L. Treating networks seriously: practical and research-based agendas in public administration. *Public Adm. Rev.* **1997**, *57*, 45–52.
5. Kickert, W.J.M., Klijn, E.H., Koppenjan, J.F.M., Eds.; *Managing Complex Networks: Strategies for the Public Sector*; Sage Publications: London, 1997; xvii pp.
6. Miller, H.T. Post-progressive public administration: lessons from policy networks. *Public Adm. Rev.* **1994**, *54*, 4.
7. Mandell, M.P., Ed.; *The Impact of Collaborative Efforts: Changing the Face of Public Policy Through Networks and Networks Structures*; Policy Studies Review: University of Tennessee Energy, Environment & Resources Center, 1999; Vol. 16, 1 pp.
8. Powell, W.W. Neither market nor hierarchy: network forms of organization. *Res. Org. Behav.* **1990**, *12*, 295–336.
9. Agranoff, R. Multinetwork management: collaboration and the hollow state in local economic policy. *J. Public Adm. Res. Theory* **1998**, *8*, 1.
10. Knoke, D. *Political Networks: The Structural Perspective*; Cambridge University Press: New York, 1990.
11. Agranoff, R.; McGuire, M. After the network is formed: process, power and performance. In *Getting Results Through Collaboration: Networks and Network Structures for Public Policy and Management*; Mandell, M.P., Ed.; Quorum Books: Westport, CT, 2001; 11–29.
12. McGuire, M. Managing networks: propositions on what managers do and why they do it. *Public Adm. Rev.* **2002**, *62* (3), 599–609.
13. Mandell, M.P. Network management: strategic behavior in the public sector. In *Strategies for Managing Intergovernmental Policies and Networks*; Gage, R.W., Mandell, M.P., Eds.; Praeger: New York, 1990; 29–53.
14. Agranoff, R. *Leveraging Networks: A Guide for Public Managers Working Across Organizations*; IBM Endowment for The Business of Government: Arlington, VA, March 2003.
15. Provan, K.G.; Britton, M.H. Do networks really work? A framework for evaluating public-sector organizational networks. *Public Adm. Rev.* **2001**, *61*, 4.
16. Meier, K.J.; O'Toole, L.J., Jr. Public management and educational performance: the impact of managerial networking. *Public Adm. Rev.* **2003**, *63* (6), 689–699.
17. Meek, J.W.; Schildt, K.; Witt, M. The future of local government administration. In *Local Government Administration in a Metropolitan Context*; Frederickson, H.G., Ed.; International City/County Management Association: Washington, DC, 2002; 145–154.
18. Peters, G.B. Governance without government? Rethinking public administration. *J. Public Adm. Res. Theory* **1998**, *8*, 223–243.
19. Rhodes, R.A. Foreword. In *Managing Complex Networks: Strategies for the Public Sector*; Kickert, W.J.M., Klijn, E.-H., Koppenjan, J.F.M., Eds.; Sage: Thousand Oaks, 1997; xii pp.
20. O'Toole, L. Implications for democracy in a networked bureaucratic world. *J. Public Adm. Res. Theory* **1997**, *7*, 443–459.
21. Klijn, E.-H.; Koppenjan, J.F.M. Public management and policy networks. *Public Manage.* **2000**, *2* (2), 135–158.
22. LaPorte, T.R. Shifting vantage and conceptual puzzles in understanding public organization networks. *J. Public Adm. Res. Theory* **1996**, *6*, 49–74.
23. Carlsson, L. Policy networks as collective action. *Policy Stud. J.* **2000**, *28* (3), 502–520.
24. Raab, J.H.; Brinton, M. Dark networks as problems. *J. Public Adm. Res. Theory* **2003**, *13* (4), 413–440.

Policy Sciences Approach

Peter deLeon

Christine R. Martell

Graduate School of Public Affairs, University of Colorado at Denver and Health Sciences Center, Denver, Colorado, U.S.A.

INTRODUCTION

This article describes the policy sciences, which have a unique approach to problem solving. The first part develops the evolution of the underlying assumptions of the policy sciences, while the second sets out an analytic framework derived from the policy sciences as a means of viewing policy decisions. Finally, a brief hypothetical example of how the policy sciences approach would be utilized in the “real world” is presented, based on a major highway and light rail construction project in Denver, Colorado.

THE EVOLUTION OF THE POLICY SCIENCES

The development of the policy sciences approach has generally been ascribed to the works of Harold D. Lasswell^[1–3] and his colleagues, soon after the end of the Second World War. The stated purpose of the policy sciences was to improve the quality of information offered to government and, by so doing, improve the quality of governance. But, the ontogeny of advice to the ruler goes back to virtually the beginning of recorded history, in which people wondered as to the application of knowledge to the public sector, to much the same end.^[4] Certainly the priests of antiquity were actively involved in political decisions, as kings would ask them to divine (and possibly sway) the “mood” of the gods prior to battle or even a royal nuptial. Later, of course, the princes of the church moved into political decision making with fewer concerns for the deities and more for their own secular livelihoods.^[5]

However, by the eighteenth century, the Age of the Enlightenment infused Western Civilization with the absolute certainty that, through study and reflection, a rational “truth” covering a wide range of subjects, including politics, could be discerned and acted upon. The history of the Progressive Era in the United States at least partially mirrors the purported marriage of good science and good government,^[6] even if politics and their administrations were somehow considered to be mutually exclusive.

The sociopolitical history of the twentieth century contributed in a sustained way to the movement of

higher education (i.e., universities) into the realm of social application. The First World War saw the emergence of psychology as a means to help winnow recruits for the American Expeditionary Force. The worldwide and sustained traumas of the Great Depression brought economics into virtually every government office. And, the horrors of World War II fascism made it apparent that governments could easily be used for ill purposes as well as good. This realization was especially clear for Lasswell, who spent the wartime years studying propaganda in the Library of Congress.

THE POLICY SCIENCES FRAMEWORK

Beginning in the late 1940s, Lasswell began to sketch out the thrust of the policy sciences, with three clear and abiding components that differentiated them almost radically from the traditional research canons as practiced in higher education; specifically, the policy sciences were meant to be:

- *Problem-oriented:* They were intended to address problems of significant social import rather than those of strictly academic concern; the policy sciences would not address the number of angels dancing on the head of any given pin, but, rather, the causes and alleviations of social unrest and inequalities. In Lasswell’s words (Ref.^[2], p. 14), “The Policy [Sciences] approach does not imply that energy is to be dissipated on a miscellany of merely topical issues, but rather that fundamental and often neglected problems, which arise in the adjustment of man in society are to be dealt with.”
- *Multidisciplinary oriented:* Lasswell^[2,3] saw the problems of the “real world” as having multiple “explanations,” such that in almost all cases, the application of a single discipline—be it law or psychology or sociology or economics—would be an unnatural constraint, be excessively narrow in scope, and therefore, in all likelihood, would fail to reveal the complete scope of the problem or its resolution. This tenet, of course, violated the entire departmental structure of the contemporary

university, which by this time was disciplinarily focused on narrow research topics.

- *Explicitly normative:* The effects of the Enlightenment were most apparent in the exposition of and general adherence to logical positivism, a “scientism” that neglected the role of norms and values and ethics in social decisions (Social Darwinism being one example). This was the situation in the social sciences, such as economics, psychology, and behavioralism in political science. Yet, Lasswell was certain that norms had an explicit place in policy making and needed to be emphasized in the policy sciences.

The third of these hallmarks was explicitly spelled out by Lasswell as he and Abraham Kaplan (Ref.^[7], p. 15) articulated that the policy sciences were designed to provide “intelligence pertinent to the integration of values realized by and embodied by interpersonal relations [such as] human dignity and the realization of human capacities.” More specifically, the next year, Lasswell wrote that “the policy sciences of democracy [were] directed towards knowledge needed to improve the practice of democracy” (Ref.^[2], p. 15), thus enshrining the normative concept of the “policy sciences of democracy.”

Two additional features of the policy sciences were their emphases on contextuality and the policy process. The first referred to the concept that no policy research could be conducted without a thorough understanding of the inherent contextuality of the problem; in other words, what were the conditions leading to the recognition of the problem, and in what environment did the problem exist? While this concept is certainly important in a world in which the policy sciences emphasize social problems, it lends itself to a fundamental, endemic concern of the policy sciences: if all knowledge is dependent upon the overall environment in which a problem exists, that naturally relegates a general *theory* of the policy sciences to a secondary position.

Lasswell also indicated a framework for understanding the events of (what he termed) the “policy process.” Formally posed in 1956, it has become the touchstone of most research in the policy process; many authors (e.g., Brewer and deLeon,^[8] Anderson,^[9] and Jones^[10]) have formulated their texts around the process, and numerous others have written on the various stages. As posed by Brewer and deLeon,^[8] the stages are as follows:

- *Initiation:* The recognition of a problem, an understanding of what goals (e.g., political, economic, and normative) are to be sought, and the preparation of possible options to attain these goals within their constraints.

- *Estimation:* The balanced assessment of the “costs” and “benefits” (broadly defined) characterizing the various policy options to be delivered to government policy makers.
- *Selection:* The stage at which decisions addressing the problem are made by the policy makers (generally, but not always and surely not exclusively) on the prior analytic work.
- *Implementation:* The stage in which the chosen option is put into practice, usually by administrators, but, more recently, through the non-profit sector, such that the programs have an effect on their intended recipients.
- *Evaluation:* The stage in which the program is assessed from a retrospective vantage point, its various aspects evaluated, and, when and where necessary, changes are made in the program and its implementation.
- *Termination:* In some cases, this is the final stage in the policy process, in which a program (for a number of reasons) is seen to have lost its effectiveness and is therefore terminated.

It is important to recognize that this framework (some have called it a “heuristic,”^[11]) is not presented as a formal “model” and does not assume any predictive capabilities. Instead, it is offered as a framework for understanding and structuring a particular problem for analysis, and, as such, serves a significant purpose.^[12] Moreover, its stages are clearly interactive with multiple feedback loops between them, rather than posing as a linear, lock-step progression.^[8]

AN APPLICATION OF THE POLICY SCIENCES APPROACH

Denver, Colorado, like so many American cities at the turn of this century, is experiencing severe transportation problems. Bisected by a major interstate (I-25), its citizens suffer, by their standards, long and crowded commutes, especially from south of the city, seemingly akin to frustrations found in Los Angeles, New York, and other major metropolises. At the same time, the development versus conservation (in Denver, known most cynically as “urban sprawl”) battle is in full tilt, with many political lances being broken over the resolution. In this fray, the newly elected Governor of the State offered what was to become known of T-REX (for TRansportation EXpansion), in which the Governor proposed a widening of I-25 south of Denver to eight to 10 lanes for approximately 19 miles, and a parallel construction and extension of Denver’s existing light-rail transit system, all at a promised cost of \$1.67 billion and a (projected) time frame of seven to

eight years.^[13,14] The State Assembly approved the entire T-REX package.

The “traditional” means of studying such a massive city works undertaking would have been little more than a textbook exercise in civil engineering (i.e., a mile of poured concrete times the number of lanes times the number of miles, with an unspecified “fudge factor,” in which time and costs were historically and heroically underestimated; witness Boston’s current “big dig” fiasco as evidence). Under contemporary standards, this process has been vastly complicated, as “new” conditions—such as endangered species, noise pollution, how it would be paid for (in this case, highway trust funds) and, of course, litigation and the delays entailed—have to be taken into account. From the perspective of the policy sciences, however, T-REX is naïve at best, and a disaster at worst.

The policy sciences pay close attention to the goals underlying a project *before* designing the project. For instance, if the goal were to reduce commuting time, then widening the interstate is only one option; others, such as developing multiple civic and commercial centers (especially with the possibilities of electronic communication) would also reduce commuting time and, probably, at a fraction of the cost. In the specific case of T-REX, the clear purpose is to expedite travel times along the I-25 corridor by expanding the freeway, but with little understanding of the long- and short-term consequences.

Assuming for the sake of the argument that the proposed T-REX plan works as advertised, its response might well be short-lived in its effectiveness. When more automobiles can be accommodated with the newly widened I-25, the housing and commercial real estate markets will again feature residential and business locations along the southeast corridor (now advertising a “speedy commute” enticement), thereby selling more property and dwellings, and thereby raising the number of residents and, concomitantly, the volume of vehicles using the corridor. In time, I-25 will become as crowded as is currently the case, and over a billion dollars will have been spent to little avail (excepting a short period of time before the real estate markets adjust). Likewise, more cars will be able to access downtown Denver’s commercial areas (even if the individual driver’s commute time is no faster), without much attention being paid to the ability of downtown to absorb the new visitors. For instance, will there be sufficient parking, or what if commercial builders opt for suburban expansion? If these were the only problems, T-REX might be understandable, but in order to build the new system, many homes and businesses will have to be literally razed, and those houses presently set back from the current I-25 route will become cheek-to-jowl to the noisy interstate. None of these are especially attractive futures.

In contrast, the policy sciences approach would have insisted on a wider variety of perspectives being examined before a contract was signed. In addition to civil engineering, these perspectives would have included, for example, urban planning, community building, sociology, political science, environmental management, and ethics.

An urban planning perspective would consider a variety of demographic shifts as well as a plan for directed growth of the city. While transportation planning (i.e., the nexus of planning and civil engineering) more narrowly focuses on moving people from point A to point B in the most expedient way, urban planners would also consider the growth consequences of highway expansion on the community and might investigate whether the areas within the expansion zone will be able to support services for the projected population increases. Moreover, urban planners actively involve the affected citizens in the planning process to determine such questions as: Will the area have a strong enough tax base, be able to accommodate new schoolchildren, be able to provide clean water, provide recreational facilities, process sewage, and collect refuse, in short, be a comfortable “home” to its businesses and residents?

In conjunction with this, a community building perspective would question whether the expansion of transportation would destroy existing communities and whether it would lead to building new and viable communities. Already, strong opposition by communities to be affected by the T-REX construction clearly signals that the project subverts the value of established community to that of future real estate growth and the concomitant perceived ills of urban sprawl.

The sociology perspective might consider the social forces to explain group behavior. What are the reasons explaining the need for expansion in such a direction? Is there a reason why people are choosing to live 10 miles outside the city when there is available housing within the city limits? Is there an urban problem that can be resolved to ameliorate the volume of suburban-to-urban commuters? A sociologist also might view the process through a series of lenses, such as gender, race, and economic class, lenses that T-REX has studiously avoided.

A political science perspective would consider the relative affect of interest groups, such as the “highway” lobby and the home owners’ association, to understand the resulting plan, specifically, the political pressure for and against transportation expansion. Political scientists would be especially concerned with the manner in which the decision was made, and who were the relevant participants. From a normative perspective, political science would ask why this particular route was chosen over alternative placements, as well as who is being displaced (or advantaged).

Supporters of environmental management might look at the expansion from the view of its affect on the environment. Would the highway be expanded into areas of fragile land? What repercussions would expansion have for urban sprawl and the use of soil and water? Would, for example, suburban development encourage the planting and maintaining of water-hungry green lawns in a desert climate (that is, its effect on urban water tables)? An environmentalist would consider the consequences that development policies have for natural resource management. In short, the environmentalist would argue strongly for the conservation side of the conservation/development coin, or what is broadly known as “sustainable development.”

An ethical perspective might question whether engaging in transportation expansion is in light of the alternatives and question why movement from point A to point B is so important when alternatives exist. Ethicists may suggest that transportation expansion is normatively problematic because of the potential negative repercussions for social mobility, environmental damage, and economic integration. Instead, they may propose solutions that are considered more justified, but in any case, alternatives will be proffered. For instance, ethicists might propose to develop a system of regional minicenters rather than a Denver-centric spoke-and-wheel metropolitan area, or they may abandon transportation expansion altogether in favor of a series of telecommute centers or education programs.

Considering these (and similar) approaches, it becomes apparent that the T-REX expansion deals with a great more than the technical engineering question of laying the road and a light-rail transit system. The policy sciences approach would more broadly recognize that the issue of traffic congestion entails many more issues than simply “moving bodies,” that is, there are many aspects to this, and most, problems. The perspectives above need not be mutually exclusive. For example, a sociologist’s view may complement the views of community builders or planners, and a planner may resonate with an environmental manager. Regardless of how well perspectives converge, viewing

the T-REX situation from different lenses illustrates how a policy problem is multifaceted, with each perspective introducing a host of issues. The more that the policy sciences approach introduces a series of related issues, the more policy goals become multidimensional, and the acceptable solutions necessarily become multidisciplinary.

REFERENCES

1. Lasswell, H.D. *Power and Personality*; Norton: New York, 1949.
2. Lasswell, H.D. The policy orientation. In *The Policy Sciences*; Lerner, D., Lasswell, H.D., Eds.; Stanford University Press: Palo Alto, CA, 1951; 3–15.
3. Lasswell, H.D. *A Pre-View of Policy Sciences*; American Elsevier: New York, 1971.
4. deLeon, P. *Advice and Consent*; The Russell Sage Foundation: New York, 1988.
5. Goldhamer, H. *The Advisors*; American Elsevier: New York, 1978.
6. Stillman, R.J., III. *Preface to Public Administration*; St. Martin’s: New York, 1991.
7. Lasswell, H.D.; Kaplan, A. *Power and Society*; Yale University Press: New Haven, CT, 1950.
8. Brewer, G.D.; deLeon, P. *The Foundations of Policy Analysis*; Brooks/Cole: Monterey, CA, 1983.
9. Anderson, J.E. *Public Policy Making*, 2nd Ed.; Holt, Reinhardt, & Winston: New York, 1979.
10. Jones, C. *An Introduction to Public Policy*, 3rd Ed.; Wadsworth: Belmont, CA, 1984.
11. Sabatier, P.A. Policy change over a decade or more. In *Policy Change and Learning*; Sabatier, P.A., Jenkins-Smith, H.C., Eds.; Westview Press: Boulder, CO, 1993; 13–39.
12. deLeon, P. The stages approach to the policy process: what has it done? Where is it going? In *Theories of the Policy Process*; Sabatier, P.A., Ed.; Westview Press: Boulder, CO, 1999; 19–32.
13. T-REX. 2001. <http://www.trexproject.com>; reviewed June 2001.
14. Anonymous. What the future hold: Transportation expansion project. *Denver Post* May 20, 2001, (pp. A1, A17–A18).

Policy Termination

Mark R. Daniels

Department of Government and Public Affairs, Slippery Rock University of Pennsylvania, Slippery Rock, Pennsylvania, U.S.A.

INTRODUCTION

Policy termination has been defined by Garry Brewer and Peter deLeon as “the deliberate conclusion or cessation of specific government functions, programs, policies, or organizations.”^[1–3] Elaborating upon this definition, another termination researcher explains that “termination is premeditated behavior, with the intent of ending a particular public organization or policy.”^[4] Both of these definitions do not recognize changes in policy emphasis or jurisdiction, which may be organizational attempts to redirect activities and to justify existence, as policy termination. In addition, both of these definitions define policy termination in a broad sense, including both organizations and their programs and policies under the same golf-course-size umbrella of policy termination. This broad definition leads to confusion because some termination researchers, notably Herbert Kaufman, have studied organization termination as a phenomenon distinct from policy termination.^[5,6]

TERMINATION RESEARCH

Eugene Bardach edited the first collection of termination studies in a special issue of the journal *Policy Sciences* in 1976, and two additional collection of articles were published in journals in 1997 and 2001.^[7–9] Bardach argues that termination is actually a special case of policy adoption: The adoption of policy A requires the elimination or curtailment of policy B. The uniqueness of this form of policy adoption lies in the vested interests of those who oppose the termination of policy B. Bardach observes that when termination does occur, it does so with either a bang or a whimper. In the former case, resistance to the termination of policy B is so intense that when it finally does occur, it does so in an abrupt, forceful fashion. In the latter case, termination occurs as a result of a long-term decline in the organizational resources that supported policy B.

Books on Policy Termination

In 1997, two books, both providing a comprehensive treatment of policy termination, were published: *An*

Operational Approach to Policy Analysis: The Craft and Terminating Public Programs: An American Political Paradox.^[4,10] *An Operational Approach to Policy Analysis* attempts to operationalize the craft of policy analysis, and each chapter provides step-by-step instructions on different approaches to policy analysis and ends with a summary section that provides a check list for policy analysts. This format is very innovative and addresses the needs of policy analysts for a how-to book. Its format and style definitely fill a gap in the policy sciences literature. Chapter Five, entitled “Implementing Termination,” follows the general format of the book and provides step-by-step instructions based on published research. This chapter makes two very original contributions to understanding policy termination. First, the Defense Base Realignment and Closure Commission (BRAC) is given as an example that demonstrates the importance of agenda control. Second, they introduce, for the first time, the concept of ethics in the termination process. The idea that a policy analyst has an ethical obligation to citizens and to the political system, in general, introduces an additional dimension to the discussion of policy termination.

Terminating Public Programs attempts to respond to the three tasks that face policy termination researchers as identified by deLeon: 1) to compile cases on termination; 2) to acquire an understanding of the factors leading to termination and build theories; and 3) to develop strategies and tactics that lead to successful termination. The book offers seven conclusions about the implementation of policy termination (see [Table 1](#)).

First, termination rarely has economic justification. Termination usually costs more than policy continuation, at least in the short run. Termination is usually a result of political rather than analytical decision making. Political values and ideologies often pave the way for termination, instead of economies or efficiencies. There is general agreement among researchers that while cost-cutting rhetoric may be used as a tactic by terminators, in reality, no short-term savings result from termination.

Second, termination is highly political and hard to achieve. The American political system is characterized by multiple access points to government. Affected interest groups can gain access to government

Table 1 Seven conclusions about policy termination

-
1. Termination rarely has economic justification.
 2. Termination is highly political and hard to achieve.
 3. Termination requires co-optation of opponents.
 4. Termination often involves changing ideologies.
 5. Termination is followed by rebirth.
 6. Successful termination is difficult to predict.
 7. Termination is an American political paradox: Everyone supports it, everyone opposes it.
-

decision-making processes through legislative hearings, the enactment of legislative bills, the executive branch, the bureaucratic apparatus of government, the courts, and even through the media. These multiple points of access provide many opportunities for organized groups to oppose changes in specific government programs. Any attempt to terminate a government program, policy, or organization is potentially at the mercy of the entire political system. The reason termination is so difficult to achieve is because it is essentially political in nature. Termination is hard to accomplish within an incrementally oriented political system.

Third, termination requires co-optation of opponents. Termination is possible only through determined, prompt action by strong political administrators who are willing to engage, if necessary, in political co-optation of opponents. Political leaders of successful terminations have successfully blocked or cut off the political access points of opposition groups.

Fourth, termination often involves changing ideologies. Successful termination is often the result of a change in the application of certain principles, or assumptions, about what services are appropriate for government to sponsor or how these services should be delivered. Because ideologies can be neither proved nor disproved, the application of ideology to the continued existence of a program does not involve issues of performance, effectiveness, cost, or economies. Instead, termination in this case is conducted on the basis of a change in principle.

Fifth, termination is often followed by rebirth. Termination appears to be a journey more than a destination, and many programs come back to life after the terminators have left the political scene. Those programs claimed by death face neither heaven nor hell, but a potential after life. There is life after death, at least in the public sector.

Sixth, successful termination is difficult to predict. The timing of termination—that is, when an agency will actually die—is something of a mystery. The interaction of social and economic events over time may create changes in an organization's environment beyond which the organization can respond effectively.

If the organization cannot respond effectively and adequately to its changing environment, the organization becomes terminally ill and will soon die. This interaction over time is so subtle that it escapes the detection and measurement by researchers. Termination researchers do not predict terminations, but instead conduct autopsies: they study the demise of programs, policies and organizations after they have been terminated.

Seventh, termination is an American political paradox: everyone supports it, and everyone opposes it. The termination of programs, policies and organizations takes place within the constraints of democratic values such as citizen participation, political representation, and old-time pork barrel politics. American government is a pluralistic democracy in which multiple interest groups, representing different and conflicting public opinions, can take advantage of multiple access points into the political system. Opposition to terminating public programs, policies and organizations is not only expected but is a political right of those adversely affected.

A Termination Process Model

Susan E. Kirkpatrick, James P. Lester and Mark R. Peterson have made a significant contribution to the study of policy termination with an article that proposes a process model of termination based on their extensive review of literature.^[11] Fig. 1 presents their model, called “A Process Model for the Termination of Public Goods.”

Under the heading “Inherent Characteristics,” the *raison d'être*, or reason for being, refers to the perceived mission of the organization or policy and the legitimacy that the mission is given in the minds of the public and political leaders. According to the public involvement literature references in their article, any organization or policy that is not seen by the public as addressing some important public problem is living on borrowed time. Longevity has been identified as an important factor in the ability of an organization to survive and to avoid termination. Mature organizations are most immune to termination. Under the heading “The Political Environment,” the prevailing ideology refers to the importance of ideological support for a program that may shift with political changes in leadership. The size, strength, and determination is also an important political variable involved in termination. Another factor is the powerful allies that may assist either termination or antitermination efforts. The opportunity for compromise possibilities is another variable in preventing termination. Finally, speed is important for would-be terminators who need to act swiftly in order to keep one step ahead of

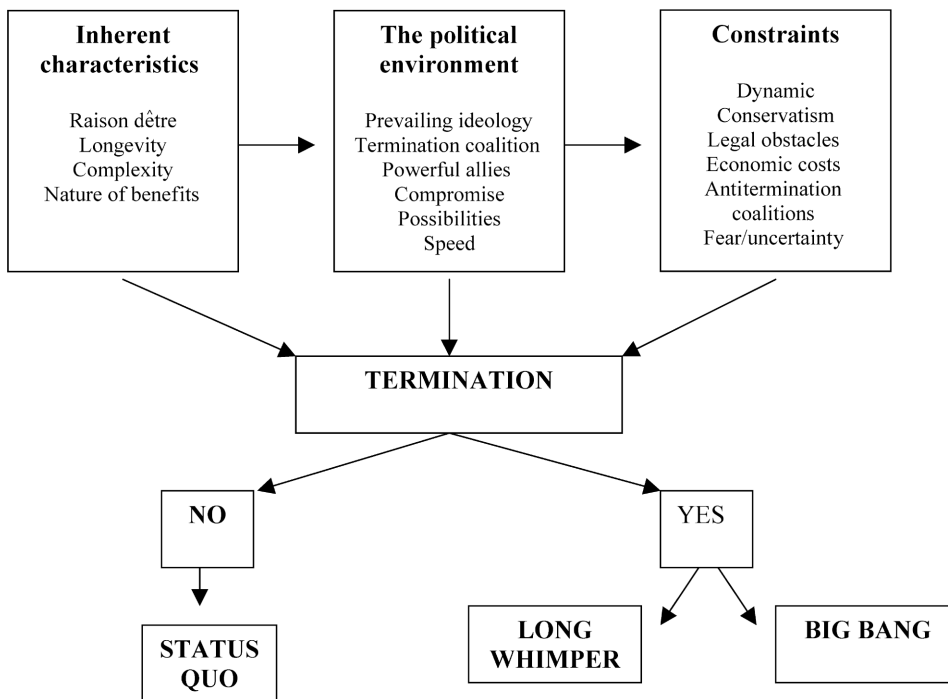


Fig. 1 A process model for the termination of public goods. (From Ref.^[11].)

termination opponents. Under the heading “Constraints,” the termination barriers of antitermination coalitions, dynamic conservatism, legal obstacles, and start-up costs are taken from the work of Peter deLeon. Antitermination coalitions are composed of constituents, both from the public and elected officials, who oppose the termination of a particular organization or policy. Dynamic conservatism refers to the ability of organizations and policies to adapt to changing environments and ensure their survival. Legal obstacles are efforts to block termination made through the legal process, for example, obtaining an injunction. Start-up costs refer to the cost of ending one program and losing the sunk costs already invested in the program. Finally, fear and uncertainty refer to the apprehension that administrators and political leaders have concerning the changes that may occur as a result of the termination of a program.

Their effort to develop this framework is an important contribution to the study of policy termination. Similar to all other discussions of policy termination, their analysis of policy termination is retrospective. While autopsies are useful in determining the cause of death, the topic of policy termination continues to lack a framework that is able to predict termination.

CONCLUSION

In his introductory article for the first symposium on policy termination, Eugene Bardach referred to policy

termination as an “underattended issue.” The so-called balancing of the federal budget and the elimination of the federal deficit during the late 1990s has turned popular political attention away from cutting government programs. For example, conspicuously absent from the 2000 presidential election were plans voiced by any of the candidates to eliminate agencies or programs, or reform government through termination of outmoded or inefficient policies. Nonetheless, as an academic topic, the factors surrounding termination are only beginning to be identified, and models of the termination process are beginning to emerge. What is lacking is a predictive equation that regresses the independent variables connected to termination and provides explanatory power with little error. While we are not yet at the point of making accurate predictions about the event of termination, recent research has brought us further down the path with new case analyses and models.

REFERENCES

1. deLeon, P. Public policy termination: an end and a beginning. *Policy Anal.* **1978**, *4*, 379–386.
2. deLeon, P. A theory of policy termination. In *The Policy Cycle*; May, J.V., Wildavsky, A.B., Eds.; Sage: Beverly Hills, CA, 1978.
3. Brewer, G.; deLeon, P. *The Foundations of Policy Analysis*; Dorsey Press: Homewood, IL, 1983.
4. Daniels, M.R. *Terminating Public Programs: An American Political Paradox*; M. E. Sharpe: Armonk, NY, 1997.

5. Kaufman, H. *Are Government Organizations Immortal?* Brookings: Washington, DC, 1976.
6. Kaufman, H. *Time, Chance and Organizations: Natural Selection in a Perilous Environment*; Chatham House: Chatham, NJ, 1980.
7. Policy Sci. **June 1976**, 7 (2), 123–243 (entire issue).
8. Daniels, M.R., Ed.; *Symposium on Public Policy and Organization Termination*; International Journal of Public Administration; Marcel Dekker, Inc.: New York, December 1997; Vol. 20, 2042–2212.
9. Daniels, M.R., Ed.; *Special Issue on Policy and Organizational Termination*; International Journal of Public Administration; Marcel Dekker, Inc.: New York, March 2001; Vol. 24, 247–339.
10. Geva-May, I. *An Operational Approach to Policy Analysis: The Craft*; Kluwer Academic Publishers: Boston, MA, 1997.
11. Kirkpatrick, S.E.; Lester, J.P.; Peterson, M.R. The policy termination process: a conceptual framework and application to revenue sharing. *Policy Studies Review* Spring 1999, 16 (1), 209–236.

Policymaking in Local Government

Jeremy L. Hall

*Department of Government, University of Alabama at Birmingham,
Birmingham, Alabama, U.S.A.*

INTRODUCTION

The public policy-making process is shaped by formal and informal institutions, rational actors positioned by values, and issues derived from the decision-making context. Local policy making follows this model and, as such, is quite distinct from state and national policy making. This essay explores each of these constructs to shed light on the process and products of local policy-making efforts.

LOCAL GOVERNMENT STRUCTURES

Local jurisdictions derive their existence and authority from sovereign states and are, consequently, only able to exercise such discretion as their parent states grant by law. They are, in essence, corporate entities. At opposite ends of a spectrum of state-granted discretion to local governments are Dillon's rule and Home rule. Dillon's rule—the traditional restrictive view of local government—provides that local governments can exercise three kinds of power: 1) those expressly granted in words; 2) those necessarily or fairly implied in the powers granted; and 3) those powers that are essential to the government's declared purpose (indispensable, not merely convenient).^[1] The doctrine is named after Judge Dillon, who authored it as Chief Justice of the Iowa Supreme Court in 1868.^[2] While some states are very restrictive in their grants of power to localities, others follow a model of Home rule, granting local governments broad discretion to self-determine policy where not otherwise restricted by state or federal mandate. Home rule or Charter governments are governed by a charter that acts much like a constitution, establishing boundaries within which local governments can exercise policy discretion.

Home rule counties are granted much broader authority than their non-Home rule counterparts, though they are still constrained by statute. For example, Knox County, Tennessee, has a unified city/county government operating under a home rule charter. In 1994, a referendum was held by Knox County voters to impose term limits on county-elected officials. The term limit requirement was not observed, however, and a court challenge led to the January 2007 Tennessee Supreme Court decision, recognizing

the charter's validity and vacating a number of county offices held by re-elected individuals, covered by the term limit amendment to the charter.^[3]

Related to the Dillon's rule/Home rule dichotomy is the structure of local general-purpose governments. Counties or parishes collectively occupy all of the territory in each state and act as the traditional administrative arm of the state. Municipalities (towns and cities) provide urban services to more densely populated incorporated areas. Cities operate under one of three general structures: the mayor-council form (with either a strong or a weak mayor); a council-manager form; or a commission form. The nature of these structural differences articulates the available entry points for policy change. And local government structures may vary considerably within a given state; Kentucky, for example, includes cities with council-manager, commission, and mayor-council forms, and the power and duties of the mayors also vary considerably among cities with the latter form of government.^[4]

Mayor-council systems of the strong mayor variety (where the mayor exercises budget authority, controls the council agenda, and possesses a veto) represent one extreme. In this setting, the council is the deliberative legislative body that sets policy (subject to a mayoral veto), and the mayor is the executive responsible for implementing and managing policy. Local policies are often so narrow in scope that there is little discretion to delegate, and, as such, administrative policy making is limited at the local level. This is not meant to imply that street-level bureaucrats in local government do not exercise bureaucratic discretion—police officers, social workers, and others certainly do. At the opposing end of the spectrum, council-manager governments lack a bona fide executive. So, the legislative council sets policy and directs a city manager (who serves at the council's discretion) to carry out policy and oversee the administration. Although cities are discussed here, similar distinctions exist in county government. A legislative body of representatives (referred to as a fiscal court in some states) and a separately elected county executive (sometimes called a judge executive) corresponds to the mayor-council system in cities, and a county manager corresponds to the council-manager system in cities. Systems that employ administrative managers in lieu of an elected executive are considered professional or reformed.

An additional distinction of local government structure is the divided executive. Whereas city mayors typically oversee all administrative operations, county executives are divided. Growing out of the Jacksonian ideal of specialization in government, county governments typically include a judge or other executive who presides over the fiscal court, a county clerk, sheriff, coroner, and various additional offices, including property valuation administrator, surveyor, attorney, and so forth. The result is increased potential for divided political control of local government operations, but the degree of specialization within each position limits the policy-making authority of any one individual. Moreover, these elected officials are often little more than administrative managers, carrying out the administrative duties delegated them by the state (including tax collection, issuing licenses, recording vital records, and similar functions). In these divided executive systems, real policy leadership ability rests with the judge executive or county executive and the legislative body.

The third branch of government—the judiciary—is also unique at the local level. Judicial preferences are expressed at the local level through the electoral system. Unlike appointed federal judges, local judges are elected to office, although usually in non-partisan elections. Judges who best represent local residents' views are elected to office, and they are then able to exercise discretion within the bounds of state law.

SPECIAL DISTRICTS, JURISDICTIONAL OVERLAP AND POLICY FRAGMENTATION

In addition to general-purpose governments, local policy making is further fragmented by the presence of single-purpose special district governments or public authorities. Special districts were created in large numbers during the 20th century, and non-education-related special districts have been the fastest growing form of local government. Special districts serve multiple purposes, among which are the professionalization of government services, a movement to make services more business-like and self-sustaining through service fees, and, perhaps, most importantly, movement of debt for capital off general fund budgets of general-purpose governments. Moving debt off-budget enhances government credit ratings and avoids debt limit requirements.

Special districts are governed by boards of directors or councils either appointed by general-purpose governments (utility boards, for example) or elected to office (school boards, for example), giving them the ability to operate autonomously under a corporate model. They have their own budgets and often have their own revenue streams derived from service fees

or earmarked tax receipts. Such districts include schools, libraries, utilities, conservation, economic development, and so on, giving each distinct policy arena its own institutional presence. They share many characteristics with cities and counties and are integrally involved in numerous substantive policy areas.^[5] Consequently, actors in each policy field and citizens interested in participating in that substantive policy process face very specialized formal institutions. The specialization may add simplicity to governance, but irregular and overlapping jurisdictional boundaries in the policy environment also lead to significant confusion for citizens seeking public officials responsive to their concerns—they often find it a daunting task to locate the accountable person with proper authority to act on their particular concern.

Fragmentation within urbanized areas has become common, as the promulgation of special districts and incorporated municipalities has led to overlapping or complex service areas.^[6] Unique communities with unique tax structures and service offerings appeal differently to different individuals. Local policy discretion provides the opportunity for cities or communities within a metropolitan area to differentiate themselves, thereby providing the bundle of services most desired by their residents, and setting the stage for savvy residents to relocate to areas that maximize their utility.^[7] Fragmentation produces choice. However, the overlapping nature of the fragmented system also creates confusion, and the cost of relocation means that the real effects of government competition are realized with either new residents or movers within an area, making locational choices.^[8]

The special district movement contributes to the “hollowing out” of the administrative state, which refers to the movement of service provision out of government agencies more generally. Contracting out to private and non-profit organizations has become an important method of local service provision. For example, waste management is now commonly contracted to private firms, as is road maintenance, snow removal, and various other public services traditionally performed by local governments. Local governments act as a source of funds, with services performed by the lowest, or in some cases the most qualified, bidder.

As one considers governments of smaller size, problems and issues that were internalized at higher levels of government become external and are more likely to fall outside of any one local government's sphere of responsibility or influence. Thus, policy solutions at the local level are dependent on coordination through regionalism and interlocal agreements. To this end, states have created regional entities to coordinate local planning. Regional planning commissions, area development districts, and similar entities bring together elected officials from the region on their

governing boards. These entities then oversee regional planning efforts, facilitate coordination among peer governments in addressing regional problems (such as growth, transportation, environment, conservation, and economic development), and provide technical assistance to small local governments that lack capacity to solve their problems independently. For example, traffic problems and transportation require coordination and agreement between jurisdictions to prevent bottlenecks and congestion at jurisdictional boundaries. In the case of waste management, economies of scale can be gained by coordinating local efforts; specifically, one large regional landfill can be operated much more efficiently than several independently operated county dumps. In some instances, local governments have even coordinated economic development initiatives, such as regional industrial parks, that pool resources in attempts to capture scarce benefits. Economic development is a traditionally competitive policy field, in which local governments compete with one another to lure firms to their area.

AREAS OF LOCAL POLICY SALIENCE

Economic development is one of the predominant policy domains at the local level.^[9] From planning and zoning decisions, redevelopment and revitalization efforts, and industrial parks and other recruitment efforts to utility and transportation expansion and amenities development, most local government efforts focus on growth, expansion, and improvement of the community; the quality of life its residents experience; and the tax base on which government solvency depends. The importance of economic development to local communities is evident in the landmark 2005 *Kelo v. New London* decision by the U.S. Supreme Court; that decision qualified a locality's use of state-granted eminent domain to allow development by a private enterprise as a public interest.^[10]

Economic development policy epitomizes the policy process at the local level. It involves residents who mobilize to support or oppose development initiatives, corporations seeking locations and investment opportunities, and elected officials actively seeking credit and publicity for flashy public works projects in the public eye. Grassroots interest groups are able to form almost instantaneously to affect local policy. For example, such groups form for such diverse purposes as protesting destruction of historic structures, pushing for more favorable decisions from an unpopular family court judge, and opposing the construction of a new interstate highway advocated by local officials.

Although economic development has traditionally been a key beneficiary of local policy attention, local governments have recently become the new arbiters

of public health. Landmark regulatory efforts by cities across the United States have limited smoking in public places, especially restaurants, bars, and government buildings. In January 2007, Bangor, Maine, issued a ban on smoking in automobiles when children are present. New York City recently passed a contentious ban on the sale of foods containing trans-fat. In addition to these actions that regulate personal behavior, other efforts are more broadly conceived. For example, appeals courts in California have upheld local ordinances restricting big box stores such as Wal-Mart from locating in their communities. The city of Arcata, California, has taken perhaps the greatest liberty with local policy making, holding the distinction as the first city to prohibit local officials from complying with the provisions of the U.S. Patriot Act. The spread of policy innovations from one government to another has been shown to be the result of a combination of internal determinants and diffusion.^[11] In studying local policies, a combination of factors have been shown to affect policy development and diffusion, including local government characteristics, establishment of interest groups, presence of regional associations, and focusing events that heighten the salience of the policy.^[12] So, where local conditions are conducive, local officials look to policies adopted elsewhere for potential solutions. Smoking bans have diffused across metropolitan areas in such a fashion.

THE INSTRUMENTS OF LOCAL POLICYMAKING

As local governments are closest to the people, they typically deal with issues of localized concern.^[13] From limiting speed on city streets to implementing stringent building codes to identifying and restricting nuisances such as junky yards, loud music, and foul smells, local governments issue ordinances in response to the concerns people face in their day-to-day lives. The common form of local policy instrument is the ordinance, and, whereas many votes can be taken by voice, ordinance votes in most places require a call of the role of the local legislative body. Ordinance votes in most local governments require readings at three meetings, the third of which is for the official vote on the measure; this effectively prevents decisions from being made in haste and provides significant time for concerned local citizens and businesses to voice their concerns and provide their input regarding the matter in question. When controversial issues arise, local government bodies may also call hearings for the express purpose of gathering public input outside the context of meetings convened for conducting business. Most states provide for local government entities to conduct business according to open meetings and administrative

procedures laws that provide for all meetings to be announced in advance and that restrict the ability of the local decision-making body to conduct business behind closed doors.

Whereas ordinances are the official actions of local legislative bodies, some states provide for policy decisions to be made by the people directly. Referenda and initiatives place policy decisions in the hands of the people; in the case of the initiative, it is directly so, while referenda are initiated through local legislative bodies and then sent to the people for a vote. Initiatives and referenda are frequently used policy tools in some states, although use varies significantly from place to place. In some areas, such as California, for example, local initiatives are widely accepted mechanisms for expressing the public will, leaving policy decisions in the hands of the people rather than their elected officials. Such was the case with Proposition 13, in which California voters capped property tax rates. At the local level, with the passage of Proposition F in November 2006, San Francisco became the first city to require businesses to provide paid sick days to workers. Highlighting the importance of mobilized interests in local policy making, this measure was pushed by an advocacy coalition led by the Coalition for Paid Sick Days.

INFLUENCE OF LOCAL INTERESTS AND INTERGOVERNMENTAL RELATIONSHIPS ON LOCAL POLICY

In 2000, Kentucky passed Senate Bill 247, which liberalized the Commonwealth's alcoholic beverage law by allowing dry counties and cities (of any size) to hold local ballot referenda to approve the sale of alcoholic beverages in restaurants by the drink—a privilege previously reserved to larger territories. This exemplifies the central role of the state in determining the scope and authority of local policy making. The presence of referenda provides formal opportunities for policy participation, leading to group mobilization and more pronounced policy conflict surrounding direct policy decisions by local residents.

Local policy making is not distinct from the political process, but whereas political parties shape public opinion at the national level, at local levels, opinion positions the parties. States or localities with more liberal opinions result in more liberal political parties.^[14] So, at the local level, opinion tends to be fairly homogeneous and parties position themselves in relation to that opinion, often making the differences difficult to discern.

Just as states formulate policy independent of federal policy activity, so local governments formulate policy where states leave discretion to them.^[15] Local

governments are able to act autonomously within the bounds established by their states, and they are able to collaborate regionally to address special problems and policy issues. They are not immune, however, from the pressure of incentives and disincentives from higher-order governments. Fiscal federalism is the relationship of fund transfers among local, state, and federal governments. The availability of grant funding often leads local governments to pursue projects and make policy decisions that do not align perfectly with local preferences, because grant funds reduce the cost to the local government of the sponsored activity. That is, local governments assume the policy preferences of the higher-order grantor governments over the preferences of local residents. Over time, governments can become accustomed to the flow of intergovernmental revenues into their coffers and become dependent on that stream of funding. This has become more difficult over time, as the use of revenue sharing and block grants has been replaced with project grants that are less fungible with general revenues. However, local governments still face the choice of increasing services or amenities by accepting state or federal funds vs. using local funds to provide services that are in greater local demand. Matching requirements for federal funds often mean that localities have to devote a portion of their own-source revenues to fund projects, taking away from projects of local priority, and, to the extent revenues are not available, local governments may resort to debt financing to meet match requirements and sustain the flow of intergovernmental funds into their budgets.^[16]

CONCLUSION

All in all, policy making at the substate level is unique to the particular government in question, shaped by state law, the form of local government, local problems and issues, mobilized interests, the presence of special districts, and the incentives or disincentives placed on those governments through funding opportunities from state and federal programs. Extending grants of home rule authority, increased use of initiatives and referenda, and creation of new special districts likely mean that local policy making in the future will be characterized by greater local discretion and increasingly complex service delivery systems and with greater attention to governance, as interests mobilize to support policy innovation.

ARTICLES OF FURTHER INTEREST

Agenda Setting, p. 71.

Financial Resource Management in Local Government, p. 766.

Home Rule, p. 933.
Metropolitan Governmental Fragmentation,
 p. 1225.
Policy Change, p. 1473.
Policy Diffusion, p. 1482.
Public Policy, p. 1632.

REFERENCES

1. Banovetz, J.M.; Dolan, D.A.; Swain, J.W. *Managing Small Cities and Counties: A Practical Guide*; International City/County Management Association: Washington, DC, 1994.
2. City of Clinton v. The Cedar Rapids and Missouri River Railroad, 24 IOWA LAW REVIEW 455, 1868.
3. Diane Jordan, et al. v. Knox County, Tennessee, et al., E2006-01377-SC-RDM-CV.
4. Miller, P. *Kentucky Politics and Government*; University of Nebraska Press: Lincoln and London, 1994.
5. Mitchell, J. The policymaking activities of public authorities. *Policy Stud. J.* **1990**, 18 (4), 928–942.
6. Lowery, D.; Lyons, W.E. The impact of jurisdictional boundaries: an individual level test of the Tiebout model. *J. Polit.* **1989**, 51 (1), 73–97.
7. Tiebout, C.M. A pure theory of local expenditures. *J. Pol. Econ.* **1956**, 64 (4), 416–424.
8. Teske, P.; Schneider, M.; Mintrom, M.; Best, S. Establishing the micro foundations of a macro theory: information, movers, and the competitive local market for public goods. *Am. Pol. Sci. Rev.* **1993**, 87 (4), 702–713.
9. Logan, J.R.; Molotch, H.L. *Urban Fortunes: The Political Economy of Place*; University of California Press: Berkeley, 1987.
10. Kelo v. New London, 04-108 268 Conn. 1, 843 A. 2d 500, affirmed.
11. Berry, F.S.; Berry, W.D. State lottery adoptions as policy innovations: an event history analysis. *Am. Pol. Sci. Rev.* **1990**, 84 (2), 395–415.
12. Godwin, M.L.; Schroedel, J.R. Policy diffusion and strategies for promoting policy change: evidence from California local gun control ordinances. *Policy Stud. J.* **2000**, 28 (4), 760–777.
13. Peterson, P. *The Price of Federalism*; Brookings Institution Press: Washington, DC, 1995.
14. Erikson, R.; Wright, G.C., Jr.; McIver, J.P. Political parties, public opinion, and state policy in the United States. *Am. Pol. Sci. Rev.* **1989**, 83 (3), 729–750.
15. Dinan, J.D.; Krane, D. The state of American federalism, 2005: federalism resurfaces in the political debate. *Publius J. Fed.* **2005**, 36 (3), 327–374.
16. Hall, J.L. The changing federal grant structure and its potential effects on state/local community development efforts. *J. Public Budg. Acc. Finan. Manage. in press*.

Policy-Oriented Social Science Research: Measuring the Benefits

Henry E. Kilpatrick

*Econpolicy, and School of Public Policy, George Mason University,
Arlington, Fairfax, Virginia, U.S.A.*

INTRODUCTION

The primary focus is on the assessment of research that eventually supports a particular policy program. Advocacy research, or research that is funded by an interested party and designed to bolster a particular position, will not be covered here.

Policy-oriented social science research is a *public good*. A public good has the following attributes: 1) it does not cost anything for another individual to enjoy its benefits, i.e., the marginal cost, or the cost of producing one additional unit of a public good, is zero and 2) it is difficult or impossible to exclude individuals from the enjoyment of a public good. The owners of social science research cannot exclude others from its use once the research is disseminated.

Public goods are an example of market failure. That is, the economic market cannot properly price them, because the benefits cannot be assigned to particular individuals. Therefore, the purchase of public goods is generally a decision made in the political marketplace.

The expected outcome of social science research is typically a public policy program, legislation, regulatory rule making, or a judicial decision. Some level of government, in response to this research finding, implements a policy that results in a new program, the elimination of an old one, a significant change in an existing program, or a significant change in a law or interpretation of law. One presumes that positive net benefits to society would be “returns,” but how does one determine what these benefits are? Clearly, there are benefits to social science research in general, because society continues to fund it, albeit at different levels in different places and times.

Both *ex ante* and an *ex post* evaluations of research are necessary to determine the returns to policy-oriented research. There must be an *ex ante* evaluation of potential benefits in order to determine whether research should be funded initially. Then, there must be *ex post* evaluations to determine if additional research should be funded, and then further *ex post* evaluations to determine whether the research has paid off and whether it should be given additional funds. These further *ex post* evaluations are often based upon evaluating programs that are put into effect as a result of the research.

Because benefits of social science research may accrue quite slowly, methods used for an initial assessment of benefits from completed research may be somewhat different than the methods employed when one is able to more fully evaluate the results. However, if one waits until all costs and benefits are fully measured before proceeding, start-up costs must be re-incurred, and institutional knowledge may be lost.

PROGRAM EVALUATION

Program evaluation is defined by Wholey as a systematic assessment of the performance, the outcomes, and the impacts of public and nonprofit programs and policies. However, linkages between specific activities and subsequent economic or social outcomes are difficult to measure.^[1] Systematic analysis is not undertaken nearly as often as it should be. Evaluations are not always welcomed, carry risks, use scarce resources, and are often a low priority for programs. Program evaluation has a long way to go before it becomes a routine practice.

The evaluation of the research that goes into developing programs is necessarily a step removed from the program evaluation process. It is a rare event that a program is adopted exactly as the original researchers intended. What is more likely is that a program will bear only a resemblance to the concepts that were originally envisioned. A program must pass through the political process before being implemented, while the research may proceed with somewhat less regard as to what is politically feasible. Benefits may be slow to accrue, and the “law of unexpected consequences” may produce commensurate costs to go with these benefits. Hence, measurement must always be uncertain.

The fact that uncertainties abound should not distract from the goal of evaluating social science research. Policy makers demand accountability for the use of scarce funds by those who carry out programs. With this goal in mind, various literatures that may be of use in assessing the benefits of social science research, proposing a limited theoretical framework to use to select an appropriate method of evaluating research will be drawn upon. The political influences,

biases, and difficulties in measuring the benefits of social science research will be explained. Methods that are used in other evaluation programs will be explained and analyzed to determine their efficacy in evaluating the benefits of this research.

Existing Literature

The literature on measuring the benefits of policy-oriented social science research is practically non-existent, but there is a growing body of public administration literature on program evaluation. Studies and methodologies applicable to research and development (R&D), and basic science research evaluation, as well as policy analysis literature are also available. There is no theoretical framework within which to place quantitative tools for measuring the output of scientific inquiry, although there is comprehensive literature that examines the issue of calculating returns to scientific research and R&D (see Refs.^[2,3]). These literatures include empirical work as well as instruction as to the methodologies underpinning the work. Some generalized theory is available in the program evaluation, scientific research, and policy analysis literature. This will form a large portion of the basis for the recommendations for social science research evaluation theory and practice that are considered here.

The Political Setting of Social Science Research

Social science research has the goal of influencing policy decisions made by government. Often, the government it is attempting to influence funds it. Due to the public good aspects of social science research, profit-seeking interests rarely fund it (an exception being advocacy research, as discussed earlier).

Let us trace a social science research project from start to implementation. First, the researcher is likely to take an interest, because existing literature in the researcher's field or public information diffused through the popular media acknowledges that an unsatisfactory situation exists or is projected to develop if not addressed with due diligence. Funding is made available through a government program or private grant, and the researcher competes for it. The researcher completes the project according to a schedule and presents the evidence in the form of a report or study to the agency or foundation. The research recommends a specific project to be implemented. The study is generally published and is announced to the researcher's professional peers. The popular press may pick it up and diffuse it to the public.

The next step is that a legislator or government executive, viewing it as something that is attractive to his or her constituents, takes an interest or a government

agency or outside group recommends legislation to carry out the program. To the extent that the program creates winners and losers or draws significantly upon the resources of the public, there will be a political battle to modify or defeat the program. Legislators will pack the legislation with their own amendments, designed to create advantages for their constituents or to mitigate damages that the proposed program will create. The legislation that finally becomes law may not implement the program in its entirety or may change the program significantly from that which the researcher proposed. Details of implementation are usually left to the agency in charge. More changes to the program may come from within the bureaucracy or as the result of a public rule-making process.

There are a few evaluation methodologies that focus directly upon research without regard to its results on society at large. However, no matter how vexatious it may prove to be, a results-oriented program to measure the benefits of research must at some point contend with the final product—the tangible results of a program.^a

The goal is to measure the benefits of the research in a quantifiable manner. However, the benefits do not often begin to flow with the performance of the research. Usually, a specific focus that takes the form of a policy program, which may be heavily modified prior to implementation, is necessary to produce benefits from the research.

The Uncertain Relationship of Research to Policy

While policy-oriented social science research rarely, if ever, gets translated directly into policy in its entirety, there are other things that evaluators must be concerned about as a prelude to the determination of how to measure its benefits. Assuming that the particular research in question is in fact propitious, evaluators are still faced with a number of problems beyond the control of the researchers. Good research may not become policy because of laws, institutions, prevailing culture, and opposition by interest groups or the public at large to the proposed program. As Quade and Carter accurately stated, "We may be able to design an efficient system on paper, but as of yet we have no algorithms for finding ways to overcome the

^a*Program* is used in a broad sense here. As is true in physical and mathematical sciences, there is social science research that is undertaken as fundamental research, whereby the goal is to acquire basic knowledge. This research is among the riskiest in the portfolio, with uncertain payoffs that tend to accumulate in the long term if at all. The benefits of this research are extraordinarily difficult to calculate, and more failure is expected at the riskier end of a portfolio, but the end result of the research should be programs.

resistance offered by tradition, legal restrictions, and a host of privileged interests that inhibit radical or even morphological change.”^[4] Competent research may be translated into policy, but the underlying conditions that the research is based upon may change. Finally, good research may be translated into policy, but unanticipated consequences may negate the positive aspects of implementation. Should research that is not used for practical purposes be evaluated for its benefits to society? Unless observers can readily identify its benefits to society or it can be evaluated in a wholesale fashion with similar research, it probably should not be. The ultimate social test of information commissioned by investors in research is its third-party utility.

Biases in Social Science Research

While research inquiries in the arena of the social sciences that are relatively value free exist, they are of far less research interest than those that are value laden. Research assumptions are often highly subject to whatever school of thought the researcher subscribes to, which, in turn, is dependent on a particular political belief or value system. Thus, social science research is often used as ammunition in political wars. Furthermore, research is subject to bias and conflict of interest, at times discovering what the financier of the research wishes to find rather than determining the correct solution to the research questions (if one correct solution indeed exists). Another variation of this conflict of interest is that in-house analyses may favor solutions that confer more resources upon the analysts. Pursuit of fame is another source of research bias. Publication bias often affects research by encouraging scientists to emphasize efforts on new approaches rather than on perfection of old methods. Data mining, or the act of fitting as many alternative equations as there are alternative subsets of potentially explanatory variables, and then picking the best data according to the researcher's criteria invalidates conventional statistical tests. However, those who evaluate the research do not necessarily know if data mining has taken place. A more serious form of data mining is the reporting only of results that the researcher wishes to report, while earlier data runs that may refute the hypothesis being tested are left on the cutting-room floor. There is simple careless, haphazard, and hurried research to contend with. There is fraudulent research and the manipulation of numbers, which are rare exceptions. A project will also require judgments to be made about the direction the research will take at critical paths when options are available.

Why does bias matter in determining how social science research should be evaluated? It has been noted that our evaluation tools are inexact, benefits may

accrue slowly, and social science is value laden. In early stages of evaluation, an individual makes inferences as to whether the research will be productive without programs in place to produce tangible results. The subjective nature of much of social science research carries over to our measurement tools and requires that we be alert to bias in the research.

Designing Social Science Research Evaluation

A design is necessary to measure the benefits and otherwise evaluate social science research. In carrying out social science research evaluation, one must first determine the purpose of the analysis. Will it be conducted to determine if the research is profitable for society, or at least for the financier? Is it to decide whether further research in the field should be funded? Next, a clear, precise, and manageable process designed to elicit information about the research to be analyzed needs to be created. Following this, the analysis must clearly define exactly what is to be measured. Research is but one step in a process, and the actual research must be isolated from the program that implements it. The subject research must also be separated from related research, as well as the rest of the surrounding policies and processes to analyze it effectively. Whether the evaluation analysis is to cover one particular item of research or analyze a group of projects in tandem must be determined. If the target research is a group of projects, it must be decided whether one evaluation methodology or one mix of methodologies will serve to measure the entire range of research. After isolating the research, a metric must be designed in order to measure or rank it against a scale. The users of the research must be clearly defined. While this may seem obvious, there may be multiple users and spillover effects. New knowledge may enhance the value of old, ignored, used up, or simply unrecognized research by lending new life to previous research results. This effect, if it is present, should be captured in any measure of benefits.

How rigorous must the analysis be? Rigor in program evaluation can be costly to achieve. The same is true in measuring the benefits of social science research. Therefore, the rigor of the specific analysis should be defined prior to undertaking the analysis.

An empirical method of analysis that will accomplish the above program must be chosen. The analysis must be carried out according to procedure with as little bias as possible, and results of the analysis must be reported clearly and precisely. This report must also specify the limitations of the evaluation methodology.

With this framework in mind, let us proceed to the “nuts and bolts” of the evaluation process. The available methodologies generally fall short of the

theoretical mark in actual operation. The ability of an evaluator to isolate a particular body of research is questionable. Rigor is a matter of judgment. In fact, the evaluator's judgment, and sometimes the judgments of others, is a crucial element of every available methodology.

The First Step in the Process: Replication of Original Studies

While replication is used extensively in science and technology research, its use in social science is paltry. Replication is much more difficult in the realm of social science, where controlled experiments are generally not available as they are in the purer sciences. Reality within social science is a social construct, it is unstable, and it is subject to continuous change. It is more dependent on culture than is pure scientific research. The fact that social sciences are value laden makes the task of evaluation harder.

One might fairly ask why replication should be used to determine the benefits of social science research, because replication is further research rather than research evaluation. But, it assures that research that at least initially appears to be beneficial is correctly reporting results prior to program development and prior to further evaluation of the research.

Replication is often left to graduate students in the social science arena, if it is conducted at all. The system of professional honors and rewards in the social science system does not place a high value on replication. The ability to reproduce a researcher's methods in order to validate results is a clearly recognized requirement of scientific inquiry and is *de rigueur* in the sciences. It should become a part of social science research evaluation as a first step in establishing the benefits of particular research.

The simplest replication is to repeat the experiment exactly using the same data set. This is often possible if the research performs statistical analyses on a particular data set consisting of historical events. One may also consider multiple cases as one would consider multiple experiments. In this case, literal replication is defined as the selection of a case, so that it predicts similar results. Theoretical replication produces contrary results but for predictable reasons. The purposes of conducting replication studies are several. First, they provide useful checks on the work of the original researchers, particularly to provide evidence as to whether the procedures were carried out as described in the replicated paper. Second, if a number of different procedures are used in addition to the ones originally used, evidence of whether different methods produce different results is provided. If different methods do not produce different results, the study is replicated

successfully. If they produce different results, it is necessary to discover why. And, evaluating the initial research is a futile exercise if there are major problems with the original inquiry, so it is useful to examine the work for accuracy before proceeding with further analysis.

Most Evaluation Methods Can Be Subjective and Difficult to Quantify

Having identified replication as the first necessary step in the evaluation process, a number of methods of evaluation are discussed below. These methods are not equal; some are better than others for particular tasks. Some will be rejected for most uses in evaluating social science research, but will be discussed because they are available. Most of the methods available are judgmental, subjective, and offer little in the way of quantifiable results that will satisfy those who wish to have a simple metric by which to accept or reject. This is unavoidable with the tools we have just as it is unavoidable in the original research we seek to evaluate.

Bibliometrics

Bibliometrics, or the measurement of published materials such as citations and publication counts, is a tool that is often used as a proxy to measure returns to R&D spending and evaluation of scientific programs. It has been used since the turn of the twentieth century. Garfield and Price pioneered the modern use of bibliometrics.^[5,6] This method is used extensively in the British university system, by the U.S. National Institutes of Health, and by the National Science Foundation. It is a measurement of research others believe to be important, because citations are an indication of a researcher's familiarity with a particular literature and the relevance of a particular article or study to a researcher's own work. There are a number of methods used to perform an analysis of citations. These range from simple counts to linking citations to each other.

Bibliometrics suffers from a number of flaws. The esteem in which a particular author holds a cited paper cannot be garnered from references of citations. Bibliometrics cannot measure quality. The discredited cold fusion paper of Pons and Fleischmann was among the top 10 of the most heavily cited papers in the scientific literature for several years. The quality of the article in which the citation appears cannot easily be measured. Some disciplines are more publication oriented than others. Finally, this analysis cannot measure benefits to society at large from the research undertaken.

One of the suitable characteristics of bibliometrics, from the standpoint of evaluation methodology, is that

it imposes measurement upon the original research rather than the programs that may result from it. However, given its flaws, it may be useful as a supplement to other methods but cannot stand alone as a method for evaluating social science research.

Similar Peer Review Measures

Measures similar to bibliometrics include convening panels of experts for evaluation based upon peer review. First used in 1665, it is the earliest evaluation method known. Peer review relies upon the expert judgment of colleagues within the same field. Many R&D laboratories as well as government agencies use it. Social science research contracts are likely to be awarded as a result of peer review in the competition process. There are a number of different forms of peer review. The reader is referred to Bozeman and Melkers for a review of the different methods as well as guidelines for enhancing this evaluation.^[7]

The results of these measures are highly dependent on the choice of reviewers. Furthermore, this methodology does not satisfy the yearning for a less judgment oriented, more direct, and more empirically based means of measurement. Peer-based appraisals such as enumerating awards and other recognition from professional societies and patent and copyright counts (which have limited use for social science research) provide a quantitative means of measuring output, but not a scalable one. Peer review methods may be used along with other methods but cannot stand alone as satisfactory means of measuring social science output.

Case Study Methodology

“A case study is an empirical inquiry that: investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used . . .”^[8] Case study methodology makes use of historical comparative data to tell a story that is much richer in detail than statistical analysis, which by its nature must be somewhat abstract.

Case study methodology is a viable tool for the evaluation of social science research when carefully applied. The evaluator can tell a story of how the research is translated into policy and how the policy then carries out the goals of the research. Modeling and advanced statistical methods are rarely used, but statistical data in the form of charts and tables are expected. The methodology may be qualitative, quantitative, or both.

The primary weaknesses of case studies are that there is potentially a lack of rigor, possibly little basis

for scientific generalization, and case studies sometimes take too long and result in massive, unreadable documents. To overcome these weaknesses, the evaluator must be specific about the research question and develop a rationale for the methods used and how these methods address the specific evaluation needs. Examining successful and unsuccessful projects may be useful, but this may use up resources that could be put to better use in evaluating research that at least gives the first impression of being successful.

Delphi Method

The Delphi method, first used in an attempt to improve betting scores at horse races in 1948, was developed at the Rand Corporation in the 1950s and 1960s as a method to increase the accuracy of forecasts. It involves picking an expert panel, which is surveyed through a questionnaire. The questionnaire is evaluated and then returned to the experts with the findings presented statistically. They are allowed to change their opinions to reflect new information and the judgments of their colleagues (but the experts are not allowed to have contact with each other). This process goes through a number of iterations until opinions converge. Convergence can be dependent on peer pressure toward conformity rather than rational processing of information. The Delphi method presently offers little for social science research analysis that cannot be gained through the use of other methodologies.

More Direct Methods: User Evaluations

User evaluations in the context of social science research evaluation entail surveying the actual users of the research to elicit information on its value. This is a costly form of analysis that may involve conducting interviews or mail or telephone surveys and the development of a survey questionnaire. The metric is ordinal rather than cardinal. There may be an inherent conflict of interest in surveying users, particularly if they receive free or low-cost benefits. They have a vested interest in expressing satisfaction in order to keep the benefits flowing.

A survey is highly dependent on the biases of those surveyed as well as those biases inherent in the questions posed to the sample of respondents. Given the costs, survey methods do not appear to be among the best means of conducting social science research evaluation other than to determine the satisfaction levels of the particular beneficiaries, particularly if the benefits to society are of greater interest.

Focus groups, small group discussions conducted with the users of the research, are another means of user evaluation. This is somewhat less costly than

survey methods, but selection of participants is not often random, and the output may be difficult to present in a systematic fashion. This technique should be reserved for program evaluation and customer service, when the provider wishes to improve delivery of its product or service.

Cost-Benefit Analysis or Measurement of Social Rates of Return

Cost-benefit analysis, which reached its peak of popularity in environmental studies, seeks to measure the benefits and the costs of a given project in a systematic and rational manner. The logic underpinning of cost-benefit analysis is quite simple. One calculates the benefits of a project or program in monetary terms. Then one calculates the total costs and compares the two. Alternatively, one may calculate and compare marginal costs and benefits. Financial methods such as the internal rate of return (the discount rate that makes the present value of estimated benefits equal the present value of estimated costs) are used in conjunction with this methodology.

While this is rudimentary in principle, it poses great challenges in practice. Quantifying costs and benefits may prove elusive. Particularly thorny is the question of determining the value of human life and health. The analyst makes judgments and critical assumptions that are buried in the calculations. Furthermore, the process is prone to manipulation by those determined to reach a particular outcome. The discount rate is never known with certainty and may be subject to manipulation. Finally, externalities are difficult to estimate with simple cost-benefit approaches.

While formal cost-benefit analysis has lost much of the glamour it had two decades ago, it is still a useful tool in policy-oriented research evaluation, as long as the premises upon which the evaluator makes judgments and calculations are clearly spelled out, even though it might prove to be difficult to separate the results of the research from the results of the implementation program. Although the analysis may have lost some of its luster over the years, all methods of evaluation require the analyst to make a determination of benefits vs. costs.

Statistical Methods: Regression Analysis

Statistical regression analysis takes the form of the standard regression equation:

$$Y = \alpha + \beta X + \mu$$

where Y is the dependent variable, α is a constant, X represents the independent variables with coefficient

β , and μ is the error term. The method fits a line or curve by the statistical means of least squares. It is a well-known, but intricate, statistical technique.

Popper refers to analyses conducted upon R&D using regression methods as production function analyses, because they involve the use of production functions in a regression equation.^[9] A production function measures the change in output due to a change in an index of inputs. The empirical production function is estimated by the specification of various models.

This method has been used with reasonable success in R&D research.^b Assessing the benefits of specific research by use of production functions may prove to be highly elusive. There is an advantage in using it to assess the benefits of R&D in private markets that do not exist in public markets—the existence of measurable input to firms and measurable output in the form of goods and services produced. However, it is possible to roughly model the results of social science research in the following manner: Use a time series of program budgets minus personnel costs as proxies for capital costs, personnel costs as proxies for labor, and the variable that the research is designed to act upon as the output variable. In particular, one may be quite interested in determining whether the research in question leads to increasing returns to scale.

Regression analysis does not need to be production function oriented in order to be useful in evaluation practices. One may partially isolate the original research from the program by properly selecting variables if multicollinearity does not present an insurmountable obstacle.

The pitfalls of regression are that it must relinquish rich details in order to create a workable model, the model may be mis-specified, and it may create a false sense of quantitative certainty. The evaluator must be clear on the variables selected and the reasons they were selected.

Robust and Substantively Weighted Regression Analysis

Oftentimes data sets contain “outliers” or individual observations that are significantly larger or smaller than most in the sample. There are a number of methods used to constrict their effects on the analysis without discarding them altogether without valid reason to do so. These methods may be found in most modern statistical texts. In addition, there is a newer method that is just the opposite: substantively weighted regression. This method gives heavier weight to the outliers as it attempts to define “best practices” or best

^bMansfield and Hall discuss the uses and pitfalls of this method in greater detail.^[10,11]

practitioners of a particular program. Primarily a tool of public administration, it also has applicability in social science research evaluation.^[12]

Operations Research Modeling Methods

Operations research modeling techniques have also been used in assessing programs for a number of years. Linear programming is a quantitative methodology that makes use of the mathematical modeling of constrained optimization. An optimal solution is sought, given certain constraints imposed by the model. A number of methods are available beyond simple linear programming. These include goal programming and data envelopment analysis (DEA).

Goal programming is based upon a paper by Charnes et al., it represents an attempt to put the concept of “*satisficing*” into operations research modeling.^[13] The technique puts goals and aspirations on one side of the equation and constraints that must be satisfied on the other. Priorities of individuals and groups may be taken into account by means of goal programming.

While this method does not appear to receive much discussion in the evaluation literature, perhaps because it is associated with applied science and engineering, it is likely to hold some promise as a quantitative methodology for the purposes specified herein. Selecting the researcher’s goals, using a goal-programming methodology rather than selecting the goals of the program, may be a means of isolating the effects of the research from the program.

DEA contains the concept of the production function that was mentioned earlier. Rather than fitting a line in a least-squares fashion, it fits a boundary that contains all points within it (the production possibilities curve). This method is likely to be successful in those cases where production function methodology is appropriate. The use of both methods will increase the rigor of the study, but of course, this has a cost.

Simulation: A More Recent Development in Operations Research

Simulation is a heuristic, probability-based methodology that conducts experiments by proxy using a model and runs that consist of selecting values of uncontrolled variables. This is accomplished by using Monte Carlo analysis, or repeated simulation of a stochastic model to investigate the properties of statistical techniques applied to it. An abstract model, but one that contains enough stylized facts to approach authenticity, is necessary. This technique has a long history in applied sciences, particularly in

defense work. Some advantages are that simulation can deal with uncertainty, it is versatile because it can be run many times, it can reproduce variability that actually occurs in the system being simulated, multiple objectives can be analyzed, and it is a rational tool. The limitations are that there may be a temptation to build a model that is too big and too expensive, and one could proceed with the modeling process only to find out that no one understands the interactions of the variables well enough to build a workable model. Judgment and elements of subjectivity enter into specifying the model as they do in other methodologies. Finally, simulation is based upon limited real-world data rather than the analysis of historical processes using only knowledge of actual events.

Simulation has been most heavily used in applied sciences. It is useful in social science research evaluation when data are hard to come by, or one suspects processes are at work that are extraordinarily difficult to model by conventional means. Additional simulation models involving the use of genetic algorithms, non-linear regression, and other modeling techniques are currently being developed in the growing field of computational economics and by complexity theorists.

Simulation is quite useful to determine system sensitivity and the range of intervention options. The major problem is that it is not based upon actual observed data, although the models may take actual data into consideration when generating scenarios. The output measures potential reality rather than actual circumstances.

Meta-Analysis

Meta-analysis is a fairly new analytical method that summarizes the studies that have the same dependent variable that have already been completed and are found in the literature. The term was coined by George Glass in 1976, who described it this way: “Meta-analysis refers to the analysis of analyses ... the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings. It connotes a rigorous alternative to the casual, narrative discussions of research studies, which typify our attempts to make sense of the rapidly expanding research literature.”^[14]

CONCLUSION

This article has outlined a rudimentary theory to be used in the application of evaluation methodologies. A number of methods to accomplish a benefits analysis of social science research have been offered. Many are

well-known research methodologies that are already extensively used in conducting social science research and program evaluation. Replication should be conducted as part of the evaluation process prior to reaching the question of whether the research is beneficial. None of the methods presented as potential tools for carrying out evaluations is completely satisfactory insofar as answering our theoretical needs. A mix of methods may prove to be the best approach in many cases. The link between research and application of results is likely to remain tenuous, which means that an investigation into the benefits of social science research will likely remain an exercise in muddling through. Decision makers will find it necessary to select among the evaluation methodologies.

Efforts into improving research benefits analysis should focus upon improving existing methods, especially because the costs of processing large quantities of data have dropped significantly and will continue to drop. In particular, operations research methodologies appear ripe for adaptation to social science research evaluation. A single quantitative solution that stands head and shoulders above the rest is unlikely to be found. Better results are likely to be obtained if programs to evaluate research are implemented rather than if we consume time and resources looking for better means of evaluation. This does not mean that we should not seek to improve our evaluation methods. However, we should not strive to perfect our methods while passing up opportunities to put the tools we have into practice.

REFERENCES

1. Wholey, J.S.; Hatrey, H.P.; Newcomer, K.E.; Eds. *Handbook of Practical Program Evaluation*; Jossey-Bass Publishers: San Francisco, 1994.
2. Alston, J.W.; Norton, G.W.; Pardey, P.G. *Science Under Scarcity: Principles and Practice for Agricultural Research Evaluation and Priority Setting*; Cornell University Press: Ithaca, NY, 1995.
3. Stephan, P.E. The economics of science. *J. Econ. Lit.* **1996**, *34*, 1199–1235.
4. Quade, E.S.; Carter, G.M. *Analysis for Public Decisions*; North-Holland: New York, 1989.
5. Garfield, E. Citation indexes in sociological and historical research. *Am. Doc.* **1963**, *14* (4), 89–102.
6. Price, D.J.S. *Science Since Babylon*; Yale University Press: New Haven, 1961.

7. Bozeman, B.; Melkers, J. *Evaluating R&D Impacts: Methods and Practice*; Kluwer Academic Publishers: Boston, 1993; 79–98.
8. Yin, R.K. *Case Study Research: Design and Methods*; Sage Publications: Newbury Park, CA, 1989.
9. Popper, S.W. *Economic Approaches to Measuring the Performance and Benefits of Fundamental Science*, Rand Domestic Research Division Report. The Rand Corporation Santa Monica: CA, 1995.
10. Mansfield, E.R. Contribution of new technology to the economy. In *Technology, R&D, and the Economy*; Smith, B.L.R., Barfield, C.E., Eds.; The Brookings Institution and the American Enterprise Institute: Washington, DC, 1996.
11. Hall, B.H. The private and social returns to research and development: What have we learned? In *Technology, R&D, and the Economy*; Smith, B.L.R., Barfield, C.E., Eds.; The Brookings Institution and the American Enterprise Institute: Washington, DC, 1996.
12. Gill, J.; Meier, K.J. Public administration research and practice: A methodological manifesto; <http://psblade.ucdavis.edu/papers/manifesto4.pdf> (accessed 1999).
13. Charnes, W.; Cooper, W.; Ferguson, R. Optimal estimation of executive compensation by linear programming. *Manage. Sci.* **1955**, *1*, 138–151.
14. Glass, G. Primary, secondary and meta-analysis of research. *Educ. Res.* **1976**, *5*, 3–8.

BIBLIOGRAPHY

1. Begg, C.B.; Berlin, J.A. Publication bias: a problem in interpreting medical data. *J. R. Stat. Soc., Ser. A Stat. Soc.* **1988**, *15* (3), 419–463.
2. Card, A.; Krueger, A.B. Time series minimum wage studies: A meta-analysis. *AER Proc.* **1985**, *8*, 238–243.
3. Cragg, J.G. Monte Carlo methods. In *The New Palgrave Time Series and Statistics*; Eatwell, J., Milgate, M., Newman, P., Eds.; W.W. Norton & Company: New York, 1990; 171–174.
4. Denton, F. Data mining as an industry. *Rev. Econ. Stat.* **1985**, *57*, 124–127.
5. Feigenbaum, S.; Levy, D.M. Testing the replication hypothesis when the data set is subject to gross error. *Econ. Lett.* **1990**, *34*, 49–53.
6. Feigenbaum, S.; Levy, D.M. The market for (ir)reproducible economics. *Soc. Epistemol.* **1993**, *7*, 215–232.
7. Romero, C. *Handbook of Critical Issues in Goal Programming*; Pergamon Press: Oxford, 1991.
8. Stiglitz, J.E. *Economics of the Public Sector*; W.W. Norton & Company: New York, 1988.
9. Woudenberg, F. An evaluation of delphi. *Technol. Forecast. Soc. Change* **1991**, *40*, 131–150.

Popular Reporting

Joy A. Clay

*School of Urban Affairs and Public Policy, University of Memphis,
Memphis, Tennessee, U.S.A.*

INTRODUCTION

Professional bodies concerned with assuring and promoting high quality government accounting and financial reporting practices, such as the Governmental Accounting Standards Board (GASB) and Government Finance Officers Association (GFOA), view quality financial reporting as an important facet of assuring government accountability (see www.seagov.org and www.gfoa.org, respectively).^[1] Accordingly, state and local governments are encouraged to issue a high quality, comprehensive annual financial report (CAFR) in conformance with generally accepted accounting principles (GAAP). This annual financial report, the CAFR, is increasingly being supplemented with a more stylish and simpler report, the Popular Annual Financial Report (PAFR), aimed at general, “popular” audiences.

Although the simpler report has less technical language and provides less detailed information than the CAFR, the popular report is expected to be credible and present an accurate, objective picture of governmental activity. Popular reports are generally aimed at lay audiences such as citizens, activists, business leaders, civic organizations, etc. who are more interested in gaining a broad understanding of a government’s activities and who often are not versed in, or want to take the time to understand, governmental accounting and bureaucratic language.

CONTEXT FOR THE REPORT: GOVERNMENT ACCOUNTABILITY

Underlying democratic government is an expectation that those in power are held accountable for their decisions and actions. Public reaction to the recent financial and governance scandals in the private sector, such as Enron, and continued public concern about government spending have added increased energy to the public management reform efforts started in the 1980s. A growing consensus on the part of public administration professional groups, including the International City/County Management Association (ICMA), GFOA, GASB, American Society for Public Administration (ASPA), has led these groups to place pressure on governments to expand vehicles and

processes of accountability. Consequently, public agencies/governments are encouraged not just to improve formal evaluation or auditing processes but to assure the transparency of governmental actions, that is, not just to evaluate programmatic results but to regularly report on them to citizens in a way that assures reliable, useful, understandable, and accurate information is being provided to a broad audience. Thus, governments are expected to generate consistent performance data and report it regularly and publicly.^[2] The popular report is one of the vehicles for governments to report on performance/results. The underlying assumption is that regular governmental attention to public accountability will in the short and long term further public trust in and credibility of government as it reduces information asymmetry.^[2]

CHARACTERISTICS OF THE POPULAR REPORT

The most specific guidance available on popular report characteristics is offered by GFOA. In summary, GFOA recognizes that popular reports can be presented in a variety of formats. Specifically, GFOA strongly encourages governments to produce a summary financial report that successfully attends to the following five characteristics:

1. Presents timely information (no later than 6 mo after the close of the fiscal year);
2. Clearly describes scope (defines what is included in the report), providing an explicit link to the CAFR to facilitate further research if desired by the reader;
3. Thoughtfully presents helpful comparative data, for example, trends or benchmarks;
4. Presents a report design/layout that is attractive, engaging, clear, and concise, with a format that avoids technical jargon and includes supportive and summative graphic material such as charts and photographs or narrative material that enhances reader understanding; and,
5. Implements an appropriate dissemination strategy (that is, the distribution approach actually gets the report to the target audience/stakeholders).

An important challenge for governments is this last characteristic, the appropriate distribution of the report. The distribution strategy must balance cost, including production and distribution costs, as well as predict volume, based on estimated readership interest. Increasingly governments make these reports available on their websites (see www.cityofmemphis.org/citizensReport2003 or PAFR award winners' sites at www.gfoa.org). Electronic posting of the popular report not only saves distribution costs but broadens accessibility.

RECOGNITION FOR EXCELLENCE

The Outstanding Achievement in Popular Annual Financial Reporting Award was established by GFOA in 1991 (J. Phillips, personal communication, September 14, 2004). This award recognizes those governments that develop a high quality popular report. Judging by the history of success that GFOA has had with the CAFR award, over 3000 governments annually apply for the CAFR award,^[3] this is a successful means of encouraging innovation and adoption of a high quality financial reporting practice. As with CAFR award winners, governments, specifically the finance and budget offices, winning the PAFR award tend to advertise receiving the award as a prestigious measure of their performance.

GFOA asks judges to assess PAFRs on their overall quality and usefulness, specifically evaluating the report's reader appeal, understandability, distribution methods, and degree of creativity/innovation. A slowly increasing number of governments are applying for the award. For fiscal year end 1996, only 29 governments applied for the award but 55 applied for 1997. More recently, the number of applicants for the award has increased approximately 6% annually (J. Phillips, personal communication, September 14, 2004). GFOA reports that 112 governments applied for the award last year and anticipate 116 next year (www.gfoa.org). The growing list of state and local governments seeking this relatively recent award attests to the number of governments producing this new financial report.

IMPLEMENTATION CONSIDERATIONS

Leadership Responsibility

Before deciding to undertake creation of a popular report, government executives need to make several key decisions, including assignment of responsibility for preparing the report and coordinating data collection, scope of the report content, dissemination guidelines, and cost parameters. Executive leaders will also

need to take into consideration the political ramifications associated with such a report, including degree of involvement of legislative leaders, capacity to continue with the report once started, willingness to receive public feedback, and the ability to present balanced and objective information. Inherently, the popular report is to be a public and accessible document. Thus, both successes and challenges will need to be presented. GFOA states very clearly that "most important, the popular report should establish its credibility with its intended readers by presenting information in a balanced and objective manner."^[3]

Report Tasks

Once the broad framework has been decided upon by government leaders, the work can be delegated to the appropriate unit and staff member(s) and responsibility for collecting the data and preparing the report can proceed. A critical decision rests on the assessment of staff capacity, that is, budget analysts may not have the requisite graphic or stylistic creativity needed to produce a highly styled report. Consequently, leaders may need to seek outside assistance, especially to create the cover design and layout, decisions which may add to the overhead cost of the report as the design becomes more sophisticated. If the governmental entity wishes to seek the GFOA excellence award, the report also needs to reflect innovative and creative practice. Consequently, leaders will need to commit time and attention to the project. This may present a challenge since these same leaders generally have the responsibility for producing the CAFR. All of these managerial decisions rest within budget constraints and feasibility considerations.

As with any report, the lead narrative and graphic information should capture the essence of the governmental unit. With the PAFR, this approach must be pleasing to the eye. Key content decisions include finalizing the scope of the report (what will be included, breadth and depth), how best to link to the CAFR, appropriate inclusion of insightful trend and/or benchmark information^[4-6] as well as graphic/photo illustrations add to the attractiveness and understandability features of the report but also to the cost of the report. The use of color enhances the presentation quality of the report but adds to production costs.

A significant feature of the PAFR also can be the inclusion of results of citizen satisfaction information. Integrating an assessment of citizen satisfaction with programmatic performance reporting will more comprehensively address various issues related to public accountability.^[7]

The PAFR, especially if available on the government website, offers the opportunity for place marketing.^[8] This makes the section that introduces the

community context and strategic visioning of the community of particular value. The popular financial report, thus, can serve as a promotional vehicle that communicates the government's assets and the essence of the community to readers. Achieving a balance between substance and presentation is a measure of a government's ability to present policy, performance, and programmatic information in a manner that encourages readers to actually read the report and thus, better understand their government.

CONCLUSIONS

As with all government reports, public administrators need to attend to data quality issues including any problems with reliability and validity of data, and objectivity. Although in "popular" format, the popular report is an official government financial report.

The decision to broaden the availability of government financial reporting recognizes that citizens are stakeholders in their community, not just customers of services, and should be able to expect a return on their tax dollars.^[9] The PAFR presents an opportunity for governments to proactively provide information that addresses policy, program, and performance accountability. The value of the PAFR lies in the nature of the report, that is, governments must present this information in human terms, not shrouded in technical and complex financial or programmatic language. To engage in the PAFR process, encourages creativity on the part of budget and finance officials and reinforces the need to be able to articulate a community

context, the vision of the community, achievements to date, and artfully position challenges.

REFERENCES

1. Government Accounting Standards Board. In *Reporting Performance Information: Suggested Criteria for Effective Communication*, Special Report; Norwalk, CT, August 2003.
2. Coe, C.K. A report card on report cards. *Public Perform. Mgmt. Rev.* **2003**, 27 (2), 53–76.
3. Government Finance Officers Association. Criteria for Popular Annual Financial Reporting Awards Program, GFOA Technical Services Center, <http://www.gfoa.org> (accessed October 2004).
4. Ammons, D.N. *Municipal Benchmarks: Assessing Local Performance and Establishing Community Standards*, 2nd Ed.; Sage: Thousand Oaks, CA, 2001.
5. Morley, E.; Bryant, S.P.; Hatry, H.P. *Comparative Performance Measurement*; Urban Institute: Washington, DC, 2001.
6. Hatry, H. *Performance Measurement: Getting Results*; Urban Institute: Washington, DC, 1999.
7. Van Ryzin, G.G.; Muzzio, D.; Immerwahr, S.; Gulick, L.; Martinez, E. Drivers and consequences of citizen satisfaction: an application of the American customer satisfaction index model to New York City. *Public Admin. Rev.* **2004**, 64 (3), 331–341.
8. Kotler, P.; Haider, D.H.; Rein, I.J. *Marketing Places: Attracting Investment, Industry, and Tourism to Cities, States, and Nations*; The Free Press: New York, 1993.
9. Smith, G.E.; Huntsman, C.A. Reframing the metaphor of the citizen–government relationship: a value centered perspective. *Public Admin. Rev.* **1997**, 57 (4), 309–318.

Portfolio Theory

Richard L.B. LeCompte

*Department of Finance, Real Estate and Decision Sciences, Wichita State University,
Wichita, Kansas, U.S.A.*

INTRODUCTION

Portfolio Theory, also known as Modern Portfolio Theory, focuses on the construction and selection of optimal portfolios of risky assets by risk-averse investors. The origin of modern portfolio theory is the seminal work of Harry M. Markowitz in the early 1950s. This work was published in his 1952 paper, *Portfolio Selection*.^[1] His body of work on portfolio theory resulted in him being awarded the Nobel Prize in economics.

Today, the work of Markowitz, and the extensions of it, provides the basis for portfolio theory as well as an understanding of the relationship between the risk and expected return on securities or assets held in a portfolio of assets. Utilizing his methodology, the investor seeks the maximum expected return for a given level of risk or the minimum risk for a given level of expected return. He altered the focus of investment analysis away from individual security selection toward investment analysis concentrated on diversification concepts and the impact of individual securities on a portfolio's risk–return characteristics. In Markowitz's formulation, efficient portfolios are formed by choosing an asset based upon its interaction with other assets in the portfolio, not on the basis of its stand-alone performance. The evidence of his impact is the widespread application of his research in quantitative portfolio management, risk management, and asset valuation. The acceptance of portfolio theory by academics, financial practitioners, and investors over the past 50 years has provided much of the impetus for the growth in the use of passive portfolio management techniques and the tremendous popularity of index funds.

Portfolio theory also forms the basis for a wide variety of academic research in diversification theory, portfolio selection, market efficiency, and capital market theory. Most notably, Markowitz's optimal portfolio selection methodology is utilized in the development of Capital Market Theory, which led to the conception and derivation of the Capital Asset Pricing Model (CAPM). CAPM was developed independently during the mid-1960s by William F. Sharpe,^[2] John Lintner,^[3] and Jan Mossin.^[4] CAPM research, in turn, formed the building blocks for another important asset pricing model, the Arbitrage Pricing Theory (APT). The APT was derived by Stephen A. Ross^[5] in 1976.

OVERVIEW

Portfolio theory is defined as quantitative analysis for optimal risk management.^[6] Markowitz's portfolio theory is based upon a mean-variance optimization (MVO) process that searches for efficient portfolios. An efficient portfolio is one that provides minimum risk for a given level of return or maximum return for a given level of risk. Rational, mean-variance optimizing, investors will construct optimal asset portfolios by determining the best risk–return opportunities available from feasible investment portfolios and then choose the best portfolio from that feasible set.^[7] A key contribution of Markowitz's research is that it mathematically quantifies, for the first time, the benefits of risk reduction achieved through portfolio diversification. Moreover, his work demonstrates that optimal diversification requires the selection of securities based upon their interrelationships or comovements, not just a random selection of a large number of securities. The MVO algorithm generates “efficient” portfolios by minimizing risk subject to a given level of expected return or maximizing expected return subject to a given level or risk. The optimization process assumes that risk-averse investors consider risk to be the variability of expected asset returns, measured by variance. These rational risk-averse investors select investments using only mean-variance criteria, i.e., given two assets with equal expected returns, they will choose the one with the lower risk, or given two assets with equal risk, they will choose the one with the higher expected return. The required inputs into the optimization model are the expected returns for each asset, the variance of each asset, and the covariances between these assets, where covariance is the statistical measure of the comovement of two securities' returns through time. Positive covariance between securities indicates that their returns move in the same direction; negative covariance indicates that the securities' returns move in opposite directions. The size of the covariance measure indicates the degree of comovement between securities.

Markowitz shows that the relevant risk of a security is determined by the covariance of its return with that of a diversified portfolio or market index, not a security's own variance in isolation. Portfolio theory provides investors with appropriate measures of portfolio and security risks and a methodology for

accomplishing optimal security selection and portfolio construction. At the same time, it quantifies the importance of portfolio diversification.

MEAN-VARIANCE FRONTIER

The mean-variance optimization process generates an efficient frontier, as shown in Fig. 1, consisting of portfolios that “dominate” the ones lying below and to the right on the basis of expected return or standard deviation.

These dominant, or efficient, portfolios are ones for which no other asset or portfolio of assets offers higher expected return with equal or less risk, or offers lower risk with equal or greater expected return. The optimal portfolio for an individual investor is the efficient portfolio that provides the highest utility based upon risk–return preferences. The utility curves, shown in Fig. 2, indicate the investor’s willingness to trade expected return for risk, while the efficient frontier shows the investor’s opportunity set, or ability to trade risk for expected return. The point of tangency between these two curves indicates the optimal portfolio choice for the investor in the MVO framework. This point of tangency identifies the efficient portfolio that provides the rational risk-averse investor with the highest utility or level of satisfaction.

EXPECTED RETURN AND RISK

To execute MVO, one first needs to compute the expected return for each asset and portfolio. The expected return on a portfolio is the weighted average of the expected individual returns of the assets held within it. The weight assigned to each asset is the

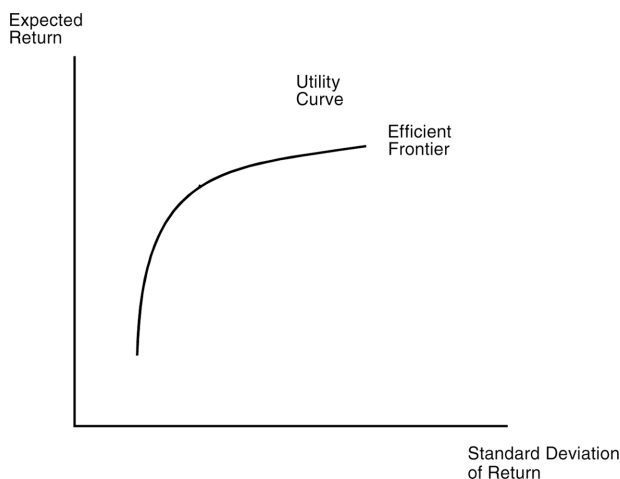


Fig. 1 Efficient set of portfolios.

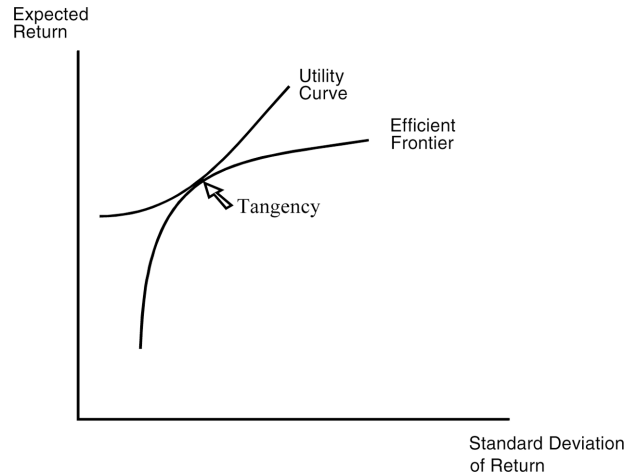


Fig. 2 Optimal portfolio selection without risk-free asset.

asset’s value as a percentage of the total portfolio value:

$$E(R_{\text{portfolio}}) = \sum_{i=1}^n X_i E(R_i)$$

where

X_i = the percent of the portfolio invested in asset i ,
 where the sum of the weights equals 1 (or 100%).
 $E(R_i)$ = the expected return for asset i .
 $E(R_{\text{portfolio}})$ = the expected return for the portfolio.

The second component required in the optimization process is a measure of portfolio risk, which is the variance of an asset’s expected returns. This is a statistical measure of the dispersion of possible returns around the mean expected return. As shown below, Markowitz’s measure of portfolio risk is a weighted function of the variances and covariances of the individual securities held within it and is computed as follows:^[8]

$$\sigma_{\text{port}}^2 = \sum_{i=1}^n w_i^2 \sigma_i^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n w_i w_j \rho_{ij} \sigma_i \sigma_j$$

where

σ_{port}^2 = variance of portfolio.
 w_i = proportion of total investment in security i .
 ρ_{ij} = correlation coefficient between security i and security j .
 σ_i^2, σ_j^2 = variance of securities i and j .
 $\rho_{ij} \sigma_i \sigma_j$ = the covariance of securities i and j .

From this equation for portfolio variance, it can be seen that when an investor invests in more than one

asset, the relevant risk of the portfolio is a function of the percentage of the portfolio invested in each asset and two distinct risk components. One of these components measures the contribution to portfolio risk by each asset's own variance. The other risk component measures the sum of the weighted covariances between the assets within the portfolio, i.e., how the returns of the assets move together. The first component is diversifiable security-specific risk, and the second is non-diversifiable market risk. The equation also demonstrates that the relevant risk of a security held in isolation is its own variance, because the covariance term in a one-asset portfolio will be equal to zero. But, as the investor diversifies by increasing the number of securities in the portfolio, the value of the first term on the right-hand side, a weighted average of individual security variances, diminishes rapidly. And in an optimal efficient portfolio, diversification will virtually eliminate this component, so that the investor's remaining risk is that attributable to market or systematic risk that cannot be diversified away. The greater this systematic risk, the greater will be the investor's expected return. In equilibrium, security markets will only reward the investor with higher-risk premiums for the systematic risk contained in a portfolio.

It can be shown theoretically and empirically that it takes only a small number of randomly selected securities in a portfolio to eliminate most of the non-systematic risk. Meir Statman's empirical research^[9] indicates that an investor must include between 30 and 40 stocks in their portfolio to eliminate non-systematic risk. Other research has indicated as little as 15 to 20 individual securities would accomplish the same task. The conclusions from portfolio theory show, as in Fig. 3, that diversifiable risk is reduced as securities are added to the portfolio, and that the

relevant, remaining risk is systematic and cannot be diversified away.

PORTFOLIO VARIANCE

Using the formula above, variance in the Markowitz two-asset portfolio is calculated as follows:

$$\sigma^2 = x_A^2 \sigma_A^2 + x_B^2 \sigma_B^2 + 2x_A x_B \rho_{AB} \sigma_A \sigma_B$$

where

$$x_A + x_B = 100\%.$$

$$\sigma^2 = \text{variance of the two-asset portfolio.}$$

$$x_A^2 = \text{percentage squared of the portfolio invested in an asset A.}$$

$$x_B^2 = \text{percentage squared of the portfolio invested in an asset B.}$$

$$\sigma_A^2 = \text{variance of return on security A.}$$

$$\rho_{AB} = \text{correlation coefficient for securities A and B.}$$

$$\sigma_B = \text{standard deviation of security B.}$$

$$\sigma_{AB} \sigma_A \sigma_B = \text{covariance of securities A and B.}$$

And, portfolio variance with three assets is computed as follows:

$$\begin{aligned} \sigma^2 = & x_A^2 \sigma_A^2 + x_B^2 \sigma_B^2 + 2(x_A x_B \rho_{AB} \sigma_A \sigma_B) \\ & + 2(x_A x_C \rho_{AC} \sigma_A \sigma_C) + 2(x_B x_C \rho_{BC} \sigma_B \sigma_C) \end{aligned}$$

The progression from two-asset to three-asset portfolios shows that as the number of securities in the portfolio increases, the number of covariance terms increases substantially, as do the computational complexity and estimation problems of the model. In the Markowitz setup, one can easily derive the minimum-variance portfolio for a small number of assets. For portfolios with large numbers of assets, however, the MVO process and construction of efficient portfolios becomes impractical, because the variance equation becomes quite complex, and the calculations become unwieldy due to the number of covariance terms that must be estimated to serve as inputs in the optimization process.

CAPITAL MARKET THEORY AND THE CAPITAL ASSET PRICING MODEL

The CAPM, developed by Sharpe, Lintner, and Mossin in the 1960s, provides an extension of Markowitz's portfolio theory framework and introduces a general equilibrium asset-pricing model with significant practical applications and theoretical implications. The CAPM optimization process for security selection is far less complex and reduces the number

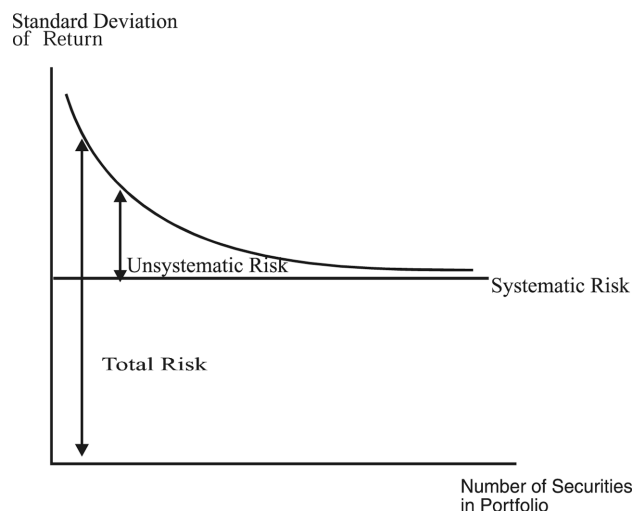


Fig. 3 Optimal portfolio selection with risk-free asset.

of calculations markedly compared to Markowitz's selection process for efficient portfolios. This simplification results from the CAPM's use of pair-wise comparisons of security covariance with the "market portfolio," rather than using the full covariance matrix between all asset pairs, as in Markowitz's optimization.

CAPM extends the Markowitz portfolio model through the introduction of a risk-free asset. Now, the investor is faced with choices among many risky assets and a risk-free asset. By definition, the risk-free asset has zero risk and zero covariance with the risky assets, i.e., its return is not correlated with the returns on risky assets. The optimization still requires the derivation of an efficient frontier of risky assets, as in Markowitz's optimization. However, with the introduction of the risk-free asset, the optimal portfolio is found at the point of tangency with a line drawn from the risk-free rate on the y-axis to the portfolio on Markowitz's efficient frontier that provides the greatest reward per unit of risk, as shown in Fig. 4.

This line creates a new efficient frontier depicting all the risk-return combinations for portfolios consisting of the risk-free asset and the market portfolio. This new set of efficient portfolios dominates all the portfolios on the Markowitz efficient frontier. These portfolio combinations create a Capital Market Line (CML) from which all investors will make their portfolio allocations between the risk-free asset and the market portfolio. Portfolio selection is a two-step process: 1) determining the optimal risky portfolio; and 2) choosing the allocation between risky and riskless asset. The second choice is based upon one's risk tolerance, and investors' allocations will differ only by the proportion of their portfolio invested in the risk-free asset. A key facet of the CAPM is that all investors choose the same risky portfolio, which is the market portfolio. In theory, the market portfolio includes all risky assets with the weight of each asset set in proportion to its

market value. In practice, however, many practitioners use a broadly diversified index such as the S&P 500 as a proxy for the "market" of all risky assets. The characteristic inducing all investors to choose the market portfolio is that it is completely diversified. Therefore, it contains only systematic risk, because all non-systematic risk is eliminated through diversification.

Sharpe shows that, in equilibrium, investors will choose the market portfolio in combination with the risk-free asset, because it provides them with the highest return per unit of risk. This equilibrium outcome paves the way for the derivation of the CAPM and the resulting Security Market Line from which expected asset returns could be determined.

SECURITY MARKET LINE AND ASSET PRICING

The CAPM equilibrium shows that an asset should earn a risk-adjusted return that is a function of the risk-free rate, the asset's Beta, and the expected return on the market. The resulting pricing relationship is known as Security Market Line (SML). Algebraically, the SML is as follows:

$$E(r_i) = r_f + \beta_i(r_m - r_f)$$

where

$E(r_i)$ = expected return on asset i .

r_f = the risk-free rate.

β_i = Beta of asset i .

r_m = expected return on the market.

If markets are efficient, all securities will be priced using the equation above.

A security's Beta is calculated as follows:

$$\beta_i = \frac{\text{COV}(R_i, \tilde{R}_m)}{\sigma_m^2}$$

where

$\tilde{R}_m$ = return on the market index.

σ_m^2 = the variance of the market return.

R_i = return on the security i .

COV = covariance.

For any asset i , the Beta is calculated by dividing the covariance of the returns on asset i and the market portfolio by the variance of the market return. It is a measure of the comovement of a security's return with the expected return of the market portfolio. The CAPM defines the Beta as the appropriate measure of a security's systematic risk for assets held within

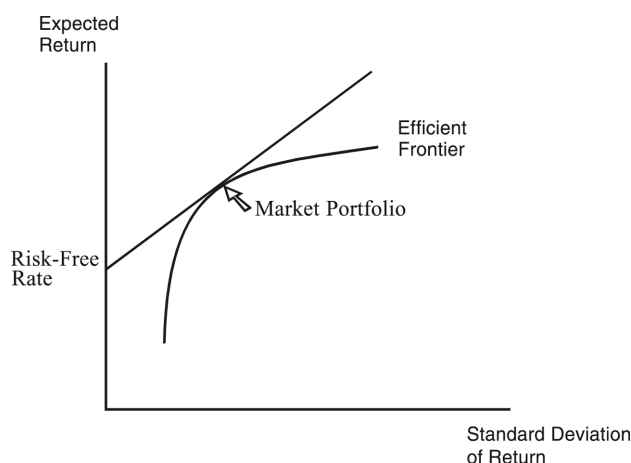


Fig. 4 Portfolio risk compared to number of securities in portfolio.

the diversified market portfolio. In theory, security Betas should be derived using *expected* return data from the individual securities within the portfolio. In practice, however, this is often done using historical security returns compared to an index such as the S&P 500. The importance of a security's Beta is in determining the required rate of return on a security and its corresponding value or price. Security Betas are widely reported and used extensively in investment and security analysis as a measure of volatility or risk. For a portfolio of securities, the portfolio's Beta is simply the weighted average of each asset's Beta.

The CAPM's contribution to portfolio theory is that it quantified the appropriate risk–return relationship for securities held within a portfolio in an efficient market. It also provided investors with a standardized measure of a security's systematic risk, i.e., its Beta. This equilibrium pricing model for individual securities allows one to compute the expected or required return for a risky asset in equilibrium, and from that, security prices and valuations can be performed. Implications from CAPM equilibrium include that there is no tendency for underpriced or overpriced securities and no expectation of superior performance from active portfolio management strategies. These implications result from the assumption that information will be fully priced and reflect all available relevant information, and the resulting equilibrium sets prices at their expected value in a diversified portfolio. This is known as market efficiency.

ARBITRAGE PRICING THEORY

The competing theory of asset pricing in the portfolio context is a multifactor equilibrium pricing model introduced in 1976 by Stephen A. Ross. It is known as the Arbitrage Pricing Theory (APT). It allows for multiple sources of risk premiums, or factors, whereas the CAPM is a single-index model based upon the market return and the security's Beta. The derivation of the APT does not require that investors select assets using mean-variance rule or that those investors are risk averse. However, using Ross's Arbitrage pricing construct, the APT also generates a linear relationship between risk and expected return.

APT allows for the pricing of risk premiums associated with multiple factors 1 through k as shown below:

$$\begin{aligned} E(r_i) = & rf + \beta_1(\text{risk premium for Factor 1}) \\ & + \beta_2(\text{risk premium Factor 2}) \\ & + \beta_3(\text{risk premium Factor 3}) + \dots \\ & + \beta_k(\text{risk premium Factor } k) \end{aligned}$$

where

$E(r_i)$ = expected return on security i .

rf = the risk-free rate.

β_k = Beta of security k for a particular risk factor.

Under APT, the expected return is determined by the aggregate of a number of systematic factors. Investors will be compensated for each non-diversifiable source of risk associated with a particular security. If factors 2 through k are 0, and the common factor generating security returns are the “market” portfolio, then the APT and CAPM will yield similar results and conclusions. The APT has fewer restrictive assumptions than CAPM, but it requires competitive capital markets and wealth-maximizing investors. Some work has been done on identifying different risk factors, but the APT does not indicate or provide guidance as to what the specific factors might be. This lack of factor identification has impeded the practical application of this methodology relative to the CAPM. In recent years, portfolio theory has been advanced by the introduction of a variety of new techniques using derivative securities such as futures and options. These developments have widened the range of tools available to optimize portfolio risk–return choices and extend the applicability of portfolio theory.

REFERENCES

1. Markowitz, H.M. Portfolio selection. *J. Finance* **1952**, 7 (1), 71–91.
2. Sharpe, W.F. Capital asset prices: a theory of market equilibrium under conditions of risk. *J. Finance* **1964**, 19 (3), 425–442.
3. Lintner, J. Security prices, risk, and maximal gains from diversification. *J. Finance* **1965**, 20 (4), 587–615.
4. Mossin, J. Equilibrium in a capital asset market. *Econometrica* **1966**, 34, 76–83.
5. Ross, S.A. The arbitrage theory of capital asset pricing. *J. Econ. Theory* **1976**, 13, 342–363.
6. Bodie, Z.; Merton, R.C. An overview of risk management. In *Finance*, 1st Ed.; Prentice-Hall Inc.: Upper Saddle River, NJ, 2000; 272 pp.
7. Bodie, Z.; Kane, A.; Marcus, A.J. Risk and risk aversion. In *Investments*, 5th Ed.; McGraw-Hill/Irwin: New York, 2002; 154 pp.
8. Reilly, F.K.; Brown, K.C. An introduction to portfolio management. In *Investment Analysis and Portfolio Management*, 6th Ed.; Harcourt Inc.: Orlando, FL, 2000; 268 pp.
9. Statman, M. How many stocks make a diversified portfolio? *J. Financ. Quant. Anal.* **1987**, 22 (3), 353–363.

Postmodernism and Public Policy

Thomas J. Catlaw

Margaret Stout

School of Public Affairs, Arizona State University, Phoenix, Arizona, U.S.A.

INTRODUCTION

While somewhat limited to date, distinct elements of postmodern thought can be discerned in the contemporary literature analyzing both the study and the making of public policy. Considered here are the meaning of postmodernism, the historical condition referred to as postmodernity, broad categories of postmodern theory, and a discussion of ways in which postmodern ideas have influenced the contemporary understanding of the study and making of public policy. In conclusion, future trends and criticisms of postmodern approaches in public policy are discussed.

WHAT IS POSTMODERNISM?

Postmodernism is a complex, multifaceted term that has been used to describe a disparate and heterogeneous array of sometimes conflicting research projects; critical discourses; artistic, architectural and cultural movements; as well as a distinct historical epoch. The prefix “post” further complicates any straightforward accounting of the postmodern, since the word presupposes a stable, consensus understanding of *modernism*, itself a term of significant contestation and debate. Indeed, some scholars argue that postmodernism is a merely acceleration of modernist tendencies rather than a distinct theoretical or historical break. This complexity, however, can be understood by examining the term in two interconnected dimensions: 1) as a new historical period or different social formation and 2) as a specific body of philosophical critique and theory that calls into question basic assumptions of modernity, reason, and the enlightenment.

The Condition of Postmodernity

The term *postmodernity* is used to describe a new historical era or different social formation in which former traditions, values, and modes of understanding no longer seem to capture or describe our experience in the world.^[1] In this regard, the term shares an

affinity with other descriptors such as postindustrial, postideological, post-Fordist, late capitalist, and, more recently, global. It bears noting that postmodernity as a term is itself not new. It has been used by scholars and artists for more than a century to name the emergence of new social, cultural, and economic formations seen to be substantively distinct from “modernity.”^[2]

Be that as it may, by the 1960s, new modes of communication, technology, and information, as well as the explosion of wealth and consumerism, seemed to be producing a new, more chaotic, unstable experience of the world, one arguably in marked contrast to the rigidity of what Max Weber called the “iron cage” of bureaucratized modernity. This new sensibility emerged together with new cultural and artistic forms and movements that emphasized different values. For example, disintegration was elevated over integration, fragmentation and deconstruction over unification and totalization, chance over design, play over seriousness, difference and otherness over identity or sameness, disparate textuality over metanarrative, eclecticism over purity, populist over avant-garde, and surface over depth.^[3]

These reorientations in value from the modern to the postmodern also challenged social, political, and cultural hierarchies. This allowed many excluded groups, discourses, and knowledge to force their way into mainstream society. As a result, modernity’s relatively coherent context for collective, authoritative meaning and identity—a stability purchased in no small part through the systematic and deliberate exclusion or marginalization of large groups of people from political, civic, and economic life—grew destabilized and people began to seek stability in smaller communities of meaning to compensate for this “thinning” of macroreality.^[4,5] While these communities can be affirmative and inclusive, they sometimes assume exclusionary forms quite at odds with the general postmodern embrace of otherness and difference, as in religious fundamentalist enclaves.

Complex changes in the political economy were also taking place, which were both producing and products of these social and cultural changes. In the industrialized parts of North America, Europe, and the Pacific

Rim, economies experienced a radical restructuring from centralized mass production to a more flexible, mobile form of capitalism that was far less accommodating to labor. In industrialized countries, new forms of communication and information technology increased focus on the production of services and cultural, informational, and symbolic commodities, shifting manufacturing to developing nations. Keynesian approaches to macroeconomic regulation of the business cycle as well as the social benefits and securities of the welfare state were gradually abandoned or seen as ineffectual, and a more limited role for government in the governing of society was advocated.

Some argued that these conditions were leading to the decline of the nation-state as the dominant form of political and social organization. While this is a debatable proposition, it remains the case that as societies fragmented culturally and economically, so did the work of governing. Several factors contributed to this. The late 1960s and early 1970s witnessed: 1) legal changes that made participation in regulatory and administrative proceedings easier for citizen groups; 2) growing demands from both the political Left and Right for local, community control; 3) an explosion in grassroots advocacy groups that grew out of the new social movements (e.g., environmentalism and women's and civil rights); and 4) in conjunction with the fragmentation and struggle for inclusion of marginalized groups noted above, the increasing pluralization of values, cultures, traditions, and styles of life.^[6]

As a result, traditional modes of policy making, so-called "iron triangles," were displaced by broader-based policy networks, as ever greater numbers of actors and points of view became involved in the formulation and execution of public policies. Moreover, fiscal uncertainty, the perceived failure of ambitious social engineering programs, and growing anti-governmental feelings and political conservatism precipitated the search for new decentralizing techniques of governing, such as privatization, contracting out, and coproduction. These reform efforts brought more actors and institutions into the governance process and reduced the visible size and scope of central governments.

Some have also contended that in these fragmented, media-saturated conditions, politics has become a game of "symbol manipulation" rather than a democratic dialogue about substantive issues.^[5] A related concern is expressed by those who feel that the policy-making process has become inappropriately dominated by interest group competition. As a result, many argue that elites rather than the broader public shape public policy through covert and manipulative activities, which restrict democratic participation and produce results that are neither in the interest of the society as a whole nor the specific targets of public policies.

Postmodern Theory

The second dimension of postmodernism is associated with a stream of radical critique that emerged from France in the 1960s. Following the Second World War, existentialism, Marxism, critical theory, phenomenology, and hermeneutics were the dominant intellectual tendencies. But by the 1960s, these ideas were challenged by linguistically based social scientific and philosophical inquiry that drew from the structuralist linguistics of Ferdinand de Saussure (1857–1913) and his critics that followed. The shift from structuralism to poststructuralism in philosophy is often used to mark the theoretical emergence of the constellation of ideas comprising "postmodern theory." These "post-" theories argue that structuralism had not gone far enough in severing connections with the assumptions of modernity, humanism, and theoreticism. Major postmodern theorists include Jean Baudrillard, Jacques Derrida (1930–2004), Michel Foucault (1926–1984), Jacques Lacan (1901–1981), and Jean-Francois Lyotard (1924–1998). American pragmatist philosophy also has been an important influence on the development of postmodern thought (e.g., Richard Rorty (1931–2007)).

While there is considerable variation among postmodern theorists and their followers, three core components in postmodern theory can be identified.^[7] First, the *critique of the human subject* calls into question the notion of an autonomous, coherent, and substantial self. In postmodernism, the self becomes an effect of power (e.g., Foucault) or discourse. It is exploded into a multiplicity of partial identifications or intensities without a fixed center for reflection, critique, or action. Second, the *critique of historicism* rejects the teleological view of history, the idea that there is an underlying meaning, pattern, or trajectory to historical events. Related to this antihistoricism is the well-known characterization of postmodernism as rejecting all *metanarratives*, or the possibility of a single, overarching narrative that can interpret all objects and phenomena. In this regard, postmodern theory prefers local micronarratives. Third, the *critique of meaning* rejects the postulate that words refer to substantial things and that we can ever have an unmediated access to reality. Instead, the world and our experience of self are socially constructed by and through language. This general stance of *antirepresentationalism* is perhaps the single most important notion in postmodern thought, arguing that language does not "mirror" reality because there is not a given object "out there" to be re-presented. Rather, objects are produced by and apprehended through language. Politically, for many postmodernists, this critique of epistemology is also linked to critique of political representation and a preference for participatory

practices. This can be a position somewhat at odds with the critique of the subject.

AFFIRMATIVE AND SKEPTICAL POSTMODERNISTS IN SOCIAL SCIENCE

Emerging primarily from the tradition of continental philosophy, the postmodern critiques share much with postpositive philosophy of science, which is associated with the analytic philosophical tradition. Like postmodern theory, postpositivism rejects the core tenets of conventional, positivist science, i.e., realist ontology, objectivist epistemology, and a strict separation of facts and values. While there is controversy concerning the difference between postpositivism and postmodernism, some postpositivists argue that the postmodern position critique of meaning and its insistence that all experience is constituted by language are too extreme. There also can be political differences. Reflecting the influence of Marxism and philosopher Friedrich Nietzsche, postmodernists frequently advance a more radical critique of existing political and social institutions; postpositivists often retain a commitment to and faith in liberal institutions. Others choose to avoid the term postmodern altogether, believing that it has acquired a negative, even derogatory, connotation.

Even among postmodernists, however, there is variation that challenges any neat categorization. Rosenau draws a distinction between “affirmative postmodernists” and “skeptical postmodernists.”^[8] Whereas skeptics may reject the idea of objectivity and possibility for discriminating among truth claims altogether, affirmative postmodernists accept that knowledge and experience are linguistically mediated and socially constructed but argue that we still can know something about the world. They seek a “philosophical and ontological intellectual practice that is non-dogmatic, tentative, and non-ideological” (on page 16 of Ref.^[8]) and adopt a reflexive, flexible attitude with regard to our concepts and constructions.^[9] Affirmatives contend that objectivity can be redefined as intersubjective agreement. This is a *consensus theory of truth* that stands in contrast to the law-like explanatory reasoning designed for prediction and control that characterizes positivism.

POSTMODERNISM IN PUBLIC POLICY

In general, the reception and application of postmodern ideas in the field of public policy research and analysis can be characterized as an affirmative postmodernism: With an acceptance of social construction and intersubjective meaning, collective understanding,

decision, and action are considered possible. Each of the postmodern critiques (i.e., of the human subject, historicism, and meaning) has made an impact on the study and making of public policy. For example, to confront the effects of power and language on the human subject, critical and interpretive approaches to policy studies and participative practices are advanced. To avoid historicism and metanarratives, local and subjective interpretations of meaning must be accommodated in both studies and participative policy analysis. To address the critique of meaning, multiple forms of knowledge, without privileging any type or social actor, are advocated in the policy process. Furthermore, challenges to representation as not only a form of meaning but also a political institution demand a more directly democratic and discursive practice in the making of policy.

Specifically, dimensions of postmodern theory’s critique as well as the conditions of postmodernity are evident in contemporary policy scholarship in three principal areas: 1) the critique of rationality in policy making; 2) postpositive approaches to policy studies (sometimes calls “postempirical”); and 3) the call for a more participatory approach to policy making. Together, the rational approach to policy making and the positivist approach to its study have been labeled the “rationality project” by Stone.^[10]

Critique of the Rational Approach to Policy Making

Postmodern critique calls into question the bases of the rationalist approach to policy formulation. As an idealized prescription for policy making, the rational approach assumes that policy develops according to an orderly five-stage process. First, objectives are identified; second, alternative courses of action are postulated for reaching the object; third, the consequences of each alternative are predicted; fourth, the consequences are evaluated; and fifth, the alternative that maximizes benefits and minimizes costs is selected. This model of reasoning assumes that there is a single vantage point from which judgments about the good can be made. It also assumes a specific conception of society, namely the market, which is viewed as “a collection of autonomous, rational decision makers who have no community life” (on page 9 of Ref.^[10]).

Both the empirical veracity and the prescriptive ideal of the model have been challenged. In regard to the former, while Kingdon should not be labeled as a postmodernist, his work shows that the process by which policy is made hardly conforms to the stages of the rational model.^[11] Kingdon analyzes the complex “primeval soup” (on page 19), in which experts, citizens, politicians, and administrators interact and

shape the way in which issues get on the agenda. This political context shapes the identification and selection of alternatives. Kingdon argues that there are largely independent streams that flow through the process—a policy stream, a problem stream, and politics stream. There is no single, overarching way to describe how these streams come together or, as Kingdon writes, how a “policy window” opens.

More clearly aligned with postmodern theory, other scholars challenge the normative prescription of economic rationality from various normative perspectives. Stone argues that the rationality project aims to purge society of political conflict. In dismissing politics, the rational approach to policy making misses the fact that policy problems and objectives are not given but are constructed through specific categories of language and schemas of classification. As such, policy problems have multiple, conflicting perspectives and interpretations and are infused with paradox, difference, and conflict. For example, Stone shows how policy can be changed by redefining the category through which people understand causality. Are people on welfare because they are inherently lazy or have been made so by the welfare system? Or are they impoverished because of pervasive disenfranchisement as marginalized groups? The two causal stories lead to very different public policies. To govern democratically in these conditions, Miller, perhaps unique in working through the full implications of the postmodern critique for public policy, argues that we need institutions and public policies that ensure against “the domination of any one truth, any singular idea, or any unifying system of social practice” (on page 99 of Ref.^[12]).

All these perspectives challenge the traditional authority of the policy expert. From this perspective, democratic governance is harmed by the usurpation of political deliberative processes by technical experts and decision-making processes that are not broadly inclusive.^[10] This changes both the nature of policy research and the policy making.

Postpositive Policy Studies

Miller argues that “we should see policy research as stemming from the necessity to form common understandings with others who are interested in the same vexations and the same social problems” (on page 67 of Ref.^[12]). There cannot, therefore, be simply one expert view or perspective that captures the complexity of a social problem. Rather, there are many possibilities for framing inquiry. Extending these concerns to policy studies, the postpositive approach to public policy adopts a general ethos of methodological pluralism, combining empirical and normative inquiry without privileging one particular form of rationality,

quantitative over qualitative methods or expert over citizen roles. This accommodates a “reconstructed methodology” that produces a more well-rounded scientific knowledge.^[13]

Consistent with the postmodern awareness that the world is socially constructed, meaning (including both fact and value) is made through language, actions, and cultural objects, including theory. Thus, all research, including policy analysis, is a process of interpreting facts and values, both of which are mediated by theory, culture, and language. When the focus is on culture and language, this approach to policy studies is often referred to as an argumentative, rhetorical, communicative, or linguistic turn. When the focus is on theory, it is asserted that theories, methods, and rules guide interpretation, explanation, prediction, and evaluation.^[14] Therefore, theory, fact, and value cannot be segregated within policy discourses. This further challenges the position of the expert.

The common denominator among these approaches is that they all take an *interpretive* view that “the creation, communication, and understanding of meaning require attention” (on page 111 of Ref.^[15]). These policy scholars focus attention on the way in which people use linguistic categories, symbols, and frames to interpret the world and seek to understand how these interpretations impact the issues considered, the policy tools preferred, and the legitimacy of the outcomes sought. Assuming goals of understanding rather than prediction and intersubjective agreement rather than objectivity, policy analysis and inquiry become an ongoing process of interpretation or critique that includes multiple forms of knowledge and models of reasoning, a process that extends to policy research itself. Scholars and policy experts must be reflexive and aware of how their own frames and theories affect their work.

Participatory Policy Analysis

Postmodern conditions can be cause for deep concern insofar as they seem to replace substantive democratic politics with empty symbolism and elite manipulation. However, for some policy scholars, the fragmented, perspectival, and uncertain nature of the “postmodernized” policy process offers the conditions for the rebirth of a robust, participatory democratic politics.^[16] Postmodern theory also points in this direction.

Postmodernized policy studies seek to answer the question “Why?” by understanding from the perspective of the subjects of inquiry. When applied to practice, then, one of the most effective ways to understand the perspectives of those impacted by the issue of concern is to include them directly in the policy-making process. Rather than enforcing sometimes arbitrary

divisions between experts, politicians, and citizens, boundaries are blurred and all policy actors become “policy analysts” together. Indeed, many scholars call for direct participation in the policy-making process as the way to ensure the best interpretive outcomes.^[17–19] This implies an approach to citizenship, governance, and public policy that is non-hierarchical and egalitarian in nature. As such, some view it as a challenge to the “broken” institutions of representative government, replacing them with deliberative and discursive forms of democracy.

Postmodern theoretical critique also points in this participatory direction. The critique of the subject suggests that people’s values and interests are not static, as assumed by the rationalist approach to policy making. Instead, through dialogue and deliberation preferences can change. Furthermore, our own understandings of our true interests may not be transparent due to the effects of social power. To avoid metanarratives, local interpretations of meaning must inform the determination of the problem and alternatives for policy action. Expertise has an important role to play, but it cannot be the only narrative that informs policy. To address the critique of meaning, policy making must make use of multiple forms of knowledge without privileging any type of social actor, and challenges to representation demand a more directly democratic and discursive practice in the making of policy.

In sum, scholars recommend participative practice to overcome the problems identified by postmodern theory as well as the conditions of postmodernity.

CRITIQUES OF POSTMODERNISM IN PUBLIC POLICY

Postmodern approaches to the study and making of public policy are not without their critics. Some suggest that postmodernism is nothing more than an extension of critiques that have been ongoing in the philosophy of science as well as public policy for some time and, as such, that postmodern theory attacks a highly caricatured image of science. For example, elite theory has suggested concerns about inclusion, authority, legitimacy, and justice in the making of democratic public policy from its beginning. However, postmodern theories would argue that these critiques are still attached to the underlying positivist philosophy and explanatory mode of research that they purport to critique and that the rational model remains the prescriptive ideal for policy making. In this connection, scholars also challenge postmodern public policy studies, analysis, and evaluation based on an epistemological and methodological resistance to non-positivist science.

A related critique is that participative and communicative practices do not ensure that valuable

scientific and technical knowledge will be considered, thus failing to produce the best policy outcomes. Another critique is that even when practical direction is offered, it fails to meet its democratic objectives. For example, a common complaint about any type of participative or discursive practice is that it is too time consuming and lengthy for “real-world” policy-making and is thus cut short or trivialized. Both of these barriers may be due to the fact that our society is simply not socialized toward political participation. Finally, some postmodernists critique the very ideas of policy and reason.^[20] For these theorists, postmodern public policy does not go far enough.

CONCLUSION

Just as it has taken several decades for the limitations of positivism to become evident, the jury remains out on what the long-term impact of postmodern policy research and applied methods will be on social science and society. However, it seems likely that so long as the conditions of postmodernity continue, postmodern theory will continue to play an important role in shaping how we understand and practice the study and making of public policy.

ARTICLE OF FURTHER INTEREST

Postpositivist Perspectives in Policy Analysis, p. 1530.

REFERENCES

1. Harvey, D. *The Condition of Postmodernity*; Blackwell: London, 1990.
2. Best, S.; Kellner, D. *Postmodern Theory: Critical Interrogations*; Macmillan: London, 1991.
3. Hassan, I. The culture of postmodernism. *Theor. Cult. Soc.* **1985**, 2 (3), 119–132.
4. Catlaw, T.J. Constitution as executive order: the administrative state and the political ontology of “We the People.” *Admin. Soc.* **2005**, 37 (6), 445–482.
5. Fox, C.J.; Miller, H.T. *Postmodern Public Administration: Toward Discourse*; Sage Publications: Thousand Oaks, CA, 1995.
6. Spicer, M.W. *Public Administration and the State: A Postmodern Perspective*; University of Alabama Press: Tuscaloosa, AL, 2001.
7. Sarup, M. *An Introductory Guide to Post-Structuralism and Postmodernism*; University of Georgia Press: Athens, GA, 1989.
8. Rosenau, P.M. *Postmodernism and the Social Sciences: Insights, Inroads, and Intrusions*; Princeton University Press: Princeton, 1992.

9. Farmer, D.J. *The Language of Public Administration: Bureaucracy, Modernity, and Postmodernity*; University of Alabama Press: Tuscaloosa, AL, 1995.
10. Stone, D.A. *Policy Paradox: The Art of Political Decision Making*, Revised Ed.; W. W. Norton & Company: New York, 2002.
11. Kingdon, J.W. *Agendas, Alternatives, and Public Policies*, 2nd Ed.; Addison Wesley Longman: New York, 2003.
12. Miller, H.T. *Postmodern Public Policy*; State University of New York Press: Albany, NY, 2002.
13. Fischer, F. *Evaluating Public Policy*; Nelson-Hall Publishers: Chicago, 1995.
14. Rein, M. Value-critical policy analysis. In *Ethics, the Social Sciences, and Policy Analysis*; Callahan, D., Jennings, B., Eds.; Plenum Press: New York, 1983; 83–111.
15. Yanow, D. Practices of policy interpretation. *Policy Sci.* **1995**, 28 (2), 111–126.
16. Danzinger, M. Policy analysis postmodernized: some political and pedagogical ramifications. *Pol. Stud. J.* **1995**, 23 (3), 435–450.
17. deLeon, P. The democratization of the policy sciences. *Public Admin. Rev.* **1993**, 52 (2), 125–129.
18. Durning, D. Participatory policy analysis in a social service agency: a case study. *J. Policy Anal. Manage.* **1993**, 12 (2), 297–322.
19. Fischer, F., Forester, J., Eds.; *The Argumentative Turn in Policy Analysis and Planning*; University of North Carolina Press: Durham, NC, 1993.
20. McSwite, O.C. *Legitimacy in Public Administration: A Discourse Analysis*; Sage Publications: Thousand Oaks, CA, 1997.

Postpositivist Perspectives in Policy Analysis

Göktuğ Morçöl

School of Public Affairs, Pennsylvania State University at Harrisburg,
Middletown, Pennsylvania, U.S.A.

INTRODUCTION

Postpositivism is a group of philosophical perspectives that are critical of positivist assumptions and methods. In philosophy and social theory, they emerged in the early- to mid-20th century and, in the policy analysis literature, in the 1980s.

Both positivism and postpositivism are vaguely defined terms. *Positivism* is sometimes used interchangeably with *empiricism*, *behaviorism*, and *naturalism*. Two philosophical movements called themselves positivist: the 19th century French sociologist August Comte coined the term and the Vienna Circle of philosophers of the early 20th century (Rudolf Carnap, Otto Neurath, Moritz Schlick, and others) adopted the name *logical positivist* (or *logical empiricist*). Some critical theorists of the Frankfurt School define positivism more broadly and label Karl Popper and the later writings of Karl Marx positivist.

The philosophical predecessors of postpositivism are the hermeneutic theories of the 19th and 20th centuries. The critiques of positivism by Karl Popper, the critical theorists of the Frankfurt School, and Thomas Kuhn established postpositivism as a philosophical perspective. Postpositivist philosophical arguments are reflected in the policy analysis and evaluation literature in five theory streams: the contextuality and presupposition theories, problem structuring and issue framing theories, the methodological critique, critical hermeneutics, and participatory policy analysis theories.^[1]

Weimer^[2] and others question the relevancy of postpositivist arguments to policy analysis, because, they think, most of policy-analytic theory and practice is not even positivistic. On the other hand, postpositivists such as Fischer^[3] argue that the failures of the positivist policy analytic theory and practice in the 1960s and 1970s led to the surge of the postpositivist critique in the 1980s and that positivism survived despite the critique because it still is professionally, organizationally, and politically expedient. Morçöl^[4] found in a survey that positivistic assumptions are prevalent at varying degrees among policy professionals.

ASSUMPTIONS OF POSITIVISM

Although the exact roots and scope of positivism are debatable, a set of assumptions can be identified as

its core: a realist and deterministic ontology, an objectivist epistemology, and the belief in a fact-value distinction. Positivists use reductionist and analytical methods and believe in the primacy and superiority of quantitative methods.

Realism is the belief that the world exists independently of and prior to the knowledge of a knowing subject. The logical positivist of the early 20th century as well as August Comte and Karl Marx were realists. Positivist policy analysts assume that policy problems exist independently and can be identified and solved objectively.

Positivists also believe that this independent reality is composed of discrete and precisely identifiable objects whose relations and motions are determined. The notion that the universe is deterministic has a long history; the most recent version, one that undergirds the modern science, was articulated by Francis Bacon, René Descartes, and, most importantly, Isaac Newton in the 17th century. In the Newtonian deterministic universe, the relations between past and future events are fixed. Therefore if a complete knowledge of the past is obtainable, then future events can be predicted precisely. The deterministic worldview is reflected in policy research in the use of the experimental method, which attempts to establish causalities between variables; the forecasting methods that use information about the past to predict future; and the theory and methods of long-term planning.

The notion that the objective knowledge of an independent reality can be acquired is a core positivist assumption. The validity of the acquired knowledge can be verified by checking it empirically against the reality itself. What follows is that factual knowledge—knowledge that has referents in reality and is verifiable—can, and should, be separated from values—inner, subjective preferences of individuals. Positivist policy analysis attempts to keep the respective realms of politics and analysis apart, on the basis of the belief that while values belong to politics, facts should be the domain of analysis.

APPLICATIONS OF POSITIVISM IN POLICY ANALYSIS

The reductionist/analytical methodology of positivism builds on the Newtonian scientific notion that objects are discrete and precisely identifiable and their

characteristics (weight, length, etc.) can be counted, measured, and compared with one another. It is also assumed that a quantitative form of knowledge is inherently and under all circumstances more scientific and objective than a qualitative form of knowledge. Such assumptions are reflected in the notion that quantitative pieces of information are “hard facts,” while qualitative ones are “soft,” and in the extensive use of statistics in policy research and policymaking.

The assumption underlying cost–benefit analysis is that costs and benefits of a program or project can be identified objectively and measured. The experimental method is considered the most scientific, primarily in psychology and policy and program evaluation, because it is assumed that experiments can isolate variables and establish deterministic relations (causalities) among them definitively.

POSTPOSITIVIST CRITIQUE IN PHILOSOPHY

The philosophical roots of postpositivism can be traced to hermeneutics. In the most general sense of the term, hermeneutics is the study of meanings in symbolic interactions (speech, written text, rituals, cultural artifacts, etc.). Hermeneutic philosophers follow Wilhelm Dilthey's (1833–1911) distinction between the natural and social sciences: while the natural sciences aim to develop causal explanations of “outer” events, the social sciences are concerned about an “inner” understanding of meaningful human actions; therefore they should use different methods of inquiry.

Hermeneutic philosophers see social phenomena as texts to be interpreted. There are differences among hermeneutic philosophers as to whether there is a “true meaning” of a text, but the position of the German philosopher Hans-Georg Gadamer (1900–2002) is the most typical one. According to Gadamer, the meaning of a text is determined by an interaction between the intentions of the author and the interpretations of the reader, both of which are embedded in their respective particular historical contexts. It is not possible for a reader to know the entire historical context of the authorial intentions and hence the “true meaning” of his/her actions.

The Frankfurt School of critical theorists—Max Horkheimer (1895–1973), Theodor Adorno (1903–1969), Jürgen Habermas (1929–), and others—synthesized hermeneutics with Hegelian philosophy and Marxism. Habermas affirms the hermeneutic distinction between the natural and social sciences and argues that if we attempt to use the methods of the natural sciences in the social sciences, we turn the latter into “social technologies.” This instrumentalist use of knowledge in the social sciences cannot be objective; in fact, it is used to perpetuate the social structures

of domination. Habermas sees the task of the critical theory as to contribute to an emancipatory knowledge of society.^[5]

In his critique of logical positivism, Karl Popper (1902–1994) aimed to show that scientific knowledge could not be positively verified and offered a *falsificationist* alternative: if a hypothesis has not been falsified, it can be accepted only as a tentative truth. Because Popper shared some of the core convictions of logical positivists (e.g., that science offers the most reliable form of knowledge because it is open to empirical testing), the critical theorists Adorno and Habermas characterized his work as an “internal critique” of positivism.^[5] However, his demonstration of its weaknesses opened the door to a deeper critique of positivism.

Thomas Kuhn (1922–1996) offered a sweeping critique of the positivist understanding of science in his book *The Structure of Scientific Revolutions* (1962). He argued that scientists work in *paradigms*, belief systems that are “incommensurable.” Paradigms are shaped in social and historical contexts. They determine which methods to be used and how the “facts” are “discovered.” Therefore the positivist principle of fact-value dichotomy and hence its objectivist epistemology are not sustainable.

POSTPOSITIVISM IN POLICY ANALYSIS

Fischer^[3] points out that the social crises and the failures of governmental policies in the United States in the 1960s and 1970s forced a rethinking of the assumptions and methods of the social sciences; those social scientists who were concerned about the role the social sciences played in the ideological manipulation of the poor and disadvantaged groups in the society began to formulate alternatives to positivist, instrumentalist, technocratic, and undemocratic forms of policy analysis. In the following decades, five theory streams emerged that are critical of positivism in general or some aspects of it: contextuality and presupposition theories, problem structuring and issue framing theories, the methodological critique, critical hermeneutics, and participatory policy analysis theories.

The *contextuality and presupposition theorists*, such as M.E. Hawkesworth, make the case that the theory and practice of policy analysis are predominantly positivistic.^[6] They also argue that policy-analytic knowledge is presupposed—mediated by analysts' preconceived notions and values—and formed in historical, cultural, and political contexts. With such assertions, this group of theorists provided the general justification for a postpositivist critique of the mainstream policy analysis.

The *problem structuring and issue framing theories* also recognize the contextuality of policy-analytic

knowledge and focus particularly on the significance of the way policy problems are structured. William Dunn points out that policy problems are typically ill-structured because they are socially constructed, dynamic, and interdependent.^[7] He argues for a conscious, systematic, and continual structuring and restructuring of problems throughout the policy analytic process. Schön and Rein^[8] point out that participants in policy processes construct problems through their mental frames in which facts, values, theories, and interests are intertwined. Schön and Rein call for a *frame-critical policy analysis*, in which the role of the analyst would be to understand frame conflicts and help participants reframe to reach policy resolutions.

In their joint and individual works, the late Donald Campbell (1917–1996) and Thomas Cook articulated a *methodological critique* of positivist methods. Campbell argued that positivists failed to recognize the equivocal and ambiguous nature of experimental research and the historical indexicality of the variables used. Cook and Campbell^[9] offered quasi-experiments as an alternative to remedy the shortcomings of the experimental method. Cook^[10] recognized the inability of positivistic methods to generate objective and precise knowledge by arguing for a *critical multiplist* methodology that would be anchored in realist ontology, but it would also use multiple methods, multivariate causal models, and multiple theoretical frameworks.

Hajer^[11] and Fischer^[3,12] represent the *critical hermeneutic* perspective in postpositivism. In his discourse analyses, Hajer shows that knowledge and power are articulated through discourse and human actors engage in struggles for discursive hegemony in policy processes. The influence of Habermas' critical theory can be seen most clearly in Fischer's works. He is critical of the instrumentalist and technocratic uses of the policy-analytic knowledge by analysts and policymakers because such uses support undemocratic practices. He proposes a participatory policymaking process, with the involvement of multiplicity of actors and interests. He emphasizes that the policy-analytic knowledge should also be transformational and emancipatory.

Participatory policy analysis has other proponents as well, although they do not necessarily come from a critical hermeneutic perspective. Kelly and Maynard-Moody,^[13] for example, suggest that policy stakeholders should be involved in framing research questions and selecting the methods of inquiry. The task of an analyst should be to facilitate rational deliberations, help policymakers and citizens understand the limitations of their perspectives, and synthesize multiple perspectives. Since the early 1990s, there is a growing literature on the use of participatory methods in the policy process—stakeholder analysis, Q-methodology, citizen panels, Internet-facilitated

dialogue and analysis, and the like—in the leading journals of the field, such as *Policy Studies Journal*, *Policy Studies Review*, and the *Journal of Policy Analysis and Management*.

CONCLUSION

Postpositivists are not unified in their critique of or alternatives to positivism, but, in general, they emphasize the historical, social, and political contextuality of the policy-analytic knowledge, argue and demonstrate that facts and values are intertwined in the policy process, call for an understanding of the discursive nature of policymaking, expose the power relations in the policy process, and call for a more democratic and participatory policy process in which analysts would play a facilitative role.

There is another theory stream that shares some of the assumptions and aspirations of postpositivism: postmodernism/poststructuralism. *Postmodernist/poststructuralist* theories also have their roots in hermeneutics. Because of this commonality in philosophical roots and the vagueness in the uses of the terms postpositivism and postmodernism/poststructuralism, it is not easy to draw clear demarcations between the two. Fischer offers a distinction: while postpositivists see a chance of establishing a common basis for a valid discourse, truthfulness of interpretations, among different perspectives, postmodernists do not.^[3]

Postmodernists/poststructuralists deny any “epistemological privilege” to any interpretation over another. In the works of Ferdinand De Saussure and Jacques Derrida, language is defined as a closed, self-referential system. As such, language does not have any reference to an “external reality”; therefore there is no basis for establishing truthfulness in knowledge. Poststructuralists use the method of deconstruction to demystify texts by revealing the arbitrary hierarchies and power bases in the seemingly rational and scientific discourses. They aim to delegitimize the socially constructed bases of authority and level the playing field for alternative, underprivileged discourses.

Schram's^[14,15] works are examples of the postmodernist/poststructuralist approach in policy studies. He aims to expose the power relations embedded in the public policy process. Schram denies any “epistemological privilege” to social scientists and policy analysts in the analytical process and argues that the policy analysis practice should be abandoned and policy analysts and academics should join the political struggle for social change.

Postpositivism and postmodernism/poststructuralism are open-ended projects. What direction each will take and whether they will merge into a larger stream remain to be seen.

REFERENCES

1. Morçöl, G. *A New Mind for Policy Analysis: Toward a Post-Newtonian and Postpositivist Epistemology and Methodology*; Praeger: Westport, CT, 2002.
2. Weimer, D.L. Comment: Q-method and the isms. *J. Policy Anal. Manag.* **1999**, 18 (3), 426–429.
3. Fischer, F. *Evaluating Public Policy*; Nelson-Hall Publishers: Chicago, 1995.
4. Morçöl, G. Positivist beliefs among policy professionals: an empirical investigation. *Policy Sci.* **2001**, 34 (3–4), 381–401.
5. Giddens, A. *Politics, Sociology and Social Theory*; Stanford University Press: Stanford, CA, 1995.
6. Hawkesworth, M.E. *Theoretical Issues in Policy Analysis*; State University of New York Press: Albany, NY, 1988.
7. Dunn, W.N. *Public Policy Analysis: An Introduction*, 2nd Ed.; Prentice Hall: Englewood Cliffs, NJ, 1994.
8. Schön, D.A.; Rein, M. *Frame Reflection*; Basic Books: New York, 1994.
9. Cook, T.D.; Campbell, D.T. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*; Houghton Mifflin: Boston, 1979.
10. Cook, T.D. Postpositivist critical multiplism. In *Social Science and Social Policy*; Shotland, R.L., Mark, M.M., Eds.; Sage: Beverly Hills, CA, 1985.
11. Hajer, M.A. *The Politics of Environmental Discourse: Ecological Modernization and the Policy Process*; Oxford University Press: Oxford, UK, 1995.
12. Fischer, F. *Technocracy and the Politics of Expertise*; Sage: Newbury Park, CA, 1990.
13. Kelly, M.; Maynard-Moody, S. Policy analysis in the post-positivist era: engaging stakeholders in evaluating the economic development districts program. *Public Adm. Rev.* **1993**, 53, 135–142.
14. Schram, S.F. Postmodern policy analysis: discourse and identity in welfare policy. *Policy Sci.* **1993**, 26, 249–270.
15. Schram, S.F. Against policy analysis: critical reason and poststructural resistance [Book review essay]. *Policy Sci.* **1995**, 28, 375–384.

Practical Organization Theory

Gerald T. Gabris

Division of Public Administration, Northern Illinois University, DeKalb, Illinois, U.S.A.

INTRODUCTION

For many public managers, the relevant question is not “why does something work,” but more simply, “what does work? For practitioners of public administration, the quest has always been toward the practical, the workable, and the doable. Defining and explaining what “practical organization theory” is all about serves as the focus for this article. Practical theory (PT) eschews simple definition. Yet for the purposes of this discussion, PT will be defined as follows. Practical organization theory provides simplified solutions to common issues and problems managers face in public organizations based on more complex and abstract theory. The simplification of theory provides managers with ready answers to practical problems. Simplification, however, can incur a price. Practical models often have somewhat limited application, they may generate contradictory advice, and they may oversimplify situations. Nonetheless, practicing managers are often willing to accept these risks given the direct applicability of PT. This article will first describe the methods and vehicles PT utilizes to become user-friendly to managers, followed by a select description of PT examples reflecting these methodologies. This article will conclude with a brief consideration of potential PT pitfalls.

VEHICLES FOR MAKING PRACTICAL THEORY USER-FRIENDLY

At the outset, it should be stated that PT encompasses such a broad range of concepts, models, and remedies that it would be impossible to categorize all these in a short descriptive essay. Thus, our purpose focuses on the common characteristics of PT rather than serving as a source for all PT possibilities. Central to this task involves identifying the “methodology” or medium through which PT is communicated and explained to practitioners, so they can use it quickly and expeditiously. With the risk of omitting some PT methods, the list below provides perspective:

- The use of principles or platitudes.
- The use of acronyms.
- The use of simple diagrams, charts, or figures.

- The use of metaphor and how-to-do-it books.
- The use of stories and anecdotes.

Although other media for PT exist, the above list captures the lions share. Let us review each of these briefly to understand how they work.

PRINCIPLES OR PLATITUDES

Principles normally involve simple statements that tend to proselytize some organizational or managerial “wisdom.” Examples might include the “Peter Principle” or notion that everyone rises to a level of their incompetence.^[1] Other off-cited principles include unity of command, span of control, and the exception principle.^[2] These ideas can be helpful to public managers. There should be only one boss, hence, avoid project management models. Limit the number of subordinates that any one supervisor is responsible for about six, and only bring exceptions to managers for decisions. Otherwise, follow standard operating procedures. These principles can be used practically to structure an organization, to firm up chain-of-command, to determine the number of managerial staff needed, and to solidify work routines. No wonder principles and platitudes are so popular among public managers, they tell them what to do and what works without a messy theoretical explanation.

ACRONYMS

Acronyms are merely words formed from the first letter or groups of letters from larger, more complicated phrases. One of the most renowned acronyms in management is POSDCORB.^[2] POSDCORB stands for “planning, organizing, staffing, directing, coordinating, reporting, and budgeting.” By using an acronym, a manager has a nifty memory device that enables one to remember their responsibilities without having to memorize each one individually. What constitutes my job, well, POSDCORB. Acronyms simplify. An acronym common to human resources managers is “KSAs.” This informs HR staff that in interviewing people for jobs they need to assess the applicants “knowledge, skills, and abilities.”

DIAGRAMS, CHARTS, OR TABLES

This is a somewhat broader category that is relatively self-descriptive. Again, PT diagrams or charts strive to simplify a complex idea or theory to make them more user-friendly to managers. An example of this would be Robert T. Golembiewski's simplified bureaucratic organization chart. This is replicated in Fig. 1.^[3]

This simplified organization chart applies to any public bureaucratic organization. There is a general manager (GMabc), specialized supervisors (S1, S2, and S3), and specialized employees (little a, b, and c). From this model, we can deduce much about bureaucratic behavior, including communications, upward mobility, chain of command, task responsibility, and interunit competition. The model cuts to the quick regarding bureaucratic behavior, thus facilitating a useful, user-friendly mechanism that public managers can understand quickly. Another figure or diagram simplifying complex ideas can be related in Fig. 2, which depicts the characteristics of degenerative and regenerative interpersonal interaction patterns.^[3]

In Fig. 2, we learn that low trust associates with low openness (communications), higher risk, and low owning (taking a stand for one's values). Put another way, a problem in one area, say trust between coworkers, usually precipitates a ripple effect that influences positively or negatively, other interpersonal variables. Thus, the diagram serves as a simple diagnostic mechanism enabling managers to clearly see how a downturn in one area of interpersonal relations may red flag problems in other areas.

METAPHORS AND HOW TO DO IT BOOKS

Practical theory can be expressed in simple metaphors and analogies that provide managers with an instant understanding of a practical concept. One good example of this is the short book by Blanchard and Johnson entitled, *The One Minute Manager*.^[4] Essentially, this book is a short compendium of multiple management principles conveyed in story form. The advice offered reflects practical managerial common

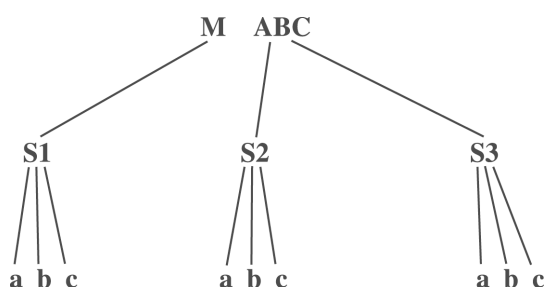


Fig. 1 The bureaucratic model.

Degenerative			Regenerative		
Status		Characteristic		Status	
Low	-	TRUST	-	High	
Low	-	OPENNESS	-	High	
High	-	RISK	-	Low	
Low	-	OWNING	-	High	

Fig. 2 Degenerative and regenerative interpersonal interaction patterns.

sense such as taking a minute to form one's goals, giving feedback, praising employees, and explaining decisions to subordinates. Most of these things take only a short time but produce big results, hence, the metaphor, "one minute manager."

A somewhat similar approach is utilized in a highly useful book by Robert Bramson entitled, *Coping with Difficult People*.^[5] This can be properly termed a "how-to-do-it" manual that utilizes metaphors to explain difficult employee behavior. Bramson utilizes such colorful metaphors as "Sherman tanks, snipers, and exploders" to describe hostile, aggressive behavior. Conversely, "clams" are persons who use unresponsiveness as a defensive routine. By explaining behavior in these colorful terms, the author paints graphic pictures in one's mind of things people can relate to, such as tanks, bombs, snipers, and clams. Then, by giving relatively simple quick fixes to these problems, managers acquire practical solutions for dealing with difficult people, without having to know very much about human relations theory or behavioral psychology.

Another aspect of PT involves the fact that it often incorporates several media to relay its messages. A how-to-do-it book may utilize principles, metaphors, acronyms, stories, and anecdotes. Several contemporary examples of this genre of literature include:

- *Reengineering the Corporation*.^[6]
- *Reinventing Government*.^[7]
- *Principled Centered Leadership*.^[8]

The list could obviously go on at length given the demand from managers for practical, useful, organizational advice. Again, the goal is to relate things to managers in ways they can comprehend and apply to specific situations and needs.

STORIES AND ANECDOTES

A final example of a PT medium relates to stories and anecdotes. Everyone likes a good story, including

managers. Stories have a way of grabbing one's attention toward an interesting topic that formal theory cannot match. Stories teach lessons, and one can then take these lessons and apply them to one's own situation. An example of a PT story is the "Abilene Paradox" developed by Jerry Harvey.^[9]

In Harvey's account, his family was relaxing on a Sunday afternoon in a small Texas town wondering what to do. It was hot, approaching 104 degrees. Someone mentioned going to Abilene. Now, several did not really want to go, but felt other family members wanted to. In actuality, no one really wanted to go, but once the idea was set in motion, all agreed to go, thinking this was the group consensus. After having a miserable time on the drive home, everyone confessed they did not really want to go in the first place but went along thinking this was what everyone else wanted to do.

The lesson of the story is this. We often engage in unhealthy agreement. We may not wish to pursue a course of action but go along, thinking this is what our coworkers want to do. They may go along thinking this is what we want. Hence, there is a "crises of agreement." The Abilene paradox is just one example of stories that teach important practical lessons. Many of the how-to-do-it books utilize this technique as well. Extremely popular management books such as *In Search of Excellence*^[10] and *Reinventing Government*^[7] are mainly comprised of short stories and anecdotes about management techniques and practices that have proven successful. These books touch on theory but rely on stories and anecdotes as their primary teaching vehicle.

So, given the value of PT for practicing managers, what could be defective with this strategy, what are its limits and potential drawbacks? We now address this issue.

LIMITS AND PITFALLS OF PRACTICAL ORGANIZATION THEORY

The saying goes, "there is a cloud with every silver lining." So it is with PT. Because the utility of PT often stems from its simplification of complex theory, it may not effectively deal with the many nuances of organizational behavior required to fully understand some issue or problem. Take, for example, span of control and unity of command. By suggesting that span of control be limited to six subordinates, the question quickly arises as to whether this advice is useful for many special situations. If tasks are simple or if employees have considerable knowledge about their work, they may need little direct supervision. Here, a much broader span of control may be effective.

Similarly, some projects may necessitate employees working together from different divisions, thus warranting multiple project supervisors. Broad principles have difficulty with nuance and tend to oversimplify.

Noted scholar Herbert Simon points out another pitfall of practical theory, namely, it frequently takes on a "proverbial" quality that leads to internal contradictions between principles.^[11] For example, unity of command if followed to the letter, denies the possibility of interunit cooperation for problem solving and creates difficulties with span of control. By striving to maximize one principle, this can lead to contradictions in other principles. The platitudes, appealing to common sense, may tell us to do different things simultaneously. Hence, which should we follow?

Practical organization theory may also oversimplify and convey a false sense of knowledge to managers that they then act upon. This could lead to serious problems. For example, some managers may learn how to handle certain problems in "one minute." Others may feel they now have a handle on how to work with certain kinds of difficult employees who display hostile aggressive behavior. Yet, the range of knowledge these condensed and simplified theories provide may be insufficient for dealing with unique issues, special cases, or deeper problems. Managers, in this sense, may overestimate their ability to handle situations, thus leading to worse outcomes.

CONCLUSION

As with any type of knowledge or technique, trade-offs exist. Practical theory provides many useful and easy to comprehend approaches to problem solving that are attractive to public managers. These techniques, principles, and simplified theories provide quick and ready answers to a large bevy of issues confronting managers on a day-to-day basis. They are, therefore, attractive, for they seem to provide a magic bullet, a quick fix, or a clear answer.

Yet, as we have also suggested, they may incur some risks. Developing skill for telling the difference between useful practical theory and faddish hoopla requires practice and experience. Fundamentally, managers must learn to balance the benefits against the potential losses PT may involve. Perhaps this becomes the key issue, because as long as we have managers who continually face streams of organizational issues on a day-to-day basis, we will have practical theory. Practical theory provides managers with an answer to the question "what works." Given the stress and constraints placed on the time of most public managers, such answers become "manna" from heaven.

REFERENCES

1. Peter, L.; Hull, R. *The Peter Principle*; Morrow: New York, 1969.
2. Gulick, L.; Urwick, L. *Papers on the Science of Administration*; Institute of Public Administration; New York, 1937.
3. Golembiewski, R.; Kiepper, A. *High Performance and Human Costs*; Praeger: New York, 1988; pp. 25 and 43.
4. Blanchard, K.; Johnson, S. *The One Minute Manager*; Berkley Books: New York, 1981.
5. Bramson, R. *Coping with Difficult People*; Ballantine Books: New York, 1981.
6. Hammer, M.; Champy, J. *Reengineering the Corporation*; Harper; New York, 1992.
7. Osborne, D.; Gaebler, T. *Reinventing Government*; Addison-Wesley; Reading, MA, 1992.
8. Covey, S.R. *Principled Centered Leadership*; Simon and Schuster: New York, 1990.
9. Harvey, J. Managing agreement in organizations: The abilene paradox. *Organ. Dyn.* **1974**, 3, 63–80. (Summer).
10. Peters, T.; Waterman, R. *Search of Excellence*; Harper and Row; New York, 1982.
11. Simon, H. The proverbs of administration. *Public Adm. Rev.* **1946**, 6, 53–67.

Pragmatism and Public Administration

Keith F. Snider

Naval Postgraduate School, Monterey, California, U.S.A.

INTRODUCTION

Pragmatism has two meanings; first, it is a turn-of-the-twentieth century movement in American philosophy, and second, it is an attitude of expediency and practicality. The contributions to and influences on public administration by the former are tenuous and ill-defined, considering the field's generally atheoretical orientation. The latter, on the other hand, is clearly evident in public administration's perennial concerns with the practical day-to-day concerns of governance.

PHILOSOPHICAL PRAGMATISM'S RISE AND FALL

Pragmatism arose in late nineteenth century philosophy as a way of thinking to mediate between two diverging trends. One trend—Darwinian, materialistic, scientific, probabilistic, and secular—portrayed humans as mere biological elements in apparently random universal processes. The other—idealistic, rationalistic, romantic, and religious—gave humans a central place in the universal scheme due to their powers of reason and their moral and religious sensibilities that infuse life with purpose and meaning. The first trend was criticized for ignoring moral considerations; the second was criticized for neglecting objective scientific facts. Pragmatism attempted to reconcile these two trends with the view that ideas dealing with the realm of values must have some practical consequence if they are to have any meaning.

The Early Pragmatists - Peirce, James, and Dewey

Charles S. Peirce (pronounced *purse*), generally recognized as pragmatism's founder, argued that ideas have clear meanings only in their operation—that is, in terms of their actions, effects, or consequences. For example, a rock is hard if it scratches something or soft if something scratches it. Apart from what is experienced as the practical consequences (the scratches) of the testing, the words “hard” and “soft” have no meaning. According to Peirce's “pragmatic maxim,” published in 1878, “Consider what effects, which might conceivably have practical bearings, we conceive the

object of our conception to have. Then, our conception of these effects is the whole of our conception of the object.”^[1]

Belief, according to Peirce, links thought and action. It is a settled state, arrived at on the basis of one's experiences of the effects or consequences of ideas. Experience is, however, always situational and contextual—any new or different sensing leads to the unsettled state of doubt. Inquiry or thought is the struggle to overcome doubt, thereby fixing belief at least provisionally. Peirce held the method of scientific inquiry to be the superior means for fixing belief. Because conceptions of things are based on sense experiences of their consequences, which should be observable and verifiable, belief may be fixed in a public sense to the extent that all observers can sense and agree upon the meaning of the consequences of ideas. Any concepts of knowledge, truth, or reality are thus socially constituted and always provisional.

Peirce's ideas received little notice, but they inspired both James and Dewey, whose work elevated pragmatism to prominence in the early twentieth century. William James, who first used the term “pragmatism” in 1898, appropriated Peirce's method for establishing meaning and extended it to a theory of truth. According to James, the test of any idea or belief is in the difference it makes in life, or as he put it, in its “cash-value.” For an idea to be true means that it must “work.”^[2]

John Dewey, whose pragmatism is also called instrumentalism or experimentalism, believed that the mind mediates between one's biological being and the various problematic situations it encounters. Thought is inspired by doubt or perplexity experienced in a particular situation. This unsettled state leads to deliberation and elaboration of competing hypotheses about how to resolve the situation. Thinking is not a quest for truth but rather an attempt to achieve adjustment with one's environment. Thought has an instrumental character in guiding action, thus ideas are plans of action or hypotheses, not representations of immutable truths or ends.^[3]

Philosophical Pragmatism's Integrative Character

The pragmatism of Peirce, James, and Dewey provided resolutions for philosophy's traditional ends-means,

fact-value, and thought-action dualisms. In place of metaphysical ends, pragmatism grounded value in human experience. People continually rearrange facts—the perceptions of experience—to form new ideas in response to their environment. The value of any idea is found in the extent to which that particular arrangement of facts serves to resolve some problematic situation. Values thus stem from and have meaning only in the facts of experience. Because actions may have value in resolving situations, values (or ends) are also inseparable from actions (or means). Finally, thinking is not isolated from experience but rather it continually reflects on experience to inform action to solve problems. Thought and action (or theory and practice) thus function in an integrated way.

Pragmatism and the Progressive Era

Pragmatism provided a philosophical basis for the progressive movement after the turn of the century. Pragmatism justified progressivism's unwillingness to accept the often appalling social and political conditions of the day as natural states of affairs and, accordingly, its vision of an active and expanded government. Legal, educational, economic, and political institutions that in the past had been used for purposes of power, greed, and oppression began to be viewed as instruments for change, amelioration, and progress.

Pragmatism's Decline

As James's and Dewey's influence spread, pragmatism took multiple and sometimes inconsistent forms depending on the background and interests of new proponents, which led to occasional conflicts and created confusion among observers. The extent of the diversity of approaches by 1908 is illustrated by one philosopher's attempt to clarify pragmatism's various interpretations in a series of essays titled *The Thirteen Pragmatisms*.^[4]

Perhaps more significant in pragmatism's decline was the dilution of its meaning as it spread beyond the discipline of philosophy into the broader public consciousness. James's catchy description of pragmatism as the "cash-value" in ideas had particularly popular appeal. Gradually, pragmatism in common usage came to mean simply doing what works or having an attitude of commonsensical practicality.

In the 1920s and 1930s, realism emerged to replace the Progressive Era's ambitious social optimism. The temper and spirit of the age changed dramatically in light of the horrors of the Great War and in the social and economic crises that followed. This new realism was exemplified in the writings of Reinhold Niebuhr,^[5]

who castigated public philosophers such as Dewey for their views of human and social perfectibility. Pragmatism had been so closely associated with progressivism that the disillusionment surrounding the failure of the movement could not fail to affect the philosophy.

In the 1930s, pragmatism came to be overshadowed by logical positivism. American philosophers, particularly at the University of Chicago, saw logical positivism as the cure for pragmatism's shortcomings, believing the two could be successfully merged. Others disagreed, noting irreconcilable differences.^[6] An especially significant incongruence was logical positivism's strict separation of facts and values, which held that value statements were meaningless because they could not be subject to empirical verification. This was of course contrary to pragmatism's view of facts and values as inseparable. Nonetheless, many were willing to sacrifice pragmatism's integrative character for logical positivism's rigor. This development heightened the general confusion and disillusionment with pragmatism, which faded as a significant force in philosophy for almost the next fifty years.

PRAGMATISM IN PUBLIC ADMINISTRATION

None of the pragmatist philosophers addressed public administration explicitly (though Dewey's *The Public and Its Problems*, published in 1927, is closest to the concerns of the field), and very few writers on public administration prior to the late twentieth century explicitly addressed pragmatism in any significant way.^[7] Pragmatism thus had little obvious or direct influence on the early field and its development. Serious treatment of pragmatism by public administration writers has only begun in the last thirty or so years.

The Founding Period

The usual account of American public administration's founding begins with the reform-minded writings and actions of Progressive Era men such as Frank Goodnow, Charles Beard, Frederick Cleveland, Luther Gulick, and Charles Merriam. This is the same general period when philosophical pragmatism emerged, but there is little evidence that it directly influenced these "early giants." Rather, they focused on the pressing problems of the day, showing little interest in philosophy. Science and its methods offered the best means for addressing and solving these problems, but politics had too often trumped science in the past. Public administrators thus had to be freed from political influences in order to employ their techniques and expertise. The proper orientation of the

emerging profession was toward value-neutral efficiency in administration.

While the founders' emphasis on problem-solving and science may seem at first blush to be coherent with philosophical pragmatism, their goal of value-neutrality actually opposed it. Pragmatism could not tolerate such a split between politics (the realm of ends or values) and administration (the realm of means and facts). That the founders thus reflected some aspects of pragmatism but not others helps explain occasional references to the field as "pragmatic."

The Founding to the Present—Traces of Pragmatism

The founders' initial value-neutrality has been described as leading in later years to an excessive emphasis on technique, rationality, and administrative expertise, which is blamed for public administration's theory-practice gap and its identity crisis. While this trend precluded serious consideration of philosophical pragmatism among most in the field, a few showed interest. Specifically, Mary Parker Follett relied heavily on William James in her essay "Community is a Process,"^[8] and John Gaus spoke approvingly of Dewey's *The Public and Its Problems*.^[9]

Apart from these few exceptions, pragmatism's influence on public administration can be characterized as either indirect or problematic. The managerial trend provides an illustration of pragmatism's indirect influence. With roots in Taylorism, this trend was reinforced by Leonard White's idea that management, not law, should provide public administration's foundation. It continued with organizational theorists such as Elton Mayo and Chester Barnard. Some aspects of the human relations movement seem based in pragmatism, for example, its contingency approaches and the instrumental roles assigned to managers and organizations in leading and caring for workers.^[10] None of the managerial theorists, however, worked directly from a basis in pragmatism. At most, their ideas were attenuated extensions of pragmatic philosophy into the arena of management; at the least, they simply reflected a few pragmatic characteristics.

An illustration of pragmatism's problematic influence on the field is seen in the policy sciences movement. Dewey has been linked as a contributor to this movement, which has origins in the 1920s among University of Chicago political science faculty such as Charles Merriam and Leonard White. This movement replaced the Progressive Era's faith in science and technology with a more constrained concern for cause-and-effect prediction and control through measurements of individual and social behavior. Yet, Dewey actually distanced himself from the movement, objecting to its

neglect of values in its empiricist behavioralism, and criticizing its inquiry as "superficial and comparatively trivial, no matter what its parade of technical skill."^[11]

Herbert Simon provides another illustration in his seminal 1947 work, *Administrative Behavior*. As a student in the Chicago political science department, Simon worked explicitly from logical positivism, which by that time had captured pragmatism, as discussed above. Simon was thus able to claim both James and Dewey as philosophical kinsmen while at the same time espousing a strict fact-value distinction. Pragmatism thus influenced Simon, but it was a corrupted and co-opted version from that of James and Dewey.

Other writers after World War II called into question public administration's separation from value-laden politics, as did Dwight Waldo in his 1948 book *The Administrative State*. Others, such as the contributors to the Inter-University Case Program,^[12] attempted to capture the rich complexity—including the politics—of administrative reality. While this turn agreed with pragmatism, none of these writers relied on it explicitly. Apart from these works, little else occurred with regard to pragmatism until the late twentieth century.

THE PRESENT REVIVAL OF PRAGMATISM

Since roughly 1970, two trends have contributed to a re-examination of pragmatism in public administration. The first is dissatisfaction among some scholars with the field's condition and a concomitant desire to rethink its foundations. This has led to historical analyses to discover or recover public administration's roots in pragmatism, as well as to reconstruction projects to reestablish the field in its ideas. The second is the influence of Richard Rorty's neopragmatism, which has contributed to a revival of interest in pragmatic thought.

The Road Not Taken—Jane Addams

The Progressive Era settlement movement, which had explicit philosophical connections to pragmatism, has recently been proposed as meriting a legitimate place in the history of public administration's founding.^[13] Jane Addams of Chicago's Hull House knew John Dewey from his frequent visits and relied explicitly on his and James's ideas in her writings. She saw the settlements as experimental approaches to some of the pressing social problems of the day. She also saw in them the potential for a more effective manifestation of democracy: participation by citizens in communities of inquiry. Addams thus embraced science and its methods, but, in contrast with public administration's

founders, she did not attempt to divorce science from questions of politics or value. This view of democracy as a participatory community of inquiry was also Dewey's central theme in *The Public and Its Problems*.

It of course remains to be seen whether this alternative account of public administration's founding in the explicit pragmatism of the settlement movement is accepted and, if so, what effect that acceptance may have on public administration's contemporary identity.

Richard Rorty and Neopragmatism

Rorty began to develop his particular form of pragmatism in the late 1970s, relying a great deal on the early pragmatist philosophers but also differing from them in important ways. Rorty's main departure from Peirce, James, and Dewey occurred in his arguments that they were insufficiently anti-foundational in their reliance on experience as a ground for meaning. That is, traditional pragmatism held that, because meanings of concepts are contingent, they must be grounded in experience. Rorty argued that this formulation made experience foundational when, in actuality, experience is simply a linguistic construction. Rorty has stated that it is better to ground pragmatism in language; hence his neopragmatism is sometimes called linguistic pragmatism. This turn sweeps away any and all foundations because language is a social construct. Progress thus obtains not from the scientific method, but rather through the creation of more appropriate vocabularies, languages, and ways of speaking.^[14]

His radical views and willingness to employ them in attacks on institutions and cherished beliefs have made Rorty a highly controversial writer. His radicalism notwithstanding, Rorty has had some influence in public administration, as some contemporary scholars see his linguistic pragmatism as holding promise as a central guiding mode of thought for the future.^[15]

CONCLUSIONS

Though attention to pragmatism among theorists and historians of public administration has increased in recent years, it is still a relatively minor topic of interest for most in the field. Public administration today still reflects, in large measure, the founders' original concern for solving problems. Much of the field, therefore, sees itself as pragmatic without recognizing either its

potential connection to or departure from the philosophical pragmatism of Peirce, James, and Dewey.

REFERENCES

1. Peirce, C.S. How to make our ideas clear. In *Pragmatism: A Reader*; Menand, L., Ed.; Vintage: New York, 1997; 26–48.
2. James, W. *Pragmatism: A New Name for Some Old Ways of Thinking*; Longmans, Green: New York, 1907; 58–73.
3. Dewey, J. *The Quest for Certainty*; Minton, Balch: New York, 1929.
4. Lovejoy, A.O. *The Thirteen Pragmatisms and Other Essays*; The Johns Hopkins Press: Baltimore, 1908.
5. Niebuhr, R. *Moral Man and Immoral Society*; Charles Scribner's Sons: New York, 1932.
6. Wilson, D.J. Fertile ground: pragmatism, positivism, and the linguistic turn. In *Pragmatism: From Progressivism to Postmodernism*; Hollinger, R., Depew, D., Eds.; Praeger: Westport, CT, 1995; 122–141.
7. Snider, K. Rethinking public administration's roots in pragmatism: the case of Charles A. Beard. *Admin. and Soc.* **2000**, *30* (2), 123–145.
8. Follett, M. Community is a process. *Philos. Rev.* **1919**, *28* (November), 576–588.
9. Gaus, J.M. *Reflections on Public Admini*; University of Alabama Press: University, AL, 1947.
10. Stever, J. Pragmatism. In *International Encyclopedia of Public Policy and Administration*; Shafritz, J., Ed.; Westview: Boulder, 1998; 1734–1736.
11. Depew, D. Introduction to Part II: pragmatism, positivism, and the linguistic turn. In *Pragmatism: From Progressivism to Postmodernism*; Hollinger, R., Depew, D., Eds.; Praeger: Westport, CT, 1995; 109–121.
12. Stein, H. *Public Administration and Policy Development: A Case Book*; Harcourt, Brace: New York, 1952.
13. Shields, P. The community of inquiry: classical pragmatism and public administration. *Admin. Soc.* **2003**, *35* (5), 510–525.
14. Rorty, R. *Philosophy and Social Hope*; Penguin: New York, 2000.
15. A recent series of articles in *Administration & Society* focused on Rorty's potential contributions to public administration. A key question emerging from this interchange was, "Which is better for the field: Rorty's language-based pragmatism or Dewey's experience-based pragmatism?" For the concluding articles in the series, see Miller, H. Residues of foundationalism in classical pragmatism. *Admin. Soc.* **2005**, *37* (3), 360–374 and Shields, P. Classical pragmatism does not need an upgrade: lessons for public administration. *Admin. Soc.* **2005**, *37* (4), 504–518.

Principal-Agent Perspective

Arie Halachmi

*Institute of Government, Tennessee State University, Nashville, Tennessee,
U.S.A. and Zhongshan University, Guangzhou, P.R. China*

INTRODUCTION

According to Barney and Hesterly, agency theory seeks to understand the causes and consequences of the conflicting goals associated with an organization.^[1] To them, agency theory is part of the literature of organization economics. Neelen suggested that the principal-agent model, neo-institutional economics on the whole, and neoclassical economics are, in fact, large families of diverse theories.^[2] These theories are related by shared analytical foundations more than they are unified theories. According to Moe, these shared analytical foundations include a focus on the individual as the unit of analysis; the assumption of rational utility maximization behavior; a concern for efficiency, optimization, and equilibrium; and a preference for mathematical modeling.^[3] As it originally developed, agency theory research focused on the relationship between managers and stockholders,^[4] drawing heavily from the property rights literature^[5] and, to a lesser extent, from transaction cost theory.^[1] The phrase “principal-agent perspective” is used here as a catch-all phrase because a complete phenomenological and epistemological discussion of the various approaches to and writings on principal-agent relationships and their antecedents in the literature on the separation of ownership and control^[4,6] are beyond the scope of this discussion.

DEFINING AGENCY RELATIONSHIPS

The principal-agent perspective views the interaction between two parties as a situation in which two actors (i.e., individuals, groups, or institutions) agree to cooperate with each other in a contractual relationship that requires the principal to pay the agent for activities the agent makes on behalf of the principal. The relationship between the two actors is characterized by the presence of agency-specific conditions:

- Conflict of interest or, more technically, negative externalities. The decisions of the agent affect both

the agent’s own welfare and that of the principal (more on that interdependency below).

- No choice when the principal cannot accomplish something on their own. The agency relationship results when the “principal” must secure the services of the “agent” through “delegation” to complete a task.
- Asymmetric information. The principal is faced with a situation of imperfect monitoring because it is impossible to observe all the actions and decisions of the agent or infer them by observing the outcomes of the agent’s decisions. Hence, the agent has some opportunity for discretionary behavior.
- The relationship between principal and agent is based on a contract.
- Actors maximize their utility functions subject to institutional constraints.
- Because principals and agents have different utility functions, activities that are mostly in the interest of one actor have adverse effect on the welfare of the other actor.
- Transaction costs are positive.

Neelen concluded that with this characterization of the agency relationship, the principal always faces questions of control in the context of information asymmetry and conflict of interest; thus, it is impossible for the principal to ensure, at zero cost, that the agent will make optimal decisions from the principal’s point of view.^[2] Earlier, Jensen and Meckling pointed out that in most agency relationships the principal and the agent will incur positive costs.^[4]

The principal incurs the cost of having to monitor the agent’s behavior. However, as noted by Arrow, some of the agent’s activities are hidden from the principal or are too costly to observe.^[7] Arrow called this “moral hazard” and pointed out that this is one source of agency problems. The other source, “adverse selection,” involves the hidden information that is available to one but not the other party.

The agent incurs the cost of bonding (i.e., of assuring the principal that certain actions inconsistent with the principal’s interest will not be taken). Jensen and

Meckling went on to suggest that the dollar equivalent of the reduction in welfare experienced by the principal due to a divergence (i.e., between the agent's decisions and those that would maximize the principal's welfare) is also a cost of the agency relationship.^[7] Jensen and Meckling referred to this cost as the "residual cost."^[4] Accordingly, they defined "agency cost" as the sum of:

- The monitoring expenditures by the principal
- The bonding expenditures by the agent
- The residual cost

As noted by Douma and Schreuder, agency relations can be found both within organizations (e.g., manager and subordinate) and between organizations (e.g., licensing and franchising).^[6] Thus, understanding the environmental variables that can influence agency cost should be of interest to practitioners and researchers alike. The problem is that most writers do not offer a conceptual framework for exploring the complex relationships among the principal, the agent, and the environmental context of their relationship.

Boorsma and Halachmi pointed out that, even though agency relations involve a hierarchy, there is a tendency in the literature to denote every dependency relation as an agency relation, including every hierarchical relation between a superior and a subordinate.^[8]

Yet, such hierarchical relations can be discussed within the time-honored Weberian framework. Second, Boorsma and Halachmi pointed out that market situations are frequently characterized by uncertainty or information asymmetry.^[8] Table 1 provides a possible summary of the different characteristics of contract and agency relationship.

A promising approach to the study of principal-agent relations was described by van der Zaal. Van der Zaal began by making two important points that help to explain the attractiveness of the principal-agent perspective for studying and understanding the challenge that can exist when services are contracted out.

First, van der Zaal claimed that principal-agent taxonomy from agency theory appears to suit the required perspective for interorganizational strategic management issues. In addition, he saw this taxonomy as a promising candidate for the development of an integrative framework in which connections can be made among environmental dimensions, such as "uncertainty" and "dependency," and intraorganizational dimensions, such as organizational resources, strategic assets, core capabilities, and core competencies.^[9] Such an inside/out approach would be complementary to strategic management, in which competitive analysis starts within the boundaries of the organization but with an outside/in perspective.

Table 1 Basic transactional patterns of contract and agency relationships

Division of labor	Contract	Agency
Formal aspect	Market	Hierarchy
Perception of actors	Isolated	Integrated
Procedural aspect of planning	Self-determination	Delegation
	Specification	Representation
		Discretion
	Discreteness	Interdependence effort
	Performance orientation	Orientation
Legal structure	Mutual consent on two individual decisions to perform	Unilateral authorization to act on behalf of principal
Necessary rules	None (fully specified contract is self-enforcing; pacta sunt servanda is not necessary but an expedient rule)	Sanctions for breach of loyalty; sanctions for non-performance of promised incentives
Revelation of individual preferences	Does not matter	Central feature
Real world paradigm(s)	Exchange of goods and services	Manager, sales agent, guardian, sharing arrangements
Basis of legitimacy	Reciprocity (mutual consent of parties)	Power delegation by 1) consent, 2) voting, 3) statute
Entitlement structure	Exchange of property rights	Horizontal split of property rights; paradigm: decision control+residual ownership/decision management

Source: From Ref.^[11].

Table 2 van der Zaal framework

	Principal	Agent
Environmental uncertainty—an information perspective	1 Uncertainty about (the relation with) the agent	2 Uncertainty about (the relation with) the principal
		3 Uncertainty about the environment surrounding the principal-agent relationship
Environmental dependency—a resource perspective	4 Dependency on (the relation with) the agent	5 Dependency on (the relation with) the principal
		6 Dependency on the environment surrounding the principal-agent relationship

Van der Zaal's second claim is that agency cost reduction is not enough to explain the cooperation between the principal and the agent.^[9] The other reason for the cooperation is the incentive to create value. Van der Zaal called this concept "agency benefit" and made two propositions: 1) that by integration of the "agency benefit" concept into the principal-agent framework, it may be possible to develop a well-balanced benefit-cost framework for analyzing cooperative interorganizational relations from a principal-agent perspective, and 2) that the principal-agent taxonomy deals with three levels of analysis.^[9] The first level is the focal organization. The partner in the relationship is part of the environment and represents the second level of analysis. The environment around the principal-agent relationship forms the third level of analysis.^[9]

The conceptual framework offered by van der Zaal^[9] is summarized in Table 2. It requires us to consider the relationships among the principal, the agent, and the environment, first by considering the uncertainty of the environment (i.e., by using an information perspective as illustrated by cells 1, 2, and 3), and second, by using a resource perspective (i.e., the dependency relations as illustrated by cells 4, 5, and 6). This framework may be useful for exploring many issues and the contingencies of various administrative innovations. For example, Halachmi suggested that a principal-agent perspective is a promising framework for studying a "make" or "buy" decision such as the federal experiment with Franchise Funds [i.e., the effort to encourage federal entities to secure (i.e., buy) administrative support services from other federal entities rather than provide (i.e., make/produce) them in-house].^[10] For that matter, the

principal-agent perspective may prove useful for studying any decision about contracting out, partnership, or cooperation.

REFERENCES

1. Barney, J.B.; Hesterly, W. Organizational economics: understanding the relationship between organizations and economic analysis. In *Handbook of Organization Studies*; Glegg, S.R., Hardy, C., Nord, W.R., Eds.; Sage Publications: Thousand Oaks, CA, 1996; 115–147.
2. Neelen, G.H.J.M. *Principal Agent Relations in Non-Profit Organizations*; University of Twente, Faculteit Bestuurskunde: Enschede, The Netherlands, 1993; pp. 63, 65.
3. Moe, T.M. The new economics of organizations. *J. Polit. Sci.* **1984**, *28*, 739–777.
4. Jensen, M.C.; Meckling, W.H. Theory of the firm: managerial behavior, agency cost and ownership structure. *J. Financ. Econ.* **1976**, *3*, 305–360.
5. Alchain, A.A.; Demsetz, H. Production, information cost and economic organization. *Am. Econ. Rev.* **1972**, *62*, 777–795.
6. Douma, S.; Schreuder, H. *Economic Approaches to Organizations*; Prentice-Hall: Englewood Cliffs, NJ, 1991; 77–78.
7. Arrow, K. The economics of agency. In *Principals and Agents: The Structure of American Business*; Harvard Business School Press: Boston, 1985.
8. Boorsma, P.B.; Halachmi, A. Introduction: warp and weft. In *Inter and Intra Government Arrangements for Productivity: An Agency Approach*; Halachmi, A., Boorsma, P.B., Eds.; Kluwer Academic Pub.: Boston, 1998; xxiii–xxxii.
9. Van der Zaal, G.A.W. *On Cooperative Organizational Relations: An Integrative Principal Agent Perspective*;

Management Report Series, School of Management, Erasmus University: Rotterdam, 1993; Vol. 164; pp. 1, 8, 9, 11.

10. Boorsma, P.; Halachmi, A. Franchising in government: an idea in search of a theory. In *Inter and Intra Government Arrangements*; Halachmi, A., Boorsma, P., Eds.; Kluwer Academic Pub.: Boston, 1998; 45–58.
11. Schanze, E. Contract, agency and the delegation of decision-making. In *Agency Theory, Information and Incentives*; Bamberg, G., Spemann, K., Eds.; Springer: Heidelberg, 1987; 470.

BIBLIOGRAPHY

1. Barney, J.B.; Hesterly, W. Organizational Economics: Understanding the Relationship Between Organizations and Economic Analysis. In *Handbook of Organizational Studies*; Clegg, S.R., Hardy, C., Nord, W.R., Eds.; Sage Pub.: Thousand Oaks, CA, 1996; Ch. 14.
2. Halachmi, A.; Boorsma, P.B. *Inter and Intra Government Arrangements for Productivity: An Agency Approach*; Halachmi, A., Boorsma, P.B., Eds.; Kluwer Academic Pub.: Boston, 1998.

Privatization

Marc Holzer

Croton-on-Hudson, New York, U.S.A.

Hwang-Sun Kang

*Department of Urban Management, Seoul Development Institute,
Seoul, Republic of Korea*

INTRODUCTION

The privatization movement in the public sector was accelerated by the New Conservatism under the Reagan and Thatcher administrations. Basically, new conservatives value economic and political liberalism that have emphasized free competition of market system. Under liberalism, small government and small tax are preferred. According to them, welfare state imposes so much tax and regulation on free-market mechanism as to distort the free flow of capital. Thus welfare state (big government) cannot but be faced with economic stagflation under which economic depression and inflation simultaneously occur. In addition to this basic belief of liberalism, new conservatives do not neglect the weak points of market mechanism. Thus they emphasize not only free-market system but also the inevitability of government intervention. In this context, new conservatives argue for “small but strong” government rather than simply “minimal government.”

The privatization movement began from this reflection on the problem of welfare state. The rationale for privatization is government failure that public choice has criticized. According to public choice theorists, decision making in government is not made based on the principle of rationality; rather, most decisions are made by political bargain. In the political bargain process, many interest groups, politicians, and business groups are involved. And public policies produced by the process are implemented by big government. Public choice theorists consider government bureaucrats as self-interest seekers just like other business groups. As business groups pursue profit-maximization, government bureaucrats, according to public choice theorists, do “budget maximization.”^[1,2] As a result, interest groups, politicians, business groups, and government bureaucrats establish a “budget-spending coalition.”^[3] Public services provided by the budget-spending coalition cannot but be inefficient.

DEFINITION OF PRIVATIZATION

Strictly speaking, privatization means that government sells its property rights to state-owned institutions to

the private sector. For example, currently government-owned companies are being sold to private companies. Thus many services that were once provided by government are now being provided by private companies. Even international companies can buy other countries’ public companies.

In a broad sense, transition of government function can be included in privatization. For instance, government can make a contract with private companies in order to enable the companies to provide public services in lieu of government.

FORMS OF PRIVATIZATION

There are several forms of privatization. Here, four representative forms of privatization are discussed based on the Savas’ classification: contracting out, voucher, franchising, and sale.^[4]

Contracting Out

Government can delegate authority of a public service to a company via a competitive bid during a limited period. Through contracting out, government can provide better services with less cost because of competitive bids between private companies. Some state governments are using contracting-out strategy in order to stimulate public employees. Public employees can see that their jobs may be privatized unless performances of their organizations are satisfactory. When a government contracts out a service with a company, it should pay much attention to whether the company can sustain good quality of service after the service is contracted out. And government should check whether the company persistently holds to its responsibility to the citizens. Most of all, government should provide fair conditions for many private companies that are participating in a competitive bid.

Possible services that can be contracted out include public car maintenance, public hospital services, garbage collection, security services, and management of public facilities such as library, cafeteria, and other

professional facilities. According to Savas' report, the average American city contracts out 23% of its 64 common municipal services to the private sector. The average American state contracts out 14% of its activities.^[4]

Voucher

There are many limitations to a government's ability to take care of a citizen's specific demand. For instance, recipients of welfare services need very specific and various cares. By giving a voucher, government allows citizens who need a specific service to choose the service that meets their needs.

In this system, government should pay much attention to whether private companies can fairly compete with each other in the voucher market. Possible services for the voucher method include food, housing, education, health care, transportation, and other social services.

Franchising

Through franchising, government can award a company specific authority to provide a service for a limited period. A company, for example, can build a public parking lot and manage it in a public land. And the company should pay government a fee. Usually, building infrastructures is a great burden to governments. Through franchising, government can take advantage of private capital and provide better service for citizens with no great financial pressure.

Sale

Government can sell a public facility to a company. Of late, some governments have sold public-owned

companies to private companies. In the globalization age, some international companies can purchase other countries' public companies.

In addition to those four methods, there are other methods of privatizations. Table 1 shows various methods of privatization based on Savas' study.^[4]

EFFECTS OF PRIVATIZATION

In general, there are two advantages in the privatization of public services. First, privatization contributes to enhancing efficiency of government management. By transferring a portion of government function that the private sector can undertake, government is able to reduce its total size. By making government small, the private sector can play a bigger role in the total national economy. Thus privatization is known to relieve financial deficit in government. In fact, the privatization movement in America is closely related to the financial crisis of the public sector in the 1970s.

Second, privatization can give multiple choices to citizens choosing public services. Government loses its monopolistic authority in providing public services and has to compete with the private sector.

WEAKNESS OF PRIVATIZATION

The ideology of privatization is efficiency of government management. It is true that privatization has contributed to increasing the total efficiency (economic efficiency) of government management. However, it should be noted that privatization should be implemented under fair competition between private

Table 1 Institutional arrangements for providing public services

Service arrangement	Arranger	Producer	Who pays?
Government service	Government	Government	Government
Government vending	Consumer	Government	Consumer
Intergovernmental agreements	Government ^[1,2]	Government ^[3]	Government ^[1,2]
Contracts	Government	Private sector	Government
Franchises (exclusive)	Government	Private sector	Consumer
Franchise (multiple)	Government and consumer	Private sector	Government and consumer
Grants	Government and consumer	Private sector	Government and consumer
Vouchers	Consumer	Private sector	Government and consumer
Free market	Consumer	Private sector	Consumer
Voluntary service	Voluntary association	Voluntary association	N.A.
Voluntary service with contract	Voluntary association	Private sector	Voluntary association
Self-service	Consumer	Consumer	N.A.

Source: From Ref.^[4].

companies. In the same context, a certain class of people should not be excluded from the benefits of privatization. This is the issue of social equity.

Although the privatization of government function is widely preferred, it does not necessarily mean that minimal state is the best form of modern government. In other words, privatization is not equated with "small government." Even in privatization periods, citizens still want government to do more work for them. Thus just as big government brings some problems, a big private sector will also bring some more problems. Stable development of modern societies would be possible in the status of equilibrium between the private and the public sectors.^[5,6]

CRITERIA OF PRIVATIZATION

Before a public service is privatized, there are several points to be considered.

First, what is the original goal of the service that is going to be privatized? All public services have their targeted goals and recipients. If the services were provided by the private sector, would those original goals be maintained?

Second, should government really be responsible for the service? As it is well discussed by many scholars, the issue of responsibility is a very important point to be considered prior to privatization of public services.

Third, is the privatization policy of a country compatible with the societal needs and national goals of the country? It should be noted that privatization movement followed reconsideration of a welfare state in America and Britain. In other words, those developed countries already have a well-established welfare system. On the contrary, citizens in other

developing and underdeveloped countries still need more extensive public services from government.

CONCLUSION

The core of privatization strategy is to imbue the public sector with competition and innovation mind. Transfer of property rights and government function from the public sector to the private sector has been a prevalent effort to institutionalize competition and innovation mind in the public sector.

Fundamentally, management strategy of public organizations should not be the same as that of private organizations because those two organizations have different organizational goals: profit maximization and pursuit of public interest. Considering the weaknesses of privatization, it would be more desirable to find more strategies to sustain competition and innovation mind along with the privatization forms mentioned above.

REFERENCES

1. Niskanen, W. *Bureaucracy and Representative Government*; Adline Publishing Company: Illinois, 1971.
2. Downs, A. *An Economics Theory of Democracy*; Harper & Row: New York, 1971.
3. Zahariadis, N. *Markets, States, and Public Policy; Privatization in Britain and France*; University of Michigan Press: Michigan, 1997.
4. Savas, E. *Privatization and Public – Private Partnership*; Chatham House Publishers: New York, 2000.
5. Mintzberg, H. Managing government management. *Harvard Bus. Rev.* **1996**, 74 (3), 75–83.
6. Nye, J. *Why People Don't Trust Government*; Harvard University Press: Massachusetts, 1998.

Privatization and Economic Development

Bill Luker

Department of Economics, University of North Texas, Denton, Texas, U.S.A.

INTRODUCTION

“Economic development” and “globalization” are the two catch phrases for this era in capitalist history. However, globalization is a relatively new term, and until recently its discussion domain has been limited to circles of professional economists, policy theorists, and senior public and private executives at national and international levels. However, among economic and policy practitioners at other levels of government (e.g., province, state, or local level), concerns about economic development arose long before the globalizing forces that now make development in their jurisdictions more of an imperative than ever.

Descriptions of the specific ends of economic development have been as varied as the places in which developmental theories and practices have been applied. Yet, in recent economic history, one general end or goal stands out—to raise or qualitatively improve the living standards of people in a particular place or nation, relative to some set of preferred living standards in Europe, the United States, or highly affluent places in those countries. Although various measures have been proposed to gauge the results of efforts aimed at these qualitative improvements, those most commonly used by economists and policy makers have been quantitative (e.g., changes in per capita income).

Per capita income change, either based on national monetary standards or on a global “purchasing power equivalent” standard, is a convenient metric for measuring progress toward developmental ends. Despite its convenience, and because it accounts only for the economic dimension of higher living standards or levels of human functioning, it has come under increasing attack by developmental theorists in recent years.^a But it is much less controversial than the *means* by which economic development itself should be achieved. It is here, since the 1980s, that the real battle lines—theoretical, practical, political, and ideological—have been drawn in a debate about whether government

or the private sector should lead in realizing economic development goals.

Much of this debate stems from the misleading notion that large-scale developmental goals have, and generally can be, attained without government involvement. This has not been the case historically, least of all in the United States, where the public sector led the planning, and provided the fiscal wherewithal, for development projects in mass electrification, telecommunications, and agriculture (through dams, irrigation, and farm supports of various kinds), and for rail, auto, truck, and air travel that helped to create the most affluent, geographically unified capitalist market on the planet. In most places around the world, whether highly articulated capitalist economies or those where “basic needs strategies” are required, it is still the case that development projects to provide infrastructure—sanitation, roads, and other non-divisible (i.e., public), goods—are planned, financed, and carried out by public or non-profit authorities.

Nevertheless, there are clear differences in practice, and in theory, between publicly and privately led development efforts. “Privatization” is the term that encompasses the latter.

PRIVATIZATION IN PRACTICE

Privatization manifests itself in economic development as the privatization of economic development. We can see it as a generalized process that has occurred more or less everywhere since the 1970s. First, the privatization of development shifts the choice of a measure for the effects of development initiatives from broadly defined improvements in living standards—even when these are narrowly measured, again, as increases in per capita income—to increases in the absolute number or growth rate of private sector jobs. The shift in metric is not merely a move from one convenient measure to another. It is emblematic of a change both in the ends and the means of development.

Jeffrey Sachs, the noted development economist, inadvertently revealed the character of this change when he stated that in parts of the developing and transitioning world (the post-Communist states), any wage-paying job, no matter what the wage, is better than no job at all. He was correct because the goal

^aSee Ref.^[1], particularly p. 10 and Chapter 2, pp. 35–53 (“The Ends and Means of Development”) for an adumbration of his assertion that the goal of development is *more freedom* for the individuals who make up a given society, and not just more economic well-being.

of the development process, even for the most economically backward or disrupted regions of the world, has shifted from more general ones such as “improving living standards” or “providing basic needs” to those in which most if not all world economies are to be ushered into a national, and ultimately international, wage, production, and consumption nexus (i.e., the cash nexus). Private employment at a cash wage is a necessary prerequisite for this to occur.

The privatization of economic development effects just such a shift in ends, perhaps because of, or in combination with, a shift that occurs in the means by which new jobs are to be created. Most often, and even when there are indigenous firms, technologies, or other physical or human resources present (Latin America is a case in point) that may allow development along an “endogenous path,”^[2] privatized development sees new job creation as the exclusive result of “foreign” private investment decisions (i.e., by national or multinational corporations from outside the nation, region, state, city, or place). In almost every case, these public jurisdictions now use public (tax) revenue to subsidize this employment-generating private investment.

Subsidies come in many forms. In highly developed capitalist economies such as those in the United States and Europe, cities, states, and regions compete with each other to induce firms to invest in their jurisdictions with property or income tax holidays (often lasting for years); outright grants of land, cash, developed facilities, or infrastructure (roads and water); state-provided assistance with screening, hiring, and training new employees, or direct wage subsidies.

In smaller or poorer places or nations, with few resources to induce a firm to make a favorable location decision, subsidies come principally in the form of cheap and pliable labor. Government plays its part in creating new employment by ensuring low or no minimum wages, no unions, minimal workplace safety and environmental standards, and minimal or no controls on capital mobility or the repatriation of profits from low-wage enterprises to the corporation making the investment.

As one local mayor in the poverty- and depression-stricken state of Tennessee put it in the 1930s, such is the process of “buying payroll.”^b Indeed, this aspect of the privatization of economic development on a global scale has its modern roots in poor communities of the South, West, and New England regions of the United States. Beginning in the early nineteenth century, they attempted to solve local poverty and unemployment problems by subsidizing the location of new manufacturing firms to their communities. Much later, as U.S. government aid to states and cities

declined throughout the 1970s and 1980s, local governments became more concerned than ever with creating new streams of revenue for needed public investments by expanding their local bases of employment and taxable property. The practice of “buying payroll” then spread throughout the United States, and ultimately to the rest of the globe, as aid from the developed to the developing world also declined relative to an apparently relentless flow of capital in the opposite direction.

Like the shift in ends that comes into view alongside the shift in metrics, the shift in the means of development, as embodied in the practice of buying payroll, is critical because it involves a drastic shift in public fiscal priorities and allocations. Its mechanisms tend to constrain government development activity both directly and indirectly, by siphoning public resources away from infrastructure investment, or from direct income or wage support aimed at increasing consumption and employment broadly throughout a nation or region.^c Through development “incentives,” at least a portion—estimates vary—of those resources are instead placed in the hands of private corporations or entrepreneurs to reduce their production costs, increase their profits, and make it more likely for them to create new jobs.

Additional pressure is brought to bear on the public fisc of smaller or poorer nations, regions, or cities when efforts to privatize the production and delivery of services that had been the province of the public sector (e.g., transport, sanitation, social services) are also pursued as part of the privatization of development. In these places, reducing the size of government, in terms of both employment and expenditures—including any income support to workers and families mentioned earlier—is often a condition for continued financial aid from the International Monetary Fund or World Bank. Government is thereby further constrained from development activity even as an employer of civil servants or as an employer of last resort. Ultimately, then, the privatization of development reduces the role of many national and subnational governments in the development process to one of support for, and dependent reliance upon, private investment to generate employment.

It could be argued that this is not a new circumstance because governments have subsidized private industry in one form or another for as long as private

^bSee Ref.^[3].

^cIn the United States, these programs have included social security, Aid to Families with Dependent Children, the “old” welfare system or individual subsidies to poorer workers like the Earned Income Tax Credit (a so-called negative income tax). In other countries, both developed and developing, the size and scope of these income and wage support programs has been much larger.

industry has existed. But among other differences, perhaps the decisive one is that subsidies granted to private industry by subnational governments under a privatized developmental scheme aim explicitly to benefit one particular place, regardless of the effects on other places within a given nation or region. More than one observer has pointed out, for example, that the great migration of manufacturing firms and jobs that occurred between 1960 and the present, from the U.S. Midwest to the South and West, and eventually, to Mexico and East Asia, has been facilitated and accelerated by the market-like process of “competition” between regions for new industries and jobs (e.g., Ref.^[4]) that the privatization of development creates. In this case, the competition for a privatized “unit of development” results in clear economic winners and losers at national, regional, and local levels.

By contrast, to return to the case of the United States, central government aid to the private sector for development purposes was traditionally disbursed with broad national goals in mind. The railroad land grants of the 1870s and 1880s furthered the settlement of the western half of the continental United States, with broad and incalculably large public benefits. Likewise, the U.S. government’s loan guarantees to the Chrysler Corporation in the mid-1970s saved hundreds of thousands of jobs not just in the industrial Midwest, but throughout the nation.

THE IDEOLOGY OF PRIVATIZATION

The public sector in a capitalist economy has been characterized as “a means of organizing the provision of some goods and services collectively rather than individually.” In 1924, however, John R. Commons saw it as much more than that:

[The] state also proportions the factors over which it has control. It opens up certain areas, localities, or resources, instead of others.... It encourages or protects certain businesses or classes of business, certain occupations or jobs, rather than others. It restrains certain activities deemed detrimental to the whole. Its proportioning of factors is the proportioning of inducements to individuals and associations of individuals to act in one direction rather than other directions.

Public economic problems thus involve the distribution and redistribution of resources, as well as opportunities for profitably appropriating those resources. These problems are inherently political; as such, they are resolved through political means. It is not surprising, then, that at the heart of privatization in economic development (and in all other economic

matters) is a political movement whose ideological substrate is a thorough-going, distinctly North American—and more recently, British—doctrine of *privatism*^d that has pervaded economic habits of mind since the founding of the United States. As Veblen^[5] pointed out, these principles

of natural (pecuniary) liberty... find [their] most unmitigated acceptance in America.... Nowhere else has the sacredness of pecuniary obligation so permeated the common sense of the community and nowhere else does pecuniary obligation come so near being the only form of obligation that has the unqualified sanction of current common sense.

The most radical and now most popular variant of this doctrine has been distilled from selected writings of Bentham, Smith, and James Mill, to wit, that all welfare flows solely from private pecuniary gain, and the right of the individual to pursue this desideratum dominates all others. In considering problems of economic development, its relevant corollary is that because government activity is not disciplined by a pecuniary motive, it can be nothing but a dead weight on national, regional, and individual economic performance. Its activities must therefore be strictly curtailed, lest they interfere with private economic endeavor. Soros^[6] called this philosophy “market fundamentalism,” and it drives all current efforts to privatize the process of economic development.

THE CASE FOR PRIVATIZATION

Like many of the sweeping changes in economic policy and practice we have witnessed in recent years, the ideological justification for privatizing economic development has been buttressed in theory by the standard microeconomic model of the market. The state of Mississippi, for example, began its privatization program in 1936, calling it the “Balance Agriculture With Industry” (BAWI) program. One well-respected regional economist^[7] later wrote that its intent was to “provide full employment for the available able-bodied manpower of working age” in Mississippi because “opportunities for off-farm employment [were] only a small fraction of the number needed.” He also reported a study^[8] of the BAWI program where 12 communities who “purchased” new industries received “an average return [on their investments]... of 800 per cent per year.” Moes asserted that these were results that “almost any community that has un- or

^dThe term is somewhat novel, but not entirely new.

underemployment could achieve if it were willing to do what it takes to 'buy' industry."

Moes's justification for this hyperbolic statement was straightforward. Industrial enterprises make decisions about where to build new plants based on production costs at different locations. Governments could attract new businesses to their areas by competing with each other to create production cost differentials large enough to influence a firm's location decision. When communities "buy payroll," local governments become "income-maximizing corporation[s] making an investment the returns of which are measured in terms of aggregate income gains."

Bartik^[9,10] and others (e.g., Refs.^[11,12]) argued that the "purchase" of private sector jobs can cause one-time increases in local employment, and these increases have long-term benefits for communities because labor markets are subject to hysteresis effects. Bartik believed that hysteresis occurs when "an economy's equilibrium unemployment rate [is] permanently increased by a temporary recession, or lowered by a temporary boom." "If this is true," he continued, "then macroeconomic policies that affect... short-run performance... may affect... long-run performance."

Bartik estimated the impact of employment growth on unemployment, labor force participation, weekly hours worked, real wages, real earnings, and income distributions in 114 local labor markets in the United States. Using both aggregate and microlevel data, his analyses supported the hysteresis hypothesis. First, he found that local employment growth—again, purchased from the private sector with tax expenditures—could lower a given area's unemployment and raise its labor force participation rate for as long as 8 years. Second, for every additional 100 new jobs created by any given job purchasing policy, 7 will go to local residents who otherwise would have been unemployed, and 16 to residents who otherwise would not have been labor force participants. The rest go to in-migrants.

Third, although "a rising tide lifts all boats" and differential impacts "are relatively small compared to the mean long-run effect for the sample," the benefits of job growth vary across races and age groups. It helps African Americans, young workers, and less-educated workers "advance to better-paying occupations," who "are [then] able to keep [them] even in the long-run." Employment growth increases real earnings and tends to equalize income distribution. These effects were so heavily biased in favor of "lower-earnings males" (e.g., African Americans and the less educated) that "overall effects of local growth on the income distribution," even after the almost inevitable increases in property values that result from job-purchasing policies are factored in, were "at least modestly progressive."

Finally, Bartik stated—and this was not tested in his empirical analysis, but justified on theoretical grounds—that job-purchasing policies by nations and communities can improve overall economic welfare at the national level. Interstate and intermetropolitan competition for the purchase of economic development can redistribute employment and employment growth from areas with low unemployment to those with high unemployment. This can result in a net gain in national economic welfare because the marginal benefit of an additional job in the latter community is higher than the marginal cost of losing a job in the former. He also averred that the probable net effect of competition for jobs increases the aggregate national level of subsidies to business, which he views as subsidies for labor demand. The higher their overall level, the larger the number of workers that businesses hire, and the faster national employment will increase.

THE CASE AGAINST PRIVATIZATION

Overall, Bartik believed that the changes wrought by the privatization of economic development will improve the lot of poor people in poor places. Yet, the market pressures imposed on the public sector in recent years have never been shown to increase the well-being of the poor. Hysteresis plausibly works in both directions.

Moreover, since its conception, the microeconomic justification for privatizing economic development with a "market for jobs" has been questioned by a host of analysts. An essential feature of the original argument was that given unimpeded subsidization by all communities and nations,

a rational subsidization policy would mean that [each] community acts as a discriminatory seller of labor, i.e., as a monopolist who has full discriminatory power. This power derives from the presence of scarce immobile resources within the community... and the community would never offer more than the minimum necessary to attract a given industry.^[7]

Thompson^[13] vigorously attacked this claim. "Had [Moes] but turned this conclusion around," he wrote, "it would have seemed more in accord with the economic facts":

Whether one considers the entire nation or a particular region, the number of new industrial plants established in any year is small in comparison to the large and constantly growing number of communities actively courting new industry. *Therefore, the matter of comparative numbers alone would place site-seeking firms in a stronger bargaining position than industry-seeking communities. The average firm would have*

greater withholding power than the typical community, for the community would ordinarily need new industry much more than the firm would need any one community for its plant site (my italics).

Thompson concluded that “if either participant in the bargaining over the terms of a subsidy would be in a position to act as a perfectly discriminating monopolist, it would almost certainly be the industrial firm.” A site-seeking firm would play one community against another, and use its superior knowledge and withholding power to extract the largest subsidy possible. The “price” of new jobs for any given community—the value of the development incentives—would not only be much higher than predicted by Moes, but also the monopoly power of the seller would not be easy to break. “The potential barriers to competitive pricing in this hypothetical ‘market’ for subsidies appear to this writer to be...formidable,” Thompson said. “If labor markets are marked by various imperfections, the same could be said about the subsidy ‘market’”:

[it] would consist of relatively few, sporadic transactions, and community representatives would tend to be poorly informed as to such factors as the market situation confronting the industrial prospect, the locational advantages and disadvantages of the community for this particular firm, the amounts of subsidy offers by competing communities, etc. This lack of relevant information on the part of communities would be aggravated by the atmosphere of secrecy in which the process of plant site selection is typically conducted.

The most important point for poor communities or nations seeking new industries and jobs, Thompson asserted, is that more affluent places would have a distinct advantage in any market for new jobs or industrial enterprises. “The poverty of many depressed, labor-surplus communities and nations would severely handicap them in bidding competitively for new industry” if the size of a subsidy were the determining factor in any specific case (see Ref.^[13]).^e This assertion is supported by data about the spatial distribution of foreign direct investment by transnational corporations in recent years, which has overwhelmingly favored rich countries and regions within them.^[16,17]

Since the 1960s, other evidence from research on the effectiveness of development incentives in improving local, regional, or national economic welfare buttressed Thompson’s case. Development incentives to

private firms were not a significant determinant of industrial location or interregional manufacturing employment growth,^[18–20] and their benefits to smaller and poorer places were in no way as unambiguously positive as Moes and later analysts such as Bartik contended. Benefits and costs were shown to vary with the assumptions and methods employed to assess the effect of development incentives, and with the type of incentive that was being evaluated.^[20–27] Thompson^[13] observed that the effects of development incentives were detectable, if at all, only at the stage of a location decision where two towns of roughly equal desirability were being compared by a site-seeking firm. Most, if not all, the studies of plant location were carried out at the interregional level, a point recognized by McMillan when he stressed the distinction between “determinants of plant location,” a set of variables that differ interregionally, and the “reasons manufacturers choose a plant site,” a decision influenced by local factors, some of which might include incentives.

WHENCE PRIVATIZATION?

Given the conflicting evidence, the benefits of the privatization of economic development are ambiguous for communities and nations seeking increases in employment and improvements in their citizens’ economic welfare. Nevertheless, in the world of economic development practitioners and government officials at national and subnational levels, there is no debate. They have completely embraced the process.

Seemingly unaware of this fact, academic researchers have continued to focus on whether development incentives, considered apart from the more general process of privatization, contribute significantly to changes in industrial location patterns. This question has become largely irrelevant, as more communities and nations have begun to compete with each other to lure new industries to their jurisdictions. In fact, participation in competitive bidding for new manufacturing plants or corporate facilities appears to no longer be a voluntary process.

Benson,^[28] arguing that “taxes matter” in the geographic distribution of employment growth, inadvertently stumbled upon this point. He reasons that because subnational jurisdictions now compete with each other on a routine basis to attract business activity, that “[this] competition prevents taxes from varying sufficiently enough to have a statistically measurable impact on relative economic growth....” He argued further that the spread of competition for new industrial enterprises imposes a “market constraint”^[28] upon policy makers worldwide, forcing

^eRubin and Rubin^[14] made this point in a more recent publication. Bowman^[15] also found that cities with high degrees of fiscal stress were much more likely to use development incentives—and a wider array of them—than cities in relatively good fiscal health.

them to keep taxes, wages, and working conditions in line with those of other regions or nations, either through legislation or through the provision of development incentives.

The net effect of the privatization of economic development, as we noted earlier, is to place constraints on the size and fiscal health of the public sector. Moore and Pagano^[29] argued presciently that these constraints create “dismal” prospects for public finance in the 1990s and beyond:

Localities [and small or poor nations] faced with already difficult fiscal problems find themselves utilizing scarce debt capabilities for purposes that may be considered less than ideal public investments (e.g., a movement away from traditional infrastructural purposes); less able to compete effectively in a growing tax exempt market; forced to raise interest rates (and therefore future debt service burdens) to compete with taxable corporate bonds; and providing tax benefits in the form of...privately-oriented subsidies that further limit current receipts.

With narrowed national or international credit options, and uncertain prospects of outside assistance in the near and long term, the only apparent alternative for local and national governments in the developed and developing world is to stimulate private sector growth so that jobs, income, and the base of public revenue-generating wealth can expand.

The market for industrial enterprises takes on a special importance in this setting. It is not simply a market for industrial plants, but in the most general sense, a market where governments attempt to buy private investment to fund public services and public infrastructure. “Development” becomes a divisible commodity whose price is determined by market-like processes. The market for industrial enterprises thus appears as a key feature of the privatization of the public sector itself. It is a way to impose market-like mechanisms^[30] upon public bureaucracies and political processes, to transform public goods into private ones.

As we said, there is conflicting evidence about the degree to which even private sector growth is spurred by development incentives. Yet, there has been almost no discussion about the effects on the public sector of the transformation of economic development into a private good through the imposition of a market for industrial enterprises and jobs. It appears that the transformation of economic development into a divisible and saleable commodity has contributed to slower public sector growth, and declining levels of investment in public needs at all levels of government. In many parts of the developing world, where public authorities are often the only source from which investments in basic needs such as clean water, transport, and health

care are available, these trends have and could be harmful to human, as well as economic, development.^[31]

Even more fundamentally, given the “private” (i.e., secretive) nature of the negotiations carried out between government officials and corporate executives seeking incentive packages in the privatized market for economic development, what effect has its emergence had on democratic public participation in decisions about development? After almost 60 years of debate, this is the crucial question, but the one least frequently asked. If we value democratic participation intrinsically, by defining development not in merely economic terms but as improvements in overall human functioning—including a person’s ability to participate in the political life of one’s community—it appears initially that the answer is not what we might want to hear.

REFERENCES

1. Sen, A.K. *Development as Freedom*; Alfred A. Knopf: New York, 1999.
2. Stohr, W.B.; Taylor, D.R.F. *Development from Below*; John Wiley and Son: New York, 1981.
3. Luker, W.A., Jr. Buying payroll: Industrial development incentives and the privatization of economic development. *West. Tax Rev.* **1991**, 2 (8), 17–40, Summer.
4. Crandall, R. *Manufacturing on the Move*; The Brookings Institution: Washington, DC, 1993.
5. Veblen, T. *The Theory of Business Enterprise*; 1904; 130.
6. Soros, G. *On Globalization*; Public Affairs Press: New York, 2002.
7. Moes, J.E. The subsidization of industry by local communities in the south. *South. Econ. J.* **1961**, XXVIII (2), 187–193, October.
8. Hopkins, E.J. *Mississippi’s BAWI Plan, an Experiment in Industrial Subsidization*; Department of Research and Statistics, Federal Reserve Bank of Atlanta, 1944.
9. Bartik, T.J. Business location decisions in the United States: Estimates of the effects of unionization, taxes, and other characteristics of states. *J. Bus. Econ. Stat.* **1985**, 3 (1), 14–22, January.
10. Bartik, T.J. *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute for Employment Research: Kalamazoo, MI, 1991; pp. 103, 155, 185.
11. Papke, J.A.; Papke, L.E. Measuring differential state-local tax liabilities and their implications for business investment location. *Natl. Tax J.* **1986**, XXXIX (3), 357–365, September.
12. Wheaton, W.C. Interstate differences in the level of business taxation. *Natl. Tax J.* **March 1983**, XXXVI (1), 83–94.
13. Thompson, J.H. Local subsidies for industry: Comment. *South. Econ. J.* **1962**, XXIX (2), 114–119, October.
14. Rubin, I.; Rubin, H.J. Economic development incentives: The poor (cities) pay more. *Urban Aff. Q.* **1987**, 1 (23), 37–62.

15. Bowman, O'M.A. *Tools and Targets: The Mechanics of City Economic Development*; National League of Cities and Lincoln Institute of Land Policy: Cambridge, MA, 1987.
16. Glickman, N.J.; Woodward, D.P. *The New Competitors: How Foreign Investors are Changing the U.S. Economy*; Basic Books: New York, 1989.
17. Graham, E.M.; Krugman, P.R. *Foreign Direct Investment in the United States*; Institute for International Economics: Washington, DC, 1991.
18. Schmenner, R. *The Manufacturing Location Decision: Evidence from Cincinnati and New England*; Harvard-MIT Joint Center for Urban Studies: Cambridge, MA, 1978.
19. Kieschnick, M. *Taxes and Growth: Business Incentives and Economic Development*; Council of State Planning Agencies: Washington, DC, 1981.
20. Wheat, L.F. The determinants of 1963–1977 regional manufacturing growth: Why the south and west grow. *J. Reg. Sci.* **1986**, 26 (4), 635–659.
21. Rinehart, J.R. Rates of return on municipal subsidies to industry. *South. Econ. J.* **1963**, XXIX, 297–306, April.
22. Sazama, G.W. A benefit–cost analysis of a regional development incentive: State loans. *J. Reg. Sci.* **1970**, 10 (3), 385–396.
23. Mulkey, D.; Dillman, B.L. Location effects of state and local industrial development subsidies. *Growth Change* **1976**, 7 (2), 37–43, April.
24. Rinehart, J.R.; Laird, W.E. Location effects of state and local industrial development subsidies: Comment. *Growth Change* **1977**, 8 (4), 44–46, October.
25. Morse, G.W.; Farmer, M.C. Location and investment effects of a tax abatement program. *Natl. Tax J.* **1986**, XXXIX (2), 229–236, June.
26. Stinson, T.F.; Dewar, M.E.; Sullivan, K. *Taxes, Jobs, and Market Growth Rates*; State and Regional Research Center, Hubert H. Humphrey School of Public Affairs, University of Minnesota, 1988. Working Paper 88-01, March.
27. Dewar, M. *Tax Incentives and Public Loans and Subsidies: What Difference Do They Make in Non-metropolitan Economic Development?* State and Regional Research Center, Hubert H. Humphrey School of Public Affairs, University of Minnesota, 1988. Working Paper 88-04, December.
28. Benson, B.L. Do taxes matter? The impact of state and local taxes on economic development. *Commentary* **1986**, 13–17, Winter.
29. Moore, R.J.; Pagano, M. Capital formation, supply-side economics, and the public sector. *South. Rev. Public Adm.* **1983**, 6 (4), 450–464, Winter.
30. Rothschild, E. *The Real Reagan Economy*; Review of Books: New York, 1988, June 30.
31. Blair, J.P.; Fichtenbaum, R.H.; Swaney, J.A. The market for jobs: Locational decisions and the competition for economic development. *Urban Aff. Q.* **1984**, 1 (20), 64–77.

Privatization and Public-Private Partnerships for Local Services

E. S. Savas

*School of Public Affairs, Baruch College, City University of New York,
New York, New York, U.S.A.*

INTRODUCTION

This article defines and describes the various forms of privatization applied to local government services, the rationale and opposition to privatization, and evidence on the results and extent of privatization. It discusses the logical expansion of privatization to public-private partnerships and the emerging concept of government by network.

Local government services are a daily necessity. People expect an abundant flow of clean water from the kitchen faucet and their household garbage to disappear from the curb. They want expedited construction of needed new roads. They demand efficient, effective, and equitable provision of local services, and this pressure to perform has forced local governments to turn to alternative service arrangements as an important strategic tool to improve service delivery.

DEFINITIONS

Introducing competition to monopolistic public agencies was proposed in 1971 as an explicit, unconventional approach to reduce costs and increase efficiency of local government services.^[1] This policy quickly caught on and focused on competitive contracting with private companies for services, and the concept was termed “privatization.”

Privatization can be defined broadly as relying more on the private institutions of society and less on government to satisfy people’s needs.^[2] It is the act of reducing the role of government and increasing the role of other societal institutions in producing goods and services and in owning assets. These institutions are the market, voluntary groupings of all kinds, and the family; together these constitute the private sector. Instead of expecting government alone to address societal needs, privatization takes advantage of the capabilities of these other institutions and aims at a new balance, a better allocation of roles and responsibilities among these four societal institutions.

Three broad methods are used to privatize government services and government-owned enterprises and

assets: delegation, divestment, and displacement. Delegation means that government is responsible for a service but utilizes the private sector to deliver it. Divestment means that government disposes of an asset, activity, or enterprise. (Often “privatize” is used narrowly to refer only to divestment.) Displacement occurs when a private for-profit or non-profit organization starts providing a service and gradually replaces a government service. The particular techniques within these three general methods are listed and illustrated in [Table 1](#). One can combine techniques, as is illustrated below.

Contracting out, also called outsourcing, is the most common privatization technique for local services. Outsourcing is a common practice in the private sector where a firm contracts to have a non-core activity provided by another firm for whom that activity is in the latter’s core area of competence. Examples include payroll processing, advertising, warehousing, and operating the company cafeteria.

Inasmuch as the basic idea is to introduce competition in monopolistic government services to improve performance, “competitive sourcing” has replaced “privatization” both as the term of choice for contracting and in practice: competition for service delivery is open to the government agency itself and not only to outside bidders.^[3] (“Managed competition” is another term for this process.) It is often found that after losing several rounds of competition and seeing its service area shrink as private contractors take over part of the service, the public agency gradually adopts the efficient practices of the private firms, becomes competitive, and starts winning some bids. The city of Indianapolis became a model of sensible privatization and managed competition during the administration of Mayor Stephen Goldsmith.^[4] When the service is retained in-house after competitive sourcing, this cannot logically be called privatization.

It is increasingly common to refer to “public-private partnerships,” a less contentious term than “privatization.” A public-private partnership is defined broadly as any arrangement between a government and the private sector in which the latter performs traditionally public activities. Increasingly, this term

Table 1 Forms and examples of privatization

Form	Examples
Delegation	
By contract	See Table 2
By franchise	Bus service, water supply, long-term operation of toll roads
By lease	Stadiums, parking garages, golf courses, airports
By subsidy (grant, low-interest loan, tax status, etc.)	Low-income housing, social services, cultural institutions, private bus lines
By voucher	Food stamps, housing, day care, schooling
By public-private partnership	Local-government infrastructure: toll roads, bridges, airports, school buildings, water and wastewater-treatment plants, correctional facilities
Divestment	
Sale	Tax liens on foreclosed buildings, excess land, city-owned businesses, public hospitals, city-owned TV and radio stations in New York
Free transfer	Example: Buenos Aires privatized by giving its buses to the drivers
Displacement	
Default	Van services and gypsy cabs when municipal bus service is inadequate; churches and other non-profits operating social services, soup kitchens
Withdrawal	A city can close payment offices and enable citizens to pay taxes, water bills, etc., over the telephone or Internet using private billing services
Voluntarism	Voluntary services when government programs are deemed inadequate; Adopt a Highway; Adopt a School, Adopt a Park
Deregulation	Appropriate deregulation of child day care to allow more private provision; elimination of rent controls to encourage housing construction and reduce the need for public housing

Adapted from Savas, E.S. *Privatization in the City: Successes, Failures, Lessons*; CQ Press: Washington, DC, 2005; 18.

is applied more narrowly to a complex undertaking involving government and several commercial partners to finance, plan, build, and operate—for up to 99 years—a large, long-lived, revenue-generating infrastructure project (e.g., highway, bridge, water-treatment plant, or airport). It also applies to an arrangement where a public agency contracts with a non-profit entity that also raises funds from private donors (including corporations) to supplement the contract payment. For example, Central Park in New York City is managed and maintained by the non-profit Central Park Conservancy—formed to rescue the Park when the government failed to maintain it properly—which raises four times as much money annually from donors as it is paid under its city contract.^[5] (This example illustrates privatization by displacement via default, voluntarism, and contract.)

Goldsmith and Eggers point out that public-private partnerships can involve relationships among a variety of public, non-profit, private, and other governmental bodies to achieve a public purpose; they call this “government by network.” Designing and managing such increasingly common organizational networks calls for a very different set of skills than conventional hierarchical government.^[6]

THE RATIONALE FOR PRIVATIZATION

The most compelling reason for contracting, or outsourcing, is increased efficiency. The benefits derive ultimately from introducing competition and challenging in-house monopolies of conventional public services. Privatization is also useful when a city wants rapid startup of a new service or to avail itself of specialized expertise.

Communities that are too small to realize economies of scale through in-house provision of a service may be able to hire contractors who service several communities and thereby can reap those benefits of scale, incur lower costs, and charge less than the cost of in-house work.

Privatization by divestment is popular because it raises money by selling surplus property and government businesses that are not core government functions. For example, New York City government sold a hotel, vacant lots, garages, thousands of residential buildings, a television station, and two radio stations. The state of New York sold golf courses, armories, and underutilized mental-health facilities. Divestment thrusts formerly government-owned enterprises and properties into the competitive marketplace.

OPPOSITION TO PRIVATIZATION

Privatization is philosophically attractive to many who believe that government has grown too dominant in society and should be limited in size and scope. It is philosophically repugnant, however, to those of the opposite persuasion. Ideological opponents fear a loss of democratic accountability under privatization, a diminution of government's role, a loss of common values, and growing inequality. They discern and reject any implication that public managers and employees are self-serving, profit-maximizing bureaucrats.^[7] They believe that government cannot control contractors and enforce contract specifications. And indeed, government must be a smart buyer of contract services, an attribute not always to be found and that must be cultivated.^[8] They point to the possibility of corruption, bribery, collusion among private firms, or extortion of private firms by government officials.

Market failure can stymie privatization because ideal markets require competition, perfect information, and no externalities. Indeed, favorable conditions may not exist, but market failure must be weighed against government failure. In any event, privatization is situation specific, that is, the advantages and disadvantages depend on the particular service and its setting.

Displacement of public employees is often cited as a barrier to privatization, but workers are rarely discharged when a service is contracted out; attrition and early retirement are the rule, and contractors are usually eager to hire most of the current employees.

The strongest opposition to privatization comes from public-employee unions, who fear a loss of jobs and dues-paying members. They challenge the evidence of the numerous studies, asserting that they ignore the true cost of the transition to a private contractor and that the performance of the contractors has not been adequately monitored and measured.^[9] Middle managers in the public sector are also apprehensive about privatization, afraid that they will be released or downgraded. Opposition takes many forms, ranging from the familiar kinds of labor actions up to and including assassination.^[10]

EXTENT OF PRIVATIZATION

It is difficult to survey local governments en masse and uncover all the privatization cases, although it has been done for Anaheim, California, and New York City.^[11] It is much easier to survey the extent of contracting, however, and comprehensive data are available. Private contractors provide almost all local government services, and essentially every local government contracts for some services.^[12] The most common ones are listed in Table 2. Local governments use only their

Table 2 Extent of privatization of most commonly contracted services

Service	Percent of cities
Vehicle towing and storage	83
Legal services	53
Labor relations	53
Residential refuse collection	49
Solid waste disposal	41
Ambulance service	37
Tree trimming and planting	37
Street repair	35
Bus system operation	30
Traffic signal maintenance	24

Adapted from Morley, E. Local government use of alternative service delivery approaches." In *Municipal Yearbook 1999*: International City/County Management Association: Washington, DC, 34-44.

own personnel for half their services, both public employees and contractors for 23% of their services, and only private firms for 18%.^[13] Contracting more than doubled, increasing by 121% between 1982 and 1992 in the 596 local governments where comparable data were available,^[14] but may have leveled off recently.^[15] Moreover privatization diversity also increased, that is, more of the different forms of privatization (contracting, vouchers, franchising, volunteers, and self-help) were being used.^[16] Several cities eschew public employees almost entirely, using contractors for all but public safety functions.^[17] A city in receivership was restored to financial health in four years through total privatization.^[18]

THE RESULTS OF PRIVATIZATION

The most studied forms of privatization are contracting for municipal services and divestment of government-owned businesses and property. The consensus is that—properly carried out—privatization by competitive contracting increases the efficiency of local services.^[19] In the most exhaustive compilation of studies, Hodge concludes, "There is little doubt the weight of evidence appears to support the notion that on average the unit costs of services are reduced through competitive tendering of public services."^[20] After studying two decades of worldwide contracting experiences, Domberger and Rimmer conclude that reduced service costs "appeared to be consistent over this period of time, throughout all the countries reviewed, and over all levels of government."^[21] Savings range as high as 35% for public works functions, which are easiest to outsource and to measure.^[22]

With respect to the other major form of privatization that has been studied extensively, divestment, the incentive is self-evident: revenue is realized as governments sell properties they do not need or businesses that do not perform core functions of government. Moreover performance of the enterprises generally improves upon divestment. Because there are relatively few government-owned business enterprises in the United States, the evidence comes from comparisons of public and private enterprises abroad^[23] and from before-and-after studies of newly privatized government-owned enterprises abroad.^[24]

The other specific forms of privatization listed in Table 1 have not been studied much because each case is different, but the forms are popular, and surveys indicate that local governments are using them more.^[25]

The observed benefits of privatization are generally achieved not by paying lower wages but by increasing productivity, that is, realizing more work output per hour of effort. This is obtained in different ways depending on the particular service. It is usually the result of more capital investment, better equipment and technology, better work organization, and better management: managers are free of the constraints typical of the public sector and can assign and supervise employees better and can reward and promote good workers and withhold rewards from—or dismiss—unsatisfactory ones.^[26] The higher productivity of a privatized service is often associated with better but fewer employees who receive higher wages.

HOW TO PRIVATIZE

The processes involved in privatization by divestment^[27] and through public-private partnerships^[28]

are well documented. The steps in contracting for services are shown in Table 3. Contracting is not a simple issue. One has to consider transaction costs, principal-agent problems, market imperfections, nature of the contract, the basis for evaluating bids, the ability to monitor and enforce contractor performance, and the basis for contract termination or renewal, for example.^[29]

Segmenting the service into smaller units or service districts and inviting bids for each district can foster competition. Limiting the number of units that any one contractor can be awarded and staggering contract start dates so that 3-year contracts are up for bid every year for some districts encourage bidders to keep competing. Of paramount importance is developing a transition plan for existing public employees who will be affected by the privatization. Common approaches are to privatize gradually—that is, only part of the service at a time—and rely on attrition; to require the contractor to hire existing employees; to give severance pay and early retirement; to give released employees preference in future openings; and to save vacancies that occur in other departments during the planning period and then fill those vacancies with the affected employees after the contract goes into effect.

CONCLUSION

Privatization is a common good-government practice, widely used by officials of all political and ideological persuasions. Many particular techniques are available. Privatization generally leads to significantly better performance provided that good management practices are followed.

Table 3 Steps in contracting for services

1. Introduce the idea of privatization by contracting
2. Select a service that is a good candidate for contracting
3. Conduct a feasibility study
4. Foster competition
5. Request expressions of interest or qualifications
6. Plan the employee transition
7. Prepare bid specifications
8. Initiate a public relations campaign
9. Engage in competitive sourcing ("managed competition")
10. Conduct a fair bidding process
11. Evaluate the bids and award the contract
12. Monitor, enforce contract performance, evaluate

Adapted from Savas, E.S. *Privatization and Public-Private Partnerships*; Chatham House Publishers: New York, 2000; 175.

REFERENCES

1. Savas, E.S. Municipal monopoly. *Harpers Magazine* **1971**, December, 55–60.
2. Savas, E.S. *Privatization: The Key to Better Government*; Chatham House Publishers: Chatham, NJ, 1987; 3 pp.
3. Martin, L.L. Public-private competition: a public employee alternative to privatization. *Rev. Public Personnel Admin.* **1999**, Winter, 59–70.
4. Goldsmith, S. *The Twenty-First Century City: Resurrecting Urban America*; Regnery Publishing: Washington, DC, 1997.
5. Savas, E.S. *Privatization in the City: Successes, Failures, Lessons*; CQ Press: Washington, DC, 2005; 170, 305.
6. Goldsmith, S.; William, D.E. *Governing by Network: The New Shape of the Public Sector*; Brookings Institution Press: Washington, DC, 2004.

7. Greene, J.D. *Cities and Privatization: Prospects for the New Century*; Prentice-Hall: Upper Saddle River, NJ, 2002; 16–20.
8. Kettl, D.F. *Sharing Power: Public Governance and Private Markets*; Brookings Institution: Washington, DC, 1993; 163, 179–211.
9. Sclar, E.D. *You Don't Always Get What You Pay For*; Cornell University Press: Ithaca, NY, 2000.
10. Savas, E.S. *Privatization and Public-Private Partnerships*; Chatham House: New York, 2000; 299 pp.
11. Savas, E.S. *Privatization in the City: Successes, Failures, Lessons*; CQ Press: Washington, DC, 2005; 44–45.
12. Touche Ross & Co. *Privatization in America*; Washington, DC, 1987; 3 pp.
13. Warner, M.; Hefetz, A. Pragmatism over politics: alternative service delivery in local government, 1992–2002. In *Municipal Year Book*; International City/County Management Association: Washington, DC, 2004; 11 pp.
14. Greene, J.D. *Cities and Privatization: Prospects for the New Century*; Prentice-Hall: Upper Saddle River, NJ, 2002.
15. Warner, M.; Hefetz, A. Pragmatism over Politics: alternative service delivery in local government, 1992–2002. In *Municipal Year Book*; International City/Country Management Association: Washington, DC, 2004.
16. Greene, J.D. *Cities and Privatization: Prospects for the New Century*; Prentice-Hall: Upper Saddle River, NJ, 2002; 79–80.
17. Greenblatt, A. Observer. In *Governing*; 17–18 (February 2006).
18. Kleiman, R.T.; Sahu, A.P. Privatization as a viable alternative for local governments. In *Contracting Out Government Services*; Seidenstat, P., Ed.; Praeger: Westport, CT, 1999; 163 pp.
19. Greene, J.D. *Cities and Privatization: Prospects for the New Century*; Prentice-Hall: Upper Saddle River, NJ, 2002, 38–50; Savas, E.S. *Privatization and Public-Private Partnerships*, 147–167; Stein, R.M. *Urban Alternatives: Public and Private Markets in the Provision of Local Services*. U. of Pittsburgh Press: Pittsburgh, PA, 1990; 187–188.
20. Hodge, G.A. *Privatization: An International Review of Performance*; Westview Press: Boulder, CO, 2000; 107 pp.
21. Domberger, S.; Rimmer, S. Competitive tendering and contracting out in the public sector: a survey. *Int. J. Econ. Business* **1994**, 1 (3), 439–453.
22. International City Management Association, *Service Delivery in the 1990s: Alternative Approaches for Local Government*; Washington, DC, 1989; Dilger, R. Jay; Moffett, R.R.; Struyk, L. Privatization of municipal services in America's largest cities. *Public Admin. Rev.* **1997**, 57 (1), 21–26; Savas, E.S. *Privatization and Public-Private Partnerships*, 148–153; Savas, E.S. Public and private refuse collection; A critical look at the evidence. *J. Urban Anal.* **1979**, 6, 1–13; Stevens, B.J. *Delivering Municipal Services Efficiently: A Comparison of Municipal and Private Service Delivery*; U.S. Department of Housing and Urban Development: Washington, DC, June 1984; 544 pp.
23. Boardman, A.E.; Vining, A.R. Ownership and performance in competitive environments: a comparison of the performance of private, mixed, and state-owned enterprises. *J. Law Econ.* **1989**, 32 (April), 1–33.
24. Galal, A.; Jones, L.; Tandon, P.; Vogelsang, I. *Welfare Consequences of Selling Public Enterprises*; World Bank: Washington, DC, 1992; Megginson, W.L.; Nash, R.C.; Van Randenborgh, M. The financial and operating performance of newly privatized firms: an international empirical analysis. *J. Finance* **1994**, 49 (2), 403–452.
25. Warner, M.; Hefetz, A. *Pragmatism Over Politics*.
26. Stevens, B.J. Comparing public- and private-sector productive efficiency: an analysis of eight activities. *Natl. Prod. Rev.* **1984**, 3 (4), 405.
27. Savas, E.S. *Privatization and Public-Private Partnerships*, Chaps. 8 and 9, Nov. 1999.
28. Stainback, J. *Public/Private Finance and Development*; John Wiley & Sons: New York, 2000.
29. Savas, E.S. *Privatization in the City: Successes, Failures, Lessons*; CQ Press: Washington, DC, 2005; 20 pp.

Privatization and Reinvention Paradigms for State Governmental Administration, Consequences of

Julia E. Robinson

J. Robinson Enterprises, Inc., Boise, Idaho, U.S.A.

INTRODUCTION

The reinvention movement burst onto the public administration scene with the publication of *Reinventing Government: How The Entrepreneurial Spirit is Transforming The Public Sector*.^[1] In their best seller, Osborne and Gaebler take up the instrumental question of “How to do government?” The term “reinventing government” quickly became the mantra for government reform advocates. The impact of reinvention themes on public policies discussion at all levels of government in the nineties cannot be underestimated.

The fundamental premise of these reinventionists is that traditional approaches to government, designed at the beginning of the twentieth century during the Progress Era and New Deal, are no longer effective for a postindustrial knowledge-based global economy. Theorists in reinvention argue for a new “paradigm” where the locus of control for design and delivery of services should be shifted whenever possible from the governmental agency to the community or empowered citizen. This new approach to government focuses on results not on rules, developing clear agency missions and goals, decentralized authority, and utilizing market mechanisms to deliver services. In entrepreneurial government, the public sector becomes the catalyst for the changes. The role of public employees is to define problems, and assemble the necessary resources to get the job done. In practice, however, the term “reinvention” has come to mean “privatization” of government services for many politicians and practitioners. The ultimate result of reinvention is that many government services are privatized, and agency personnel become contract managers.

Reinvention is a public administration theory and an approach to government service delivery. In their writings, public administration theorists in general have expressed a sense of unease that the reinvention theory moves American public administration further away from fundamental democratic values of citizen involvement, social justice, and equity while over-emphasizing efficiency and self-interest.

The popularity of reinvention is only understandable within the political and social context of the early

nineties. The popularity of reinvention can be explained by a confluence of influences, resulting in a shift from government solutions to market solutions. The millennium brings a new confluence of variables, which suggests that while reinvention strategies have been successfully enacted in federal and state governments, U.S. citizens are more willing to debate the “appropriate” role of government in providing public goods and public protection today than they were even a year ago.

OVERVIEW OF THE REINVENTION PARADIGM

The reinvention movement burst onto the public administration scene with the publication of *Reinventing Government: How The Entrepreneurial Spirit is Transforming The Public Sector*.^[1] In their best seller, Osborne and Gaebler take up the instrumental question of “How to do government?” Their focus is day-to-day operations. Osborne and Gaebler argued that in a time of declining public resources, government must move from a bureaucratic model to an entrepreneurial model.

Osborne provided further clarification of the reinvention approach through actual examples in *Banishing Bureaucracy*.^[2] Osborne’s definition of reinvention is: “The fundamental transformation of public systems and organizations to create dramatic increases in their effectiveness, efficiency, adaptability and capacity to innovate. This transformation is accomplished by changing their purpose, incentives accountability, power structure and culture.”^[2,13,14]

The fundamental premise of these reinventionists is that traditional approaches to government, designed at the beginning of the twentieth century during the Progress Era and New Deal, are no longer effective for a postindustrial, knowledge-based global economy. They see the traditional governmental model as bureaucratic based on formal rules and regulations, with line-item budgets and a large number of employees providing the services in hierarchical structures. Theorists in reinvention argue for a new “paradigm,” where the locus of control for design and delivery of services should

Table 1 Contrast of traditional bureaucratic and reinvention themes

Traditional bureaucratic government	Reinvention government
• Rational-legal authority based on the rule of law • Hierarchical administrative structure with formal lines of authority. Top-down administration • Division of labor • Authority lies with the position rather than the individual • Established channels for communication • Ordered by written rules: apply general rules to specific cases, resulting in fair and impartial results • Standardized and impersonal • Administrative processes used to deliver services • Specialized administrative functions are provided for based on objective criteria	• Catalytic government: steering rather than rowing • Community-owned, focused on empowering rather than serving • Service delivery through competitive processes • Mission driven rather than rule driven • Results oriented: funding outcomes rather than inputs • Customer-driven: meeting the needs of the customer, not the bureaucracy • Enterprising government: earning rather than spending • Anticipatory government: prevention rather than after the fact cure • Decentralized government: emphasizes employee participation and teamwork • Market-oriented government leveraging change throughout the market

be shifted whenever possible from the governmental agency to the community or empowered citizen. This new approach to government focuses on results not on rules, developing clear agency missions and goals, decentralized authority, and utilizing market mechanisms to deliver services. In entrepreneurial government, the public sector becomes the catalyst for the changes. The role of public employees is to define problems and assemble the necessary resources to get the job done.

Reinventionists call for entrepreneurial managers who have authority over budgets to shift resources from areas of lower productivity to areas of greater yield, provide incentives to employees and to programs, and focus on leading (steering) rather than service provision (rowing). In this model, government becomes a purchaser of services through competitive processes, meaning that many traditional government roles are bid out. In the pure reinvention model, the public sector may use the public sector, private sector, voluntary non-profit sector, or a combination of any of these as the actual service provider. In practice, however, the term “reinvention” has become synonymous with “privatization” of government services for many politicians and practitioners. The ultimate result of reinvention is that many government services are privatized, and agency personnel become contract managers.

Table 1 presents a summary of major reinvention themes contrasted with traditional bureaucratic themes.

IMPACT OF REINVENTION ON THE PUBLIC AND ON GOVERNMENT AT ALL LEVELS

Osborne’s and Gaebler’s book, *Reinventing Government*, had a broad-based impact, appearing on the New York Times’ best-seller list for four editions in

the first year of publication. During the first two years after publication, there were more than 400 newspaper articles, journal articles, books, conferences, and even advertising highlighting “reinvention” at all levels of government.^[3] By the late 1990s, the book had sold around half a million copies and had been translated into 15 languages.^[4] Reinvention themes in the early 1990s struck a popular cord with elected officials and members of the general public, who had grown frustrated by what they perceived to be large bureaucratic structures at all government levels (federal, state, county, local), and which were unresponsive, inefficient, and not focused on citizen needs.

The term “reinventing government” quickly became the mantra for government reform advocates. The term “reinvention” was adopted by the general public and was incorporated into then President Bill Clinton’s policy agenda through the first National Performance Review headed by Vice President Gore with consultation from Osborne.^[5]

The reinvention paradigm also permeated state and local government discussions. Scores of state and local politicians ran for office on reinvention themes across the country. Numerous efforts were made by governors and state legislators to incorporate reinvention concepts into state government management. The most well-known theoretical discussion of the potential for reinvention strategies in state and local government is the work of the Winter Commission, named after its chair, the Honorable William F. Winter, former Governor of Mississippi. The Commission began in 1991 and submitted its report to President Clinton in 1993.^[6]

As illustrated by the previous discussion, the impact of reinvention themes on public policies discussion at all levels of government in the 1990s cannot be underestimated. In addition, there was a great deal of debate

in the public administration literature on the positives and negatives of reinvention concepts in terms of future directions of American public administration. (For some examples of this literature, see Refs.^[4,7-9].)

Key questions raised by public administration theorists include but are not limited to the following:

- Do the themes of reinvention correctly capture the current body of public administration literature and create a body of sustainable theory?
- How much decentralization is enough?
- Is a government customer the same as a citizen? Can one assume that the collective self-interest of individuals will result in the overall public good? And, if not, how is citizen involvement sustained in the reinvention approach?
- In service agencies, how are decisions of public goods versus market mechanisms made to insure public accountability? Who determines what and from whom?
- What is the role of regulatory agencies in the reinvention model? Are the bodies being regulated (probably business) the customers, and, if so, what impact does that have on public safety and cost effectiveness for citizens and not for businesses?

Public administration theorists in general have expressed in their writings a sense of unease that the reinvention theory moves American public administration further away from fundamental democratic values of citizen involvement, social justice, and equity, while overemphasizing efficiency and self-interest. At the beginning of the new millennium, the jury may still be out on the value of reinvention for American public administration theory. It is clear to every government practitioner that reinvention has left a mark in the structure and process of government.

ROOTS OF REINVENTION

The above discussion demonstrates that reinvention is both a public administration theory and an approach to government service delivery. It is important to note that the concept of making government more like business is not new. In the 1980s, Graham Allison,^[10] in a famous article comparing the differences between public and private management, used the experiences of managers who have served in both public and private capacities to draw a different conclusion than “reinvention.” He concluded that public and private management are different, and public management is harder.

Thus, the popularity of reinvention is only understandable within the political/social context of the early 1990s. The popularity of reinvention can be

explained by a confluence of influences resulting in a shift from government solutions to market solutions. Invisible and omnipresent, cultural values impinge upon and guide societal choices. By the early 1990s, an array of factors converged to enhance the appeal of reinvention as a viable political approach. The shift in cultural acceptance of reinvention can be traced to these changing dynamics.

First, in the 1980s, the business management philosophy writers began emphasizing change agents and the search for excellence in business settings. (For some examples of this genre, see Refs.^[11,12].)

Second, since the late 1960s, there has been an ongoing political transformation in our country. Citizens increasingly have moved away from voting for politicians who supported big government, such as former President Lyndon Johnson, to conservative politicians who see government as a necessary evil, typified by former Presidents Ronald Reagan and George Bush. This political transformation was documented with tax revolts by citizens in the late 1980s and early 1990s; the rise of the religious right as a powerful political presence in the state and national political arenas; and, the election of a centrist Democrat, Bill Clinton, as President. This transformation of public attitudes toward conservative themes of limited government increased fiscal accountability and devolution of authority from the Federal government to the states and paralleled the rise in popularity of the reinvention paradigm.

Third, American business was beginning the longest running bull stock market in American history, and Americans were witnessing the seemingly overnight phenomenal economic success of high-tech start-ups, which made young inventors and investors millionaires many times over. The business sector came to be viewed as the “place with all the answers” in a fast-changing world.

EXAMPLES OF REINVENTION EFFORTS IN STATE GOVERNMENTS

Along with reinvention theory, the 1990s brought increasing power and prominence to states in shaping the national agenda. The zenith of this state government movement may well prove in hindsight to have been the 1995 Republican Congress “Contract for America,” in seeking to turn the federal social programs back to the state. There is great variation among states in constitutional structure, resulting in differences in service delivery design and implementation. These differences make it difficult to generalize how reinvention has impacted states. Recent research as well as national initiatives such as welfare reform

and the implementation of the Children's Health Insurance program, however, provide some glimpses of how reinvention has penetrated the structure of state government.

At the beginning of the 1990s, many states undertook studies to determine how reinvention principles might be applied. The most well-known effort to implement suggested reinvention reforms was Minnesota's redesign of its Department of Administration, as chronicled in *Breaking Through Bureaucracy*.^[13] Research at the end of the 1990s suggests that many states have taken major steps to redesign their basic administrative units with a focus on human resources.^[14] The emerging research suggests that states are using reinvention principles with new computer technologies to streamline human resource processes. Many states are delegating authority for personnel functions to agencies and managers, shifting human resource missions to focus more on collaboration with agencies, and adopting performance management systems.^[14]

Osborne and Gaebler^[1] provided numerous examples of potential modifications of the welfare and Medicaid systems. The Winter Commission^[6] echoed the need for change in these areas, with particular emphasis on Medicaid. Former President Clinton and Congress addressed these requests for change through two major federal initiatives that essentially devolved authority for federal programs to state level. These initiatives were national welfare reform in 1996 and creation of the State Children's Health Insurance Program (SCHIP) by Congress in 1997.

Welfare reform was enacted with the passage of the Temporary Aid for Needy Families block grant (TANF). TANF represented a transfer of responsibility for welfare design from the federal to the state level and a major shift in this country's low-income safety net. Aid to Families with Dependent Children (AFDC), an ongoing cash entitlement program with federal eligibility rules, was transformed into a capped, block grant allowing states to design and implement a time-limited work program. States were given broad flexibility in the design of their respective welfare programs as long as federal work participation rates, reduced welfare caseloads, and increased participants' earnings were met. States have taken the opportunity to re-craft their programs into broad-based, workforce programs using a myriad of private contractors to provide various employment programs in combination with social supports such as transportation and day care. State efforts combined with a strong economy have resulted in large reductions in welfare caseloads along with substantial increases in the number of former welfare recipients who are working.^[15]

SCHIP was created as part of Congress's effort to delink Medicaid from welfare. SCHIP provides states

the opportunity to enroll children with family incomes up to 200% of poverty level in Medicaid, and, in turn, receive a favorable matching rate of 80% federal funds and 20% state funds to insure this population. SCHIP also allowed parents of children to be enrolled in innovative state-designed insurance programs. The original SCHIP legislation also provided some flexibility in program design, which included limited cost sharing by recipients, simplified applications, and annualized enrollment. All of these changes reflect the customer and real-world service delivery emphasis of the reinvention movement.

Enacting the SCHIP program was clearly an effort by Congress to address the growing numbers of uninsured in this country. A number of states, such as Wisconsin, New York, Texas, Florida, Vermont, and Minnesota, have been able to use the new flexibility in the program to further expand coverage to uninsured children and families.

Newly elected President Bush chose Governor Tommy Thompson of Wisconsin to be his cabinet head of the federal Department of Health and Human Services (DHHS) in the winter of 2000. As Wisconsin's governor, Thompson was one of the leaders in state welfare and Medicaid reform. As Secretary of the Department of Health and Human Services, Thompson announced in the spring of 2001 further efforts by DHHS to allow states flexibility in SCHIP program design and, thus, encourage expanded state coverage to the uninsured.

With the start of the new millennium, a clearer picture is developing with regard to the impact of reinvention on state government. Most states have made internal efforts to streamline administrative processes. But, possibly more important in terms of fundamental state services, the federal government has wholeheartedly endorsed concepts of decentralizing services to the states, providing more flexibility, and monitoring results based on measurable outcomes and targets rather than enforcing standardized processes. These changes are core values of the reinventionists, and when enacted on the federal level, have fundamentally changed the work of state health and human services agencies. Even with the economy starting to downturn in the middle of 2001, there are no current political discussions of revisiting this fundamental transformation in service delivery in these areas.

CONCLUSION

On September 11, 2001, the United States experienced the worst terrorist attack on its own soil in its history. Three commercial airlines manned by terrorists crashed into the twin towers of the World Trade Center and the Pentagon causing thousands of deaths. A fourth

airline crashed into a field in Pennsylvania. As a result of these attacks, the federal government is substantially increasing its role in public protection at U.S. airports, along coastlines, water systems, dams, power plants, and other potential terrorist sites.

An unexpected result of this terrible tragedy has been that Americans have rapidly come to accept the need for government and have developed a newfound respect for government employees from firemen to policemen, to military officers, and to national leaders. As the economy plummets in the wake of these events, the federal government under a conservative administration is initiating not only a "War on Terrorism" but also an economic stimulus package.

This article began with an examination of social and political history in the 1990s that resulted in a large group of government leaders and the public who were willing to accept "reinvention" tenets and implement those tenets into government practice. The millennium brings a new confluence of variables which suggests that, while reinvention strategies have been successfully enacted in federal and state governments, U.S. citizens are even more willing now than they were to debate the "appropriate" role of government in providing public goods and public protection today.

REFERENCES

1. Gaebler, T.; Osborne, D. *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*, 4th Ed.; Addison-Wesley Publishing Company: Reading, MA, April 1992.
2. Osborne, D.; Plastrik, P. *Banishing Bureaucracy: The Five Strategies for Reinventing Government*; Addison-Wesley Publishing Company: Reading, MA, 1997.
3. Hollings, R. *Reinventing Government: An Analysis and Annotated Bibliography*; Nova Science Publishers, Inc.: Commack, NY, 1996.
4. Thompson, F.; Riccucci, N. Reinventing government. *Annu. Rev. Polit. Sci.* **1998**, *1*, 231–257.
5. National Performance Review. *From Red Tape to Results: Creating a Government That Works Better and Costs Less*; U.S. Gov. Print. Off: Washington, DC, 1993; 244 pp.
6. National Commission on the State and Local Public Service. *Hard Truths/Tough Choices: An Agenda for State and Local Reform*; Nelson A. Rockefeller Inst. Gov: Albany, NY, 1993; 88 pp.
7. Nathan, R.P. Reinventing government: what does it mean? *Public Adm. Rev.* **1995**, *55*, 213–255.
8. deLeon, L.; Denhardt, R. The political theory of reinvention. *Public Adm. Rev.* **March/April 2000**, *60* (2), 89–97.
9. Lenkowsky, L.; Perry, J. Reinventing government: the case of national service. *Public Adm. Rev.* **2000**, *60* (4), 298–307.
10. Allison, G., Jr. Public and private management: are they fundamentally alike in all unimportant respects? In *Current Issues in Public Administration*; Lane, F.S., Ed.; St. Martin's Press: New York, 1982; 13–33.
11. Peters, T.; Waterman, R., Jr. *In Search of Excellence: Lessons from American's Best Run Corporations*; Harper and Row Publishers: New York, 1982.
12. Naisbit, J.; Aburdene, P. *Reinventing the Corporation: Transforming Your Job and Your Company for the New Information Society*; Warner Books: New York, 1985.
13. Barzelay, M. *Breaking Through Bureaucracy*; University of California Press: Berkeley, CA, 1992.
14. Selden, S.; Ingraham, P.; Jacobson, W. Human resource practices in state government. *Public Adm. Rev.* **2001**, *61* (5), 598–608.
15. Van Lare, B. Congress to reevaluate TANF in 2002. *PA Times* **July 2001**, *24* (7), pp. 1, 2, 10.

Probability Theory

Bradley J. Best

Department of Political Science, Buena Vista University, Storm Lake, Iowa, U.S.A.

Paul J. Brennan

Department of Accounting, Austin Peay State University, Clarksville, Tennessee, U.S.A.

INTRODUCTION

Public administrators use probability theory to account for uncertainty in many aspects of the policy process, especially the important functions of project planning and decision making. When engaged in project planning, public administrators employ probability theory to establish the likelihood of events that effect the allocation of scarce public resources. As a decision-making tool, probability theory is used in many techniques of policy analysis, especially when administrators are called upon to assess the expected benefits of alternative courses of action. Furthermore, probability theory is used to make inferences about the characteristics of a city, state, or nation's population and the presence of relationships important to the administration of public programs.

BASIC PRINCIPLES OF PROBABILITY

The fundamental principle of classical probability provides that the likelihood of an event is equal to the frequency of its occurrence divided by the frequency of all possible outcomes. For example, if each of 12 horses has an equal chance of winning a race, the probability that any one horse will win is 0.083 (1/12). In the same way, the probability of drawing a queen from a standard deck of 52 playing cards is equal to 0.077 (4/52). Moreover, the probability of any event lays somewhere between the values of 0 and 1. The total probability of all events is 1. An event that is certain to occur has a probability of occurrence equal to 1. Events that are certain not to occur have a probability of occurrence equal to 0. Although important theoretically, events that are either certain to occur or certain not to occur require no reference to the concept of probability.

It is important to note that probability theory is a key source of probability expectations in statistics. The examples of the horse race and the drawing from a standard deck of playing cards demonstrate that, although theoretically derived, probability expectations in statistics may also be gathered through

empirical experience (i.e., observation of “real world” events). Even the notion of subjective probability—probability established by the hunch of an expert—is founded in empirical evidence because the expert's hunch is based on past experience, even if the expert has never bothered to precisely quantify that hunch. Thus, a full appreciation of probability expectations in statistics is derived from both the implications of probability theory and the insights gained from everyday experience.

When estimating the likelihood of uncertain events, public administrators and policy analysts consider the related and important concepts of independence and conditional probability. Conditional probability refers to the likelihood that one event will occur contingent upon the occurrence of another event. The concept of conditional probability is related in the following example. Consider the hypothetical data presented in [Table 1](#). Assume these data reflect the outcome of a survey presented to the Democratic and Republican party members of a state legislature. Suppose that each respondent was asked to register an “agree” or “disagree” response to a policy proposal.

Table 1 reflects that of the 300 legislators surveyed, 90 issued an “agree” response and 210 indicated that they “disagreed” with the proposed policy change. Further, Table 1 reveals that of the 300 legislators surveyed, 200 identified themselves as Republicans and 100 indicated that they were members of the Democratic party. Most important, Table 1 indicates that the probability of selecting a legislator that “agrees” with the policy proposal from the list of Republican identifiers is equal to 0.30 (60/200). The likelihood of this event given the known existence of another event is known as its conditional probability. If two events are related, knowing conditional probabilities allows a policy maker to better predict a situation from prior knowledge of a related event.

Statistical independence implies that the probability that one event has occurred does not influence the probability that a second event will occur. If two events are independent, they are unrelated and knowing that one is present does not help a policy maker predict the likelihood of the other event occurring. Expressed

Table 1 Relationship between party membership and issue position

Party	Agree	Disagree	Totals
Republican	60	140	200
Democrat	30	70	100
Totals	90	210	300

mathematically, two events are statistically independent if the probability that both events occur is the product of their individual probabilities.

STATISTICAL INFERENCE: WORKING WITH SAMPLES AND POPULATIONS

The principles of probability outlined above are essential to the practice of statistical inference. As Anderson and Zelditch explained, probability is “the mathematical domain out of which virtually all statistical inference has emerged.”^[1] Statistical inference is used for making decisions for a wider population based on observations of a sample. A sample refers to a small subset of a population. When engaging in statistical analysis, social scientists typically examine the characteristics of a sample. Known as statistics or estimates, these characteristics are used as substitutes for the typically unknown, if not unknowable, qualities of the population (called parameters). When using estimates to make inferences about the characteristics of a population, social scientists prefer to gather probability samples. Welch and Comer related that, “In a probability sample, every object or unit of population has a known probability of appearing in the sample.” In addition, the authors explained that use of a probability sample “allows you to judge how well sample results reflect the population.”^[2] Moreover, knowledge of the “fit” between a sample and its population is the foundation of all exercises in statistical inference.

Statistical inference is used by social scientists because it allows for reliable, educated guesses regarding the characteristics of a population in the absence of an ability to examine all units that comprise the population. In addition, probability theory is closely linked to the concept of statistical inference in that social scientists are frequently called upon to draw tentative, yet reliable conclusions regarding relationships among elements that comprise a population. In short, statistical inference allows social scientists to draw these tentative conclusions and express, using the basic features of probability theory, the likelihood of an erroneous conclusion. Applications of probability theory and the related logic of statistical inference allow public administrators and other social scientists to make decisions with a known degree of confidence.

Specifically, then, how might a public administrator use probability theory and the logic of statistical inference to answer a practical question? Suppose public officials in a large community want to market services to the local population in the most cost-effective way. In addition, suppose they want to conduct a survey to test their suspicion that older residents in the community demonstrate a higher propensity to use city-owned recreational facilities. Yet, polling all residents of a large community would be either unduly burdensome or prohibitively expensive. What is a public manager to do? The answer involves gathering a representative sample of community residents, measuring their ages and tastes in recreational facilities, and using statistical inference to draw tentative, though accurate, estimates of the relationship between these two variables in the entire population.

The capacity to draw these conclusions is rooted in the concept of a probability distribution. If a large number of repeated samples are drawn from a population and if a particular characteristic is measured from each sample, a distribution of these measures—known as a probability distribution—emerges. This particular form of probability distribution is known as a sampling distribution. If policy makers understand the behavior of the sampling distribution, they may apply rules of probability to estimate a range of outcomes. In short, within a probability distribution, the likelihood of any particular value occurring may be reliably determined. Thus, public administrators are able to discern how likely it is that a particular result will occur merely by chance.

All sampling distributions are defined by two key features—their mean, or arithmetic averages, and standard deviation. The standard deviation refers to the average amount of variation of values in a group of data vis-à-vis the mean of the group of data. In other words, all else being equal, distributions with large, as opposed to small, standard deviations are characterized by a high degree of dispersion of values around the mean. By contrast, all else being equal, distributions with small standard deviations are said to reflect a high density of values immediately around the mean of the distribution. Thus, assuming equivalent means, any two distributions will vary in apparent “shape” to the extent that their standard deviations are unequal.

THE IMPORTANCE OF THE NORMAL PROBABILITY DISTRIBUTION

A basic understanding of probability theory and its applications requires familiarity with the idea of the normal probability distribution and the Central Limit Theorem. The importance of the normal probability

distribution in statistical inference is derived from the Central Limit Theorem. This theorem provides that the means drawn from repeated random samples of a population will follow a particular pattern, known as a normal probability distribution, provided that the sample sizes are sufficiently large. Considering the nature of a sampling distribution and the dictates of the Central Limit Theorem, as the number of repeated samples in a distribution increases, the “shape” or basic features of the sampling distribution closely resembles those of the population.

The Central Limit Theorem also dictates that, as the number of samples increases, the distribution of sample means will follow what is known as a normal distribution.^[3] When depicted graphically, the normal distribution takes on the familiar shape of the symmetrical, bell-shaped curve possessing a single mode. The particular case of the distribution of sample means reflects a distribution with a mean equal to the population mean and a standard deviation equal to the standard error of the mean (population standard deviation divided by the square root of the sample size). It is important that the Central Limit Theorem will hold regardless of the shape of the underlying population, but the samples drawn from populations departing substantially from normality will require larger sizes before the means of the samples can be assured to follow a normal distribution.

The Central Limit Theorem is the foundation of all exercises in statistical inference. Further, it is the reason why the normal probability distribution is so often used in inferential statistics. Yet, much data in the world does not follow a normal probability distribution and unless one knows something about the underlying characteristics of a probability distribution, one cannot predict the probability of getting a certain occurrence in a range of occurrences. Fortunately, the Central Limit Theorem tells us that, even if the underlying population is not normally distributed, means from repeated samples will follow a normal distribution provided that the sample size is sufficiently large. Armed with a sample mean and an estimate of the standard error of the mean, one can make a prediction about the population mean and the probability of obtaining a given sample mean from that population. This condition holds due to fact that the sample means will follow a normal probability distribution with a mean equal to the population mean and a standard deviation equal to the standard error of the mean. In addition to the sample mean, other sample statistics follow normal probability distributions or closely related distributions, such as the t distribution.

Often, researchers want to estimate, with a known measure of confidence, a particular characteristic of a population. Beyond the Central Limit Theorem, an additional concept, the principle of large samples, is

of particular importance to students of probability theory. An example illustrates this point most effectively. Suppose we wanted to know the average age of all citizens of China. In our effort to do so, we might compile a list of the ages of 10,000 randomly selected Chinese citizens. After calculating the average age of these 10,000 persons, can we use this number as our estimate of the average age of the population of China? Considering that our sample is very large, the answer is yes. The principle of large samples tells us that as sample size increases, our estimates derived from those samples more closely resemble the true population characteristics or parameters. Despite this, our confidence in the accuracy of this estimate is still an open question. What might we do to enhance our confidence in the accuracy of our estimate?

One strategy for increasing our confidence in the accuracy of our estimate is to conduct repeated sampling procedures and, on each occasion, identify the average age of persons in each sample. Based on our discussion of the Central Limit Theorem, if a large number of samples are gathered and sample averages calculated, a distribution of these sample averages could be formed. If that distribution were presented graphically, it would resemble a bell-shaped curve or what statisticians refer to as a normal curve. The peak of this distribution indicates the value calculated most frequently in the repeated samples. This value is known as an estimate, or a “good guess” as to the average age of the entire population. In short, as repeated samples are conducted, the value of our estimate approaches the value of the average that would result if we were able to examine the entire population.

PROBABILITY THEORY AND CONFIDENCE INTERVALS

The lessons of probability theory are especially important when constructing confidence intervals from statistical estimates. A confidence interval is a probability boundary for a population parameter estimated from a sample. Furthermore, Manheim and Rich describe a confidence interval as “an indicator of the accuracy with which a population parameter can be predicted from a sample statistic stated in terms of the range of values above or below the sample statistic the population parameter is likely to fall.”^[4] Many types of confidence intervals can be estimated from a sample. A confidence interval may be estimated for a population mean, a difference between means, a population proportion, a difference between population proportions, a population variance, a population standard deviation, a difference between two population standard deviations, a correlation coefficient, or a regression coefficient. This list is not exhaustive.

Confidence refers to the degree of certainty that the estimated interval contains the population parameter. Confidence refers to the degree of certainty that a certain percentage of similar statistics from repeated samples would fall within the interval. This is nearly the same as saying that, within a certain specified probability, the population parameter would fall within this interval. Confidence intervals are generally two-sided and constructed as follows:

$$V \pm Z(SE)$$

where V = the value of the statistic computed from the sample; Z = the value of the multiplier for a stated probability level obtained from a probability distribution; and SE = the standard error of the estimate.

The following example reflects the application of probability theory to the concept of confidence intervals. Suppose the director of the parks and recreation department of a large municipality decides to perform a survey to discover whether a proposal for expanded hours for the municipality's indoor recreation facilities is justified by community demand. The administrator plans to propose additional evening and weekend hours, but the expanded hours will require additional staffing and facilities expenses. The administrator decides to survey regular or occasional users of the facilities. The administrator will make the proposal if it is believed that at least 30% of facilities users indicate that they would at least be "quite likely" use the facilities during the proposed new hours.

The administrator's population of interest is users of the city's indoor recreation facilities. The director has made the decision to ignore non-users by assuming that increased hours will not have a significant impact on their behavior. The administrator wants to have some idea of whether expanded hours can be justified by increased demand. Because of time constraints, the administrator is not able to survey all users but chooses a sample of users instead. Based on the sample information, the administrator will infer the level of demand in the overall population.

Assume that a sample of 200 users is obtained. Those responding to the survey are asked how likely they would be to use the facility during the expanded night and weekend hours. The question is answered on a 5-point scale with scores of 4 and 5 representing respective choices of "quite likely" and "extremely likely." Thirty-seven percent of the users answering the survey fall into those two categories.

Based on the survey evidence, may the administrator infer that at least 30% of users would use the facility during expanded hours? The administrator must employ probability theory to test these hypotheses and estimate a confidence interval for the expected percentages in the overall municipal population.

Suppose the administrator uses the confidence interval approach to estimate a population proportion. We know that sample proportions will vary randomly around the true population proportion from sample to sample. Sample proportions of this type may be modeled as random variables following a binomial probability distribution. The normal distribution may also be used to approximate the binomial distribution of sample proportions from this population of samples.

Most likely, the director would use a computer statistical package to estimate the confidence interval of the population proportion from the sample estimate. Assume that a 95% confidence interval is chosen by the director as an adequate level of statistical confidence to estimate the population proportion. If the confidence interval includes values less than 30%, the director will be unable to conclude within the desired level of confidence that more than 30% of the overall user population will be "very likely" to use the facilities during expanded hours. However, if the confidence interval does not include values less than 30%, the director will infer that the demand level in the overall population is met.

The obtained 95% confidence interval using the exact binomial probability distribution is (0.276, 0.472). Although the confidence interval includes estimates of the population proportion well in excess of the 30% demand-level target, the interval also includes estimated values less than 30%. The administrator cannot conclude within the desired level of confidence that the demand level in the population is greater than 30%. In fact, the demand level could be as low as 27.6% and one could obtain a sample proportion as high as 37% from a sample of 100 with a small, but not extremely unlikely frequency.

Note that the confidence interval obtained in our example is rather large (almost twenty percentage points). The director could obtain a narrower confidence interval by either sampling a larger number of individuals or choosing a lower level of confidence. The obtained confidence interval if the sample is expanded to 200 individuals and the same proportion of 37% fall into the top two categories of responses is (0.303, 0.441). Here, the lower bound of the confidence interval exceeds the target figure, and the director may be confident that sufficient demand exists in the user community. If a 90% confidence interval had been chosen for the sample of 100, the estimated interval would be (0.289, 0.457). Choosing the later alternative is easier but does not shorten the confidence interval as much and still does not allow the director to conclude that the population demand is in excess of 30%. In addition, choosing a lower level of confidence increases the researcher's chances of making a mistake in inferring a population proportion from a sample estimate.

CONCLUSION

Probability theory is the foundation of statistical inference, a necessary process to decision making in public policy. Statistical inference requires knowledge of a set of concepts and tools springing from probability theory, including an understanding of the normal probability distribution, other important probability distributions, the concepts of hypothesis testing, and construction of confidence intervals for estimates of population parameters. Public administrators use all these basic concepts when engaging in scholarly pursuits and making resources allocation decisions. The application of probability theory has played a central role in modern and contemporary efforts to rationalize the tasks of governing. Probability theory is, and

will remain, an important source of guiding principles in the study and practice of administration.

REFERENCES

1. Anderson, T.R.; Zelditch, M. *A Basic Course in Statistics with Sociological Applications*, 3rd Ed.; Holt, Rinehart and Winston, Inc.: New York, 1975; 225 pp.
2. Welch, S.; Comer, J. *Quantitative Methods for Public Administration: Techniques and Applications*, 3rd Ed.; Harcourt College Publishers: Orlando, FL, 2001; 168 pp.
3. Knoke, D.; Bohrnstedt, G.W. *Basic Social Statistics*; F.E. Peacock Publishers, Inc., 1991; pp. 92–98, 163.
4. Manheim, J.B.; Rich, R.C. *Empirical Political Analysis: Research Methods in Political Science*, 4th Ed.; Longman Publishers: White Plains, NY, 1995; 441 pp.

Procurement

Charles K. Coe

*Department of Political Science and Public Administration, North Carolina State University,
Raleigh, North Carolina, U.S.A.*

INTRODUCTION

State and local governments annually purchase over \$850 billion of goods and services. Because of the large sums involved, procurement is a fundamental aspect of governmental operations. Professional policies and procedures reduce costs; conversely, shoddy practices cause waste and excessive costs. Well-managed procurement programs should be soundly grounded with laws, regulations, and ethical standards; professionally acquire goods and services; securely operate a fixed-asset system; follow sound inventory control and management procedures; and be regularly evaluated. Let us discuss each requisite.

A SOUND FOUNDATION

A purchasing materials management program must have a sound foundation that includes well-trained personnel; laws and regulations adhering to federal government directives, basic contract law, the Uniform Commercial Code, state law, and local ordinances; a system centralized in a single office; an operating manual; and clearly defined ethical standards. Governments should inform the public and vendors about their policies and procedures in writing and on their website.

State laws direct most, if not all in some states, procedures that local jurisdictions must follow in purchasing goods and services. Purchasing officers should, therefore, be intimately familiar with their state's requirements, paying particular attention to provisions covering acceptance of bids and proposals, preparing specifications, advertising of the letting of contracts, withdrawing bids, and executing contracts.

The American Bar Association has published two guides for preparing a local purchasing ordinance or code: *The Model Procurement Code for State and Local Government* and *The Model Code for Local Governments* that is intended to meet the needs of smaller governments.

ROUTINE PURCHASING

Most goods purchased follow the basic purchasing cycle that includes preparing a purchase requisition;

determining if bids are required, and if so, preparing an invitation for bids; evaluating and awarding bids; preparing and administering a purchase order; receiving goods; making payment; and evaluating vendors' performance. Contracts are awarded to the lowest responsible and responsive bidder. A "responsible" bidder must be able to perform the contract specifications, meeting the material terms of the invitation for bids. Accordingly, bids are usually awarded to the lowest bidder. The predictability of the normal purchasing system may give policy makers the false impression everything works smoothly. To the contrary, the purchasing system, like any other system, can always be improved.

In particular, administrators should vigilantly evaluate bids and purchase orders for their completeness; guard against unwarranted changes proposed by vendors; and carefully inspect delivered goods to insure that they meet contractual requirements. After a vendor has received a specified number of complaints, the government should take corrective action.

The purchase description, called a specification, is at the heart of the procurement process, informing the vendor what is required and providing the basis for awarding the purchase order or contract. The purchasing office prepares most specifications; however, agencies may take the initiative. In this case, the purchasing office should reserve the right to review specifications that the agency prepares. Purchasing officials should avail themselves of all resources in preparing specifications. National and professional societies have standards for some commodities. Likewise, federal, state, and local governments have a wealth of expertise that is one call away. Other helpful sources for specification writing and purchasing are found in [Table 1](#).

Purchasing and materials management should be computerized if possible. To ensure that vendor bidding is competitive and track vendor performance, governments should register vendors in a database that maintains vendor financial data (such as tax status), tracks vendor performance, and is the basis for making payments. A contracting module generates bid lists, tracks purchase orders, and assists in preparing, awarding, and monitoring contracts.

The Internet likewise streamlines purchasing. Suppliers offer their catalogs on the Internet, governments send bid notices and solicitations electronically, and a

Table 1 Resources for purchasing assistance

Source	Address	State	Telephone
National Institute of Governmental Purchasing	11800 Sunrise Valley Drive, Suite 1050	Reston, VA	703/715-9400
National Purchasing Institute	7910 Woodmont Ave.	Bethesda, MD 20814	800/695-1025
National Association of Purchasing Management	2055 East Centennial Circle	Tempe, AZ 85284-1898	800/695-1025

wide variety of purchases are made online. The Internet increases the pool of bidders, while saving vendors and governments money by reducing transaction costs. Electronic commerce is paperless and integrates cataloging, ordering, and exchanging financial transactions.

Using the Internet, governments can find help to prepare specifications, understand recommended purchasing procedures, jointly purchase with other governments, and train purchasing officers. Particularly useful are the following websites:

- The National Association of State Purchasing Officials (www.naspo.org).^[1]
- The National Institute of Governmental Purchasing (www.nigp.org).^[2]
- The National Association of Purchasing Management (www.napm.org).^[3]
- The National Contract Management Association (www.ncmahq.org).^[4]
- Procurement “jump” sites (www.fedmarket.com/statejump.html#35).^[5]
- Federal Electronic Commerce Program (www.ec.fed.gov).^[6]

Similarly, using credit cards to purchase small items results in savings. To establish a credit card system, purchasing officers should first conduct a cost analysis to determine which situations and which price ranges warrant their use. Strict accounting controls should be enforced to avoid misuse. Credit card advantages include:

- Instead of paying scores of vendors, only one credit card company is paid.
- Payments are scheduled to optimize payment, thereby increasing the amount of time to invest funds.
- Most credit card companies provide up-to-date reporting and query capabilities, which are more accessible than most governmental accounting systems.

NON-ROUTINE PURCHASING

Purchases falling outside of the normal purchasing cycle include purchasing services, cooperatively purchasing with other governments, volume purchasing,

emergency purchasing, sole-source purchasing, petty cash fund purchasing, and environmental purchasing. Of these, the most challenging, technical, and complex is purchasing services, which is more complicated than routine purchasing for two reasons. First, contracts for services are awarded not principally on lowest cost but on firms’ experience. Second, some services are technical, and contracts for them must be negotiated. There are four types of services: professional services (e.g., legal), operating services (e.g., solid waste collection), technical services (e.g., MIS installation and software), management services (e.g., motor equipment maintenance and repair), and client services (e.g., halfway houses and counseling). Defining the scope of technical services, such as legal, engineering, management information service (MIS) software and hardware installation, and auditing, is challenging. For more technical services, a work team—typically comprised of the contract manager, a departmental representative, and representatives from purchasing, budget, and legal—prepares the scope of work specifying the contractor’s and government’s responsibilities and general contractual conditions including payment procedures, insurance requirements, and bonding requirements. Contracts are then negotiated and require careful governmental oversight.

Looking at other non-routine types of purchasing, cooperatively purchasing with other governments saves money due to larger economies of scale. There are three types of cooperative purchasing: two local units joining together to purchase a good or service (e.g., a city jointly purchasing cleaning supplies with a school system); a joint association of two or more governmental units (e.g., all states permit local units to purchase goods on their state contracts); and a new legal entity, such as a regional purchasing council. Volume purchasing is made for repeated purchases. Rather than repeatedly issue purchase orders, one purchase order for an item is issued for a specified period, thereby reducing costs and eliminating inventories. Sole-source purchasing should only be done after a good faith review determining that only one source is available. Reasons for sole-source purchases should be documented in writing. A petty cash fund for small amounts, typically from \$10 to \$100, eliminates purchase orders, thereby saving time and money. Petty

cash funds need strict accounting controls, ironically, because of the small amounts involved, which sometimes cause employees to forget to account properly for withdrawals. Finally, governments should make environmentally responsible purchasing decisions, including, where cost effective, buying recycled products, setting recycling content standards for paper and paper products, not buying disposable products, not buying batteries or other mercury-containing products when alternatives are available, purchasing energy-efficient lightbulbs and fixtures, and buying non-toxic, biodegradable cleaning products.

FIXED-ASSET INVENTORY SYSTEMS

A fixed-asset inventory system that maintains accounting controls over buildings, land, and equipment is a useful capital planning tool, stops tools and equipment from “walking” off with employees, and informs managers about the amount of capital that should be protected (through insurance or self-insurance) against risks of theft, fire, or vandalism. A fixed asset has a useful life of at least one year. The government must decide the capitalization limit, which is the dollar limit over which fixed assets must be accounted for and recorded. Two competing considerations govern the decision: 1) the higher the limit, the fewer the assets, and the less cost of creating and maintaining the system; but 2) the less accounting control is exercised.

The Governmental Accounting Standards Board (GASB) Cod. states that fixed assets should be valued at their initial cost (Cod. Section 1400). However, in creating a system, governments likely have fixed assets with no record of their initial costs. In which case, GASB permits using estimated historical costs. Next, the government should form and train a team to conduct the initial inventory, tagging physical assets and accounting for depreciation. Facility assets should be depreciated on a composite straight-line basis at a 2% annual rate. Other assets should be depreciated at the following rates: furniture and office equipment (10%); maintenance and construction equipment (12%); medium and heavy motor vehicles (30%); and data-processing equipment (20%). Finally, the government should fix personal responsibility for assets, prescreening purchases to see if they are fixed assets, and following procedures for additions, changes, and disposals.

INVENTORY CONTROL SYSTEMS

Whether to have an inventory is a benefit–cost decision. Keeping supplies and equipment in an inventory reduces downtime to get parts, decreases theft due to having accounting control, decreases the paperwork of making numerous purchases, and can give

management information about the costs of supplies on particular jobs. On the other hand, inventories have costs including staffing the facility, carrying costs including obsolescence and deterioration, investment interest lost due to goods tied up in inventory, insurance costs, and equipment (e.g., forklifts, shelves, and pallets).

When costs exceed benefits, governments should instead volume purchase using blanket contracts. If benefits exceed costs, creating a system entails fixing responsibility for creating and operating the system, deciding what to stock, deciding how to stock (e.g., two bins, periodic inventory, perpetual inventory), purchasing software, adopting a commodity classification code identifying items and cross-referencing them against manufacturers’ parts numbers and the user’s parts numbers, designing a stores plan, and adopting an inventory procedures manual.

EVALUATING THE PURCHASING PROGRAM

The purchasing manager should routinely evaluate the effectiveness of the purchasing program. The customers of a purchasing department are the operating departments. Accordingly, the purchasing departments should regularly survey them to assess satisfaction levels and solicit recommendations for improvement. Likewise, purchasing agents should annually conduct customer satisfaction surveys with a representative sample of vendors who do business with the government. Agencies and vendors are concerned with the amount of time taken from the time of the requisition to the time of awarding a contract or completing a purchase order. Efficient purchasing departments should be able to issue routine purchase orders in a matter of a few days. Contracts requiring formal bids take more time, of course.

REFERENCES

1. The National Association of State Purchasing Officials (www.naspo.org).
2. The National Institute of Governmental Purchasing (www.nigp.org).
3. The National Association of Purchasing Management (www.napm.org).
4. The National Contract Management Association (www.ncmahq.org).
5. Procurement “jump” sites. (www.fedmarket.com/statejump.html#35).
6. Federal Electronic Commerce Program (www.ec.fed.gov).

BIBLIOGRAPHY

1. *A Guide for Vendors 7th Ed.: How to do Business with the States*; The National Association of State Purchasing Officials: Lexington, KY, 1997.

2. Burt, D.N.; Pinkerton, R.L. *A Purchasing Manager's Guide to: Strategic Proactive Procurement*; AMACOM: New York, 1996.
3. Coe, C. *Purchasing and Materials Management*; Sheshunoff Information Services, Inc.: Austin, 1998.
4. Harney, D.F. *Service Contracting: A Local Government Guide*; International City/County Management Association, 1992.
5. *Inventory and Warehouse Management*; The National Institute of Governmental Purchasing: Reston, VA, 1991.
6. Lallatin, C.S. *Public Purchasing and Materials Management*; The National Institute of Governmental Purchasing, Inc.: Reston, VA, 1995.
7. MacManus, S.A. *Doing Business with Government: Federal State, Local and Foreign Government Purchasing Practices for Every Business and Public Institution*; Paragon House: New York, 1992.
8. Short, J. *The Contract Cookbook for Purchase of Services*, 2nd Ed.; The Council of State Governments: Lexington, KY, 1990.
9. Short, J. *Issues in Public Purchasing: A Guidebook for Policymakers*; The Council of State Governments: Lexington, KY, 1992.
10. *State and Local Government Purchasing: Principles and Practices*, 5th Ed.; National Association of State Purchasing Officials: Lexington, KY, 1997; Vol. 1.
11. *Survey of State and Local Government Purchasing Practices*, 5th Ed.; National Association of State Purchasing Officials: Lexington, KY, 2001; Vol. 2.

Productivity and Effectiveness

Patria de Lancer Julnes

Department of Political Science, Utah State University, Logan, Utah, U.S.A.

INTRODUCTION

Productivity and effectiveness are important concepts for public administrators. However, there seems to be no agreed-upon definition of the terms or clarity of the relationship between them. These issues are addressed here. In addition, this article includes discussions of the history of productivity improvement efforts, the measurement of productivity, and some strategies for improving productivity.

DEFINING PUBLIC PRODUCTIVITY AND EFFECTIVENESS

Productivity

Although the concern for productivity has a long-standing tradition, there has been some ambiguity as to what it means. According to standard usage, productivity refers to the level of output—amount of services delivered, given a certain level of inputs. Defined this way, productivity is calculated as a ratio that attempts to answer the more general question: how well are resources used? If an agency increases outputs without requiring more inputs or if it maintains the same level of outputs after reducing inputs, then that agency has increased its productivity. Economists call this technical efficiency.

However, this simple definition of productivity does not capture all the nuances of common usage of productivity as applied to producing an output or delivering a service. Productivity is affected by many issues; accordingly, this has made it necessary to find different ways of thinking about it. For example, some think of public service productivity not as a simple ratio but as the responsiveness to the needs, desires, and resources of the community.^[1] This emphasis on responsiveness starts to overlap with effectiveness, the other topic discussed in this article.

Effectiveness

The term “effectiveness” has been emphasized by those interested in more than technical efficiency—not just outputs, but desired outcomes. In common parlance, effectiveness in the public sector refers to the

overall impact and outcome of policies or programs. An effective program achieves the desired outcomes, which typically follow from the program’s mission, goals, and objectives without creating other problems. This definition implies causality, and some argue that assessing something as abstract as a causal relationship is outside the realm of the day-to-day monitoring of operations of an agency. Given this concern, a definition of effectiveness that has been argued to be more appropriate for public productivity is implementation effectiveness, which refers to the extent a program is implemented in a way that produces and delivers the *right* amount and quality of services.^[2] Others have gone further, defining an agency’s effectiveness in terms of the relationship between internal processes and mission achievement. They argued that an agency is effective if it “does well that which it is supposed to do, whether people in the agency work hard and well, whether the actions and procedures of the agency and its members are well suited to achieving its mission, and whether the agency actually achieves its mission.”^[3]

The emphasis on effectiveness is an important contribution to the field of public productivity. But note that, as with productivity, the term effectiveness also has some ambiguities. Although some define it in terms of outcomes and impacts, others argue for output- and process-based definitions.

Linkage Between the Two

In focusing on output and quality, implementation effectiveness provides a complementary answer to the question of the extent to which resources are being used appropriately. For example, if an agency becomes more productive—that is, delivers more services (output) on a weekly basis with the same number of case workers (input)—but does so at the expense of quality and poorly completed paper work, this is not a good outcome and thus resources are not being used well. This dictates that the managers need to pay attention to the process.

Consider the following question as a second example. What does an increase in the number of arrests (outputs) by officers in a police precinct tell us about the success of a newly implemented productivity improvement program (e.g., a new promotion system)? Based on these numbers alone, one would

conclude that the productivity-enhancing system is good and effective in motivating the police officers to work harder. However, what if one consequence of the policy is that innocent citizens are being arrested? This will decrease the conviction rate and decrease citizens' confidence of the police. In such scenario, this productivity-enhancing program is not effective. This is a simplistic example, but it helps to illustrate that even if productivity is increased without compromising the quality of arrests, for example, one needs to know that there is not a negative impact on society. Thus, focusing on output, quality, or technical efficiency alone is not enough. Productivity and effectiveness must go hand in hand. However, as discussed in the next section, this conceptual partnership has proved problematic.

THE PRODUCTIVITY MOVEMENT

At its inception as a field, the study of public productivity was based on a rational/technocratic approach emphasizing technical efficiency. The study of public productivity has been traced back to an interest in efficiency as the vehicle to wealth.^[4] Since the 1800s, public and private productivity have been quantified in the United States, but more attention has been paid to productivity since World War II. In the United States, an increased interest in efficiency in the public sector was stressed in 1887, when, in his seminal essay, Woodrow Wilson supported a civil service reform that emphasized professionalism in public administration to change the corrupt behavior of public officials.^[5] Although not directly emphasizing efficiency, the belief at the time was that politics had to be separated from administration because the intervention of politics inhibited good administration. Political immorality precluded the development of a professional, accountable, and well-trained public administration, which, in turn, led to inefficient public administration.

Such ideas served as the foundation of a desire for "good" government, and the adoption of administrative practices with procedures and measurement techniques to identify and increase the productivity of workers. From the late 1800s to 1940, economists used productivity to describe the link between resources and products; however, efficiency and economy were more popular terms. The comparison of inputs to outputs is the basis of applications such as scientific management.^[6] Nonetheless, this emphasis on efficiency was questioned by many because it did not pay enough attention to satisfaction or the benefits or results produced.^[7] That is, the definition of efficiency as the accomplishment of the work with the least expenditures of manpower and materials neglected the concept of effectiveness, or allocative efficiency, as it is known in the economics literature.

From the 1940s to the 1970s, the period Bouckaert called "government by administrators," the focus of government was to control expenses through the budget process.^[7] Although productivity, doing more with less, remained of interest, improving decision making through the application of positivist, scientific techniques were the goal.^[2] It was during this era that the planning programming budgeting system and zero base budgeting were developed, and suggested for adoption throughout government.

Furthermore, in 1962 numerous studies were commissioned under the Bureau of the Budget to assess the feasibility and usefulness of productivity measurement. But nothing was implemented until after the 1970s.^[7] Effectiveness, and particularly institutional effectiveness, became a core value in the field of public administration because of the work of leading scholars of this period.^[8]

In the 20-year span between 1970 and 1990, triggered by the growth and cost of government programs and general problems in the economy, a heightened cynicism on the part of the public toward government led to the development of measures of public productivity and appropriate techniques for improving it. As explained by Schein, the public became cynical about the money spent by public organizations on social services; tax dollars spent did not seem to be providing an appropriate level of return as judged by the apparent failure of programs to solve social problems.^[9] Improving government productivity—more yield per dollar spent—became the mantra.

In 1970, under President Nixon, the National Commission on Productivity was created (later changed to National Center for Productivity and Quality of Working Life). One of the goals of this commission was to improve government productivity. Similar efforts followed at the state and local levels. Academia also contributed to the productivity improvement effort. One of the most recognized leaders in the field is Marc Holzer who in 1972 created the National Center for Public Productivity, NCPP (<http://www.ncpp.us/index.php>). The NCPP contributes to the improvement of productivity in the public sector through research and public service. Two major publications of the center include the journal *Public Performance and Management Review*, and the *Public Productivity Handbook* (1st and 2nd editions).

Government productivity improvement efforts of the 1970s were followed by somewhat opposite efforts with similar goals, increasing efficiency by reducing government's involvement in delivering services. This reduction was believed to be best accomplished through privatization and the contracting out of government functions. The assumption was that the private sector could do a better job than the public sector and still deliver services at a lower cost. At

the same time, public administration scholars became more concerned with the ethical, legal, and economic constraints on government. Effectiveness per se was folded in these three issues.

Embedded in the assumption held by many Americans that a large portion of public funds are wasted^[10] while the private sector continues to improve quality, initiatives to improve public productivity now focus on increasing constituents' satisfaction with the way tasks are performed.^[11] The period beginning in the 1990s has seen an expansion of the meaning of productivity to include such things as results or outcomes and quality of services. Governments now attempt to show their constituents what they are getting for their tax dollars, how efficiently and effectively their tax dollars are spent, and how expenditures benefit their lives.^[12] Performance measurement has been touted as the vehicle for accomplishing these public goals.

MEASUREMENT AND PRODUCTIVITY

At the core of productivity improvement is performance measurement. The basic reason for conducting performance measurement is to strengthen government's resource allocation and administrative decision-making capabilities.^[13] The importance of performance measures derives from their empirical basis for decision making.^[4] Measures of productivity can be used to maintain internal and external accountability; to provide a rationale for the development of goals and objectives, resource allocation, and monitoring results; and to support modifying plans to enhance performance.^[1,14]

Because performance measurement is considered an important aspect of governmental management, several organizations including the Governmental Accounting Standards Board (GASB), the National Academy of Public Administration, and the International City Managers Association, are encouraging and working with governments to experiment with them.^[15] Notwithstanding these efforts, establishing a performance measurement system often faces a wide range of obstacles, ranging from individuals' resistance to measurement to an organization's lack of technical or analytical skills.^[16]

Furthermore, the establishment of a performance measurement system can have serious drawbacks if it is not done properly. In particular, a system is not good if it does not provide appropriate information to decision makers, and it can hurt an organization if it misrepresents, misleads, or introduces perverse behavioral incentives.^[13]

The GASB has suggested that public organizations assess performance in ways that go beyond the relationship of outputs to inputs. It recommends that organizations use various measures including those

that relate their efforts to their accomplishments. The GASB adds to the measures of inputs (amount of resources used) and outputs (amount of workload accomplished) measures of outcome and efficiency.

STRATEGIES FOR IMPROVING PRODUCTIVITY

The amount of available resources and their effective use have been central concerns in productivity improvement efforts. However, if public organizations are to improve their productivity as defined in the broader sense, they need to employ a management framework that allows them to develop and reinforce their production capacity. In discussing award-winning government agencies, Holzer and Callahan concluded that these organizations tended to^[12]

- 1) apply quality management principles; 2) use measurement as a decision-making tool; 3) invest in human resource development and organizational learning; 4) adapt new technologies; and 5) develop partnerships with the private sector, with other governmental and non-profit agencies, and between management and labor.

These recommendations appear reasonable; the question, then, is how to put such framework into practice? Rosen offered some specific guidelines for making productivity improvement decisions that come from lessons learned in practice.^[2] First, there is the realization that productivity improvement takes time. Therefore, strategies must be thought through thoroughly and carefully. Reforms should be kept simple, and projects that present the least risk but have high potential of success should be started with. Pilot projects should be conducted and evaluation should be ongoing. This, however, must be accompanied by a participative orientation and open communication. Second, it is important to anticipate resistance and to develop the appropriate strategies for dealing with it. Those strategies would include attaining top-level support, maintaining open discussions with stakeholders, and building rewards into the productivity improvement plan. Third, you have to choose a problem that most people see as a problem. Fourth, the solution must be the right solution in that it should be a proven approach to the problem. At the same time, it must be feasible and timely. Fifth, the goals must be clearly defined and method for measuring success devised. Finally, the productivity improvement program must be thought of in terms of its effect on the organization as a whole.

As highlighted above, productivity improvement efforts must be conducted in a systematic way. Approaches that have been successful elsewhere must

be adjusted to the local context. Also, it is important to recognize that productivity improvement must be a team effort. Productivity improvement efforts are more likely to succeed if they have the support from individuals at all levels of the organization, including elected officials.

CONCLUSION

Traditionally, productivity improvement efforts focused more on quantity (outputs) than on quality and impact. It's been argued here that such an approach is needlessly simplistic. To be more useful, productivity improvement efforts must also involve efforts to improve effectiveness. But as was discussed here, the practical and conceptual connection between productivity and effectiveness has been problematic. To overcome this impasse, it is recommended that organizations adopt a systematic managerial framework based on a broadened conceptualization of productivity that includes measures of effectiveness.

REFERENCES

1. Epstein, P. Measuring the performance of public service. In *Public Productivity Handbook*, 1st Edn.; Holzer, M., Ed.; Marcel Dekker: New York, 1992; 161–194.
2. Rosen, E.D. *Improving Public Sector Productivity. Concepts and Practice*; Sage Publications: Newbury Park, CA, 1993.
3. Rainey, H.; Steinbauer, P. Galloping elephants: developing elements of a theory of effective government organizations. *J. Public Admin. Res. Theory* **1999**, *9* (1), 1–32.
4. Holzer, M. Mastering public productivity improvement. In *Public Productivity Handbook*, 1st Edn.; Holzer, M., Ed.; Marcel Dekker: New York, 1992; 1–11.
5. Wilson, W. Study of administration. *Polit. Sci. Quart.* **1887**, *June*, 197–222.
6. Martin, D.W. The management classics and public productivity. In *Public Productivity Handbook*, 1st Edn.; Holzer, M., Ed.; Marcel Dekker: New York, 1992; 47–62.
7. Bouckaert, G. Public productivity in retrospective. In *Public Productivity Handbook*, 1st Edn.; Holzer, M., Ed.; Marcel Dekker: New York, 1992; 5–46.
8. Stillman, R. *Public Administration. Concepts and Cases*; Houghton Mifflin Company: Boston, MA, 2000.
9. Shein, E. Culture: the missing concept in organization studies. *Admin. Sci. Quart.* **1996**, *41*, 229–240.
10. Fischer, R.J. An overview of performance measurement. *Public Manager* **1994**, *7*, S2–S8.
11. Romzek, B.; Dubnick, M. Accountability. In *International Encyclopedia of Public Policy and Administration*; Shafriz, J.M., Ed.; Westview Press: New York, 1998; 6–11.
12. Holzer, M.; Callahan, K. *Government at Work: Best Practices and Model Programs*; Sage: Thousand Oaks, CA, 1998; 27 pp.
13. Kravchuk, R.S.; Schack, R.W. Designing effective performance measurement systems under the Government Performance and Results Act of 1993. *Public Admin. Rev.* **1996**, *56* (4), 348–358.
14. Hatry, H.P.; Fountain, J.R., Jr.; Sullivan, J.M.; Kremer, L. *Service Efforts and Accomplishments Reporting: Its Time Has Come*; Governmental Accounting Standards Board (GASB): Connecticut, 1990.
15. Fountain, J. Are state and local governments using performance measures? *PA Times* **1997**, *20* (1), PM 2, PM 8.
16. de Lancer Julnes, P.; Holzer, M. Promoting utilization of performance measures in public organizations: an empirical study of factors affecting adoption and implementation. *Public Admin. Rev.* **2001**, *61* (6), 693–708.

Program Budgeting (PPBS)

Katherine G. Willoughby

*Department of Public Administration and Urban Studies, Georgia State University,
Atlanta, Georgia, U.S.A.*

INTRODUCTION

Program budgeting is a reform of traditional budget practice that is based on a program structure and the systematic analysis of government goals and objectives in order to inform decision making. This method of budgeting requires that costs be associated with the conduct of agency activities, that alternatives to meeting program objectives are generated, and that a multiyear focus on program operations is instituted. Within a program budget framework, “budget material is arranged in such a way as to aid the executive and legislature in understanding the broader policy implications of their decisions.”^[1]

OVERVIEW

Program budgeting received heavy scrutiny in governments in the United States during the mid-twentieth century. It represented a dramatic departure from a traditional focus of budgeting on the objects of expenditure. That is, a program budget focuses on agency outputs and the results of government activities, while a line item budget focuses on the resources necessary for input to government operations. The planning orientation of program budgeting pushed managers and budgeters past traditional roles of prudent stewardship and efficient administration.

TERMINOLOGY

Program budgeting, or planning, programming, and budgeting system (PPBS), has unique characteristics that set it apart from prior budgeting approaches. These include a program versus functional budget structure that emphasizes a more comprehensive approach to understanding government activity. It is at the program category level that the broad goals of government are declared. Goals are distinguished from objectives as being more general—“a desired state of affairs based upon current knowledge and values” and thus timeless.^[2] For example, the goal of the education program as stated in the State of Pennsylvania 2001–2002 budget is “to provide a system of learning

experiences and opportunities that will permit each individual to achieve his or her maximum potential intellectual development.”^[3]

A tiered structure subdivides program categories into subcategories and then elements. Goals become more specific as programs are subdivided and as the mission becomes more directly actionable. By the subcategory or subprogram level, associated objectives require the quantification of program affects “within a given timeframe and employing available resources.”^[4] Whereas goals are idealized, representing a valutive and desired state of affairs, objectives are measurable and must be both realistic and achievable.

The program structure is the foundation for a system that requires strategic planning, a multiyear focus in budget preparation, information gathering, cost assessment, alternative generation, and reporting. Results from the analysis of programs as tiered is delineation of government activities presumably to foster better understanding of the linkages between program goals and objectives, resource input, and the results of government spending. Program budgeting is premised on the notion that systematic collection and analysis of information and reporting about what government does can enhance decision making about what government should do. This approach is a concerted effort to examine program impact (a good indicator of government effort) rather than program input (a less than satisfactory indicator of government effort). A program approach to budgeting that emphasizes the “measuring, monitoring, analyzing” and reporting of government operations helps to focus attention on program performance, outputs, and outcomes.^[5] The planning focus and systematic analysis require both administrators and budgeters to think beyond what happened last year, past an incremental perspective.

HISTORY

Although formal introduction of a program approach to federal budgeting was made with great fanfare in the 1960s, experience with this approach existed prior to that time by public and private sectors both here and abroad. Novick’s compilation of cases of program budgeting system applications and experiments

includes nine countries and the United States, two state governments, one local government, a school district, and two private companies.^[2] Explanations of local government applications with special reference to the Los Angeles and New York City experiences, and mention of efforts in Washington, DC, Dayton, Denver, Nashville, San Diego, Dade County, Florida, and Nassau County, New York, also exist.^[6] General Motors, then Ford, and finally Chrysler automobile companies experimented with program budgeting beginning in the 1920s. Case studies of use by General Electric and the John Hancock Life Insurance are also in evidence.^[2]

Certainly pressure from the progressive movement for more rational (*vis-à-vis* political) administration of public programs and activities underscored the need for system centralization, budget comprehensiveness, and information collection and analysis that is evidenced in a program budget approach. Scientific method, systems, and operations research also served as foundations for program budgeting analyses requirements, at least in ideal form. Early governmental influences to future program budgeting initiative at the federal level include the Taft Commission on Economy and Efficiency of 1912 (argued against a line item budget format) and the Budget and Accounting Act of 1921 (created the federal executive budget process, a central budget office, and analytical support via the General Accounting Office).

Government fiscal policy also influenced the budget reforms in this century with increasing spending and more complex administrative structures necessitating attention to the budget process. For example, federal spending jumped from \$7 billion in 1934 to \$91 billion by 1944. The Brownlow and Hoover Commissions in the 1930s and 1940s, the Reorganization Act of 1939, and the Budget and Accounting Procedures Act of 1950 reoriented government, arming it with a strengthened executive and more centralized budget process as well as advanced information and data-gathering requirements of public managers.

Early application of program budgeting at the federal level occurred first in military agencies, beginning with the Army in the 1950s, and then in the Department of Defense by 1961. The Johnson administration called for implementation of program budgeting in 25 civilian agencies by 1965. The focus of the initiative was to advance development of program missions and goals, subprogram objectives, and then information and data related to those subprograms and elements to provide systematic cost assessment and cost-benefit analyses. "The objectives of PPBS are to improve the chances that tax dollars will be allocated rationally and will have maximum effect in achieving stated purposes. These objectives require that effects can be measured."^[7]

Although much of the analytical research associated with the program approach, especially trend analysis, was sustained in military programs and in those related to welfare and housing well into the 1970s, program budgeting in practice died out fairly quickly. Controversy related to the war in Vietnam and Watergate, then subsequent stagflation and the energy crisis diverted the attention of the American public and government officials from a concentration on rational means for determining spending. The Reagan administration presented somewhat of a death knell on rational, comprehensive, and top-down assessment of what government should be doing, given its primary agenda of cutting both the size and presence of the federal government.

LASTING EFFECTS

It is difficult to calculate the success or failure of any given program budgeting system. No two systems are identical, and even application in one government is likely to be inconsistent across departments and divisions. Systems are rarely comprehensive. Realistically most budget reforms, program budgeting included, can lay little claim to effecting dramatic change in actual budgeted amounts or proportions. Yet, there is no doubt that program budgeting has left lasting impressions that flavor current budgeting initiatives. For example, strong components of modern performance-based budgeting reforms in the United States include requirements of budget justification, management improvement, and program success that are inherent to a program approach. What is different today with performance-based initiatives is an "enhanced emphasis in public organizations on quality, variety, customization, convenience, and timeliness of services and programs over and above economic efficiency concerns."^[8] Also, modern public organizations have equity concerns versus the cost assessments that overshadowed much of the analysis generated within the program budget format.

Today, the federal government maintains an emphasis on performance measurement and reporting begun in earnest during the first Clinton administration. The Government Performance and Results Act of 1993 requires performance measurement and tracking in federal agencies and established a schedule for performance reporting by these agencies. This system seemingly extends a strictly program approach by incorporating strategic planning, multiyear budgeting, and goal determination, along with more specific measurement development, information gathering and data analysis, and results assessment and reporting.

At the state level, most governments have some type of performance-based budgeting process formulated, if not completely instituted. Still, nine states characterize their budget approach as solely programmatic (Idaho, Illinois, Louisiana, Maine, Maryland, Massachusetts, Nevada, New Jersey, and Pennsylvania), while 13 states characterize their approach as both program and incremental. Nineteen states include program budgeting along with at least one or two other types of approaches, such as zero based or performance. Finally, nine states do not include program budgeting as an approach currently in use by their government (Alaska, Colorado, Georgia, Indiana, Missouri, New Hampshire, New Mexico, New York, and Washington).^[1]

Pennsylvania offers perhaps the most enduring example of program budgeting. This was originally an executive-driven initiative by Governor-elect Raymond P. Shafer in 1967 to support “good management.” At that time, eight program categories were created, seven that remain today (Table 1).

Factors supporting budget reform, in general, and program budgeting, specifically, are also enduring. Consideration of the many case assessments of program budgeting initiatives across governments in the United States yields several conclusions about the factors that can facilitate successful implementation. These include:

- The process must recognize and even encompass the political environment.
- There must be sustained leadership for the initiative, regardless of how prescribed.

Table 1 Program categories in the Pennsylvania state budget

Began with the following programs ^a	Today there remain these programs ^b
• Direction and supportive services • Protection of persons and property • Health-physical and mental well-being • Intellectual development and education • Social development • Economic development and income maintenance • Transportation and communication • Recreation and cultural enrichment	• Direction and supportive services • Protection of persons and property • Health and human services • Education • Economic development • Transportation and communication • Recreation and cultural enrichment

^aCommonwealth of Pennsylvania, Office of Administration, Bureau of the Budget, *The Budget Process in Pennsylvania*, December 1970.

^bCommonwealth of Pennsylvania, Office of the Budget, *The Budget Process in Pennsylvania*, May 1999. (These seven programs, as named, start with the 1988–1989 fiscal year.)

- There should be staggered introduction of the system to agencies.
- There must be ownership of the new/changed system via involvement up and down the hierarchy.
- There must be concerted application of resources:
 - There may be costs associated with the connection of the program structure with the budgeting and accounting system,
 - There will be costs associated with even minimal training of administration and staff regarding system requirements.
- There must be a realistic time frame for institutionalization of the system.
- Any reform is a continually evolving process that must allow for some flexibility to remain viable.

Perhaps the greatest “gift” related to the introduction of program budgeting is its requirement of systematic application of analysis to the study and interpretation of budgets. “Without accurate, comparable, and informative data, decision making will perforce continue to be based on guesswork and intuition.”^[7] Today, the plethora of performance-based budgeting initiatives, as well as advancement of accounting standards at all levels of government, can only enhance the capabilities of government officials for more rational budgeting. Of course, nothing will or should displace the political aspects related to public spending. Rather, the concern should be with providing an environment that supports the presentation and use of clearly formulated information and analysis related to government operations. “Reporting information on the results of government programs in government financial statements can improve the capacity to create such information in addition to developing a demand for it.”^[9]

REFERENCES

1. NASBO. *Budget Processes in the States*; October, 1999; <http://www.nasbo.org/Publications/PDFs/bupro99.pdf> (accessed September 2001).
2. Novick, D., Ed.; *Current Practice in Program Budgeting (PPBS)*; Crane, Russak & Co., 1973.
3. Commonwealth of Pennsylvania, 2001–2002. Governor’s executive budget. February 2001.
4. Mowitz, R.J. Pennsylvania’s Planning, Programming, Budgeting System. In *Current Practice in Program Budgeting (PPBS)*; Novick, D., Ed.; Crane, Russak & Co., 1973; 169–180, Chapter 21.
5. Lee, R.D., Jr.; Johnson, R.W. *Public Budgeting Systems*, 6th Ed.; Aspen Publications, 1998; 171 pp.
6. Hirsch, W.Z.; Sonenblum, S.; Teeple, R.K. *Local Government Program Budgeting: Theory and Practice*; Praeger Publishers, 1974.
7. Hatry, H.P.; Hirsch, W.Z. *Program Budgeting in State and Local Governments: The Practitioner’s View*;

University of California, Institute of Government and Public Affairs, 1971; 2 pp.

8. Willoughby, K.; Melkers, J. Performance Budgeting in the States. In *Quicker, Better, Cheaper: Managing Performance in American Government*; Forsythe, D.W., Ed.; The Rockefeller Institute Press, 2001; 335–364, Chapter 13.
9. Joyce, P.G. The reiterative nature of budget reform: Is there anything new in federal budgeting?. *Public Budg. Finance* **1993**, *13* (3), 46.
8. Lee, R.D., Jr.; Johnson, R.W. *Public Budgeting Systems*, 6th Ed.; Aspen Publications, 1998.
9. Mosher, F.C. *Program Budgeting: Theory and Practice with Particular Reference to the U.S. Department of Army*; Public Administration Service: Chicago, 1954.
10. Mowitz, R.J. Pennsylvania's Planning, Programming, Budgeting System. In *Current Practice in Program Budgeting (PPBS): Analysis and Case Studies Covering Government and Business*; Novick, D., Ed.; The Rand Corporation, 1973; 169–180, Chapter 21.
11. Mundel, D.S. PPB: clearing the record. In *Discussion Paper Number 6*; Harvard University, Kennedy School of Government, May, 1972.
12. NASBO. *Budget Processes in the States*; October 1999; <http://www.nasbo.org/Publications/PDFs/bupro99.pdf> (accessed September 2001).
13. Novick, D., Ed.; *Current Practice in Program Budgeting (PPBS)*; Crane, Russak and Co., 1973.
14. Schick, A. A death in the bureaucracy: the demise of federal PPB. *Public Adm. Rev.* **1973**, *33*, 146–156.
15. Schick, A. PPB: the view from the states. *State Gov.* **1972**, *45*, 12–18.
16. Schick, A. The road to PPB: the stages of budget reform. *Public Adm. Rev.* **1966**, *26*, 243–258.
17. *The Budget Process in Pennsylvania*; Office of the Budget: The Commonwealth of Pennsylvania, 2001–2002; http://sites.state.pa.us/PA_Exec/Budget/ (accessed September 2001).
18. Willoughby, K.; Melkers, J. Performance Budgeting in the States. In *Quicker, Better, Cheaper: Managing Performance in American Government*; Forsythe, D.W., Ed.; The Rockefeller Institute Press, 2001; 335–364, Chapter 13.

BIBLIOGRAPHY

1. Grossbard, S.I. *PPBS for State and Local Officials*; Research Series No. 15; University of Rhode Island, Bureau of Government Research, 1971.
2. Hatry, H.P.; Hirsch, W.Z. *Program Budgeting in State and Local Governments: The Practitioner's View*; University of California, Institute of Government and Public Affairs, 1971.
3. Hinrichs, H.H.; Taylor, G.M. *Program Budgeting and Benefit-Cost Analysis*; Goodyear Publishing Co., Inc., 1969.
4. Hirsch, W.Z.; Sonenblum, S.; Teeple, R.K. *Local Government Program Budgeting: Theory and Practice*; Praeger Publishers, 1974.
5. Joyce, P.G. The reiterative nature of budget reform: Is there anything new in federal budgeting? *Public Budg. Finance* **1993**, *13* (3), 36–48.
6. Kettl, D.F. *The Global Public Management Revolution*; The Brookings Institution, 2000.
7. Lee, R.D., Jr. The use of program analysis in state budgeting: changes between 1990 and 1995. *Public Budg. Finance* **1997**, *17* (2), 18–36.

Program Evaluation and Review Technique (PERT) and Critical Path Method (CPM)

Bradley J. Best

Department of Political Science, Buena Vista University, Storm Lake, Iowa, U.S.A.

Jeremy R.T. Lewis

Department of Political Science, Huntingdon College, Montgomery, Alabama, U.S.A.

INTRODUCTION

Program evaluation and review technique (PERT) and the critical path method (CPM) are used to plan, coordinate, and control the performance of activities that comprise complex projects. Designed as a tool for executing projects that are unique and non-repetitive, PERT methods typically include the preparation of a network diagram. Constructing a PERT diagram requires a manager to identify the interdependent phases of a project and the correct sequences of activities that lead to completion. Furthermore, when combined with applications of probability theory, PERT may be used to predict the duration of activities and forecast the complete date of a project.

Closely related to PERT, CPM techniques are used to guide project managers in making resource allocation decisions during the life of a project. Critical path method guides managers in identifying the sequence of activities that most directly influence the length of a project. Moreover, CPM enhances a manager's ability to use resources in an efficient way by providing a strategy for determining which activities can be expedited to minimize total project costs. Together, PERT and CPM rationalize the central tasks of project management—forecasting the completion date of projects, while coordinating resource allocations to maximize efficiency and effectiveness.

PERT: ORIGINS AND APPLICATIONS

Firmly rooted in the field of operations research, PERT/CPM shares a common theoretical lineage with the familiar techniques of linear programming and decisional analysis. Each method is dedicated to improving managerial decision making. Devised in the late 1950s by the firm of Booz-Allen Hamilton, PERT was first employed to schedule and control the performance of tasks necessary to construct the U.S. Navy's Polaris missile system. This project required the coordinated work of scores of contractors and

subcontractors. In addition, the large number of system components and the technological complexity of the project demanded a carefully orchestrated plan for construction. Moreover, PERT emerged as a method of sequencing and estimating the duration of activities in complex projects.

Program evaluation and review technique methods are not appropriate for all projects, but are best employed in projects that, for analytical and planning purposes, can be dissected into a series of distinct activities and events. Program evaluation and review technique is useful in planning for projects in which some or all activities must be performed in prescribed sequences. The task of identifying the proper sequencing and precedence relationships in a project requires that managers systematically reduce the project to a collection of distinct activities. When an activity can be performed only after the completion of another activity, there is said to be a precedential or dependent relationship between these tasks. Furthermore, PERT and CPM methods are jointly applicable to projects wherein a delay in the completion of one phase requires some adaptation, such as the dedication of additional resources to a key activity to realize a specified completion date for the entire project.

Program evaluation and review technique methods are gainfully used in projects that have identifiable points of initiation and conclusion. Thus, PERT is most useful in planning and controlling non-repetitive, non-cyclical projects. As Levin and Kirkpatrick observed, repetitive projects, such as the assembly-line manufacture of automobiles or the continuous provision of trash disposal services by a local or county government, are little improved by applications of PERT analysis.^[1] This is not to say that planning and controlling work processes is unnecessary in instances of repetitive projects. Rather, with each repetition data are generated to inform the planning and control functions that managers must perform in overseeing future iterations of production or service delivery. By contrast, managers of nonrepetitive projects must plan, coordinate, schedule, and control activities in the absence of experiences on which to

estimate the time necessary to complete those activities.^[1] Thus, planning for non-cyclical projects requires a system of scheduling and coordinating activities characterized by both multiple precedence relationships and complex streams of uncertainties. Program evaluation and review technique is a network planning technique developed to address these unique managerial challenges.

EVENTS AND ACTIVITIES IN PERT

When conducting PERT analysis, a project is defined as a collection of distinct, yet often interdependent activities that when performed in prescribed sequences attain some specific goal. Projects are typically represented using a network diagram—a visual depiction of the events and activities that link events. Preparation of a network diagram requires managers to both conceptualize an entire project and identify its many distinct activities. Construction of a network diagram calls upon project managers to identify the sequences of activities that will affect the efficient completion of a project.

An event is defined as the point of origin or completion of an activity. Events are depicted on the network diagram using letters either enclosed or adjacent to circles. If PERT methods were used to plan the construction of a large ship, one event on the network diagram may represent the beginning of work on the hull. A subsequent event would represent the completion of the hull and the commencement of the next phase of the project. Moreover, events consume no time and incur no personnel or material costs. Thus, events are included in the network diagram to reveal the boundaries of activities in project design.

Activities are defined as the processes that link events. Unlike events, activities consume time, require personnel, and incur other costs. An example of a simple project, such as the one depicted in Fig. 1, is useful in illustrating the distinction between events and activities in PERT analysis. Presume that our entire project consists of planting a small tree on the campus green and decorating it with small light bulbs. Event 1 is identified as the beginning of the project—the point at which we begin digging a hole suitable for a small tree. If the process of digging the hole is denoted as activity A, event 2 refers to the point at which the hole is complete.

Activities are the principal focus of PERT analysis. Consider the example presented in Fig. 1. Because we

must know when to have the light bulbs available at the site of the newly planted tree, it is essential that a manager produce a reliable estimate of the length of the digging and planting process, labeled activity B. Estimating the time necessary to complete activity B requires an elementary knowledge of probability theory.

APPLICATIONS OF PROBABILITY THEORY

A principal task of the project manager is to accurately forecast the completion date of a project. The accuracy with which project completion dates are identified is contingent on the planner's ability to assess the duration of critical tasks. Furthermore, the planning and scheduling functions of PERT are highly sensitive to the accuracy of estimated activity completion dates. The conventional method of identifying the expected length of an activity involves an application of probability theory.

Levin and Kirkpatrick explained that probability theory affords the planner a method of estimating activity completion dates in the absence of data on past iterations of the activity.^[1] Project planners know that, theoretically, a particular activity could be performed 1500 times and, on each occasion, the length of time needed to complete the activity could be recorded. As the authors pointed out, probability theory indicates that some completion times would occur more frequently than other completion times. In addition, some completion times would occur infrequently; in a few instances, the activity would be completed quickly and, in a small number of cases, the activity would consume an extraordinary length of time. Moreover, if an activity were performed a large number of times, a probability distribution of activity completion times would emerge. Considering Levin and Kirkpatrick's discussion, it follows that a project manager could use elementary statistics and the concept of a probability distribution to determine the likelihood of the activity being completed at any particular time.^[1]

Unique public projects are, by definition, performed on only a few occasions or only once. As a result, managers have no data with which to construct a probability distribution of activity completion times. However, the logic of probability theory can be applied to estimate the time needed to complete an activity. This estimate is commonly denoted as the expected completion time of an activity, and many scholars^[2-4]

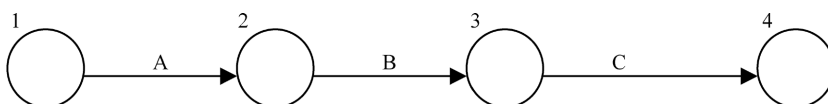


Fig. 1 Events and activities in PERT diagram.

suggest that it may be calculated as follows:

$$t_{etc} = t_{opt} + 4t_{ml} + t_{pess}$$

In this mathematical expression, t_{etc} refers to the expected time for completion of the activity; t_{opt} refers to an optimistic forecast of the time needed to complete the activity; t_{ml} refers to the project planner's estimate of the most likely time necessary to complete the activity; and t_{pess} refers to a pessimistic forecast of the time needed to complete the activity. It is important that t_{etc} is a weighted average—the expected time for completion of an activity is, logically, most influenced by the value of t_{ml} .

Estimations of project completion dates can be made only when project planners are confident in their calculations of the expected completion times of activities. Wiest and Levy noted that, as the range of values between optimistic and pessimistic estimations increases, project planners should be less confident in their calculation of expected activity completion time.^[4] This reflects another implication of probability theory: as the statistical variance of estimated completion times increases for any particular activity, planners must adapt project schedules to the uncertainty that this introduces into estimations of activity completion dates. In these instances, project planners must prepare for activities to be completed in much less or notably more time than suggested by the value of t_{etc} .

THE CRITICAL PATH IN PERT ANALYSIS

The tasks of planning, scheduling, and controlling the performance of activities become more complicated when projects must be completed by a predetermined date. When a project must be completed by a particular date, managers must determine the most effective allocation of personnel and other resources among a large number of distinct tasks. The process of making these decisions is simplified by identifying the sequence of activities most likely to cause delays in the completion of the project. These activities are said to comprise the critical path on the network diagram.

In any project, the critical path is regarded as the most time-consuming sequence of activities and events. On simple PERT charts, the critical path may be identified by finding the route of activities from the project's origin to the project's completion that yields the largest sum of expected times for completion. It follows that the total amount of time needed to complete a project is equal to the total amount of time necessary to complete all critical activities. An important principle guiding PERT/CPM is that a delay in the completion of

any activity on the critical path, often referred to as a critical activity, will cause a delay in the completion of the project. This is to say that critical activities possess none of what is known as “slack.”

Understanding the concept of slack clarifies the distinction between critical activities and non-critical activities. Unlike critical activities, an activity that possesses slack may be delayed without triggering a delay in the completion of the entire project. For example, suppose that the latest time that activity X can be finished without jeopardizing the ultimate completion date of the project is the conclusion of week 10. If the earliest possible finish date of activity X is the conclusion of week 8, activity X possesses 2 weeks of total slack. Thus, assuming that the activity immediately preceding activity X is completed on or before its latest finish date, activity X could begin late or progress could be temporarily suspended provided that the activity is completed by the close of week 10. As a project is underway, planners typically recalculate the slack time associated with events on the PERT chart, and refer to these figures when making scheduling and resource allocation decisions.

How do managers identify the critical path and assess the amount of slack associated with non-critical tasks? Computer software packages are available to project managers for the purpose of quickly ascertaining the critical path and slack times associated with complex projects. However, it is also possible to generate the same information using simple mathematics. The ability to do so reflects a more thorough understanding of key concepts used in PERT/CPM. A simple project with multiple sequences of dependent tasks is useful in demonstrating the process of finding the critical path and ascertaining the amount of slack in noncritical activities.

The process of finding the critical path involves calculation of the total slack associated with activities on the PERT diagram. Critical activities are those activities with zero slack. Furthermore, the critical path consists of the sequence of dependent activities that have zero slack. Identifying those activities with zero slack requires calculation of the latest start time and the earliest start time or, alternatively, the latest finish time and earliest finish time, for each activity on the network diagram.

The first step in calculating total slack is to employ the forward pass procedure.^a This involves the calculation of the earliest finish time for each activity, beginning with the event that signals the initiation of the project. In the project depicted in Fig. 2, the leftmost

^aA detailed explanation of the forward and backward pass procedures can be found in Ref.^[4]. For an alternative explanation of the forward and backward pass procedures, see Ref.^[5].

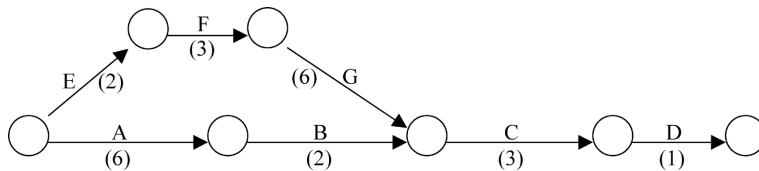


Fig. 2 PERT diagram with expected duration of activities.

node represents the start of the project and the commencement of the first activities. For each activity, the earliest finish time is equal to the earliest time that the activity could begin in addition to the time needed to complete the activity. For example, the earliest start time for activity E is, obviously, zero; at the initiation of activity E, no time has elapsed on the project schedule. Considering that the duration of activity E is 2, its earliest finish time is also 2. Due to the precedence relationship between activities E and F, activity F may begin only when E is complete. Therefore, the earliest finish time for activity F is 5, or the sum of its earliest start time and its duration, which are 2 and 3, respectively. A complete listing of early finish times for all activities is presented in Table 1.

The latest start time for each activity is determined using what is commonly referred to as the backward pass. Beginning with the event that signals the completion of the project (i.e., the terminal node), the latest finish time for each activity is determined. The latest start time for each activity is calculated by subtracting the duration of the activity from the latest finish time. For example, as indicated by Fig. 2, the project must end with the fifteenth time period, the moment that activity D is finished. Considering this, the latest finish time for activity D is 15 and its duration is 1. Thus, the latest start time for activity D is 14. Activity C, the task immediately prior to activity D, requires three time periods for completion. In view of the fact that the latest start time for activity D is 14, activity C must be completed no later than 14. Subtracting 14, the

latest finish time for activity C, from the length of time needed to complete C, which is 3, produces a latest start time of 11 for this activity. Table 1 relates the latest finish and latest start times for each activity in the PERT diagram.

Also depicted in Table 1 is the slack associated with each activity presented in the PERT diagram. Activities A and B possess slack. Activity B, for example, could be finished as early as time period 8 or as late as time period 11. Provided that activity B is completed by time period 11, a delay in the initiation of this activity will not adversely affect the completion date of the project. By contrast, activities C, D, E, F, and G possess no slack. For each, the difference between their earliest and latest finish times is zero. A delay in the start or completion of one or more of these activities will compromise the completion date of the entire project. Therefore, activities C, D, E, F, and G comprise the critical path.

Recall that when constructing a PERT chart, an expected time of completion—a weighted average of optimistic, most likely, and pessimistic projections of activity completion times—is generated and used to schedule sequential activities. The search for the critical path differs in that for each activity project planners employ only one time estimate. As a result, estimations of the critical path are dubious where there is much uncertainty about the duration of critical activities. In fact, mistakes in estimating the duration of activities may cause managers to incorrectly assign tasks to the list of critical activities.

It is assumed that projects must be completed with limited resources and that delays in project completion dates are to be avoided. Thus, managers often want to redirect resources to critical activities to shorten the critical path or remedy an anticipated delay in the completion of a critical activity. Knowing the location and quantity of slack in a project facilitates a manager's need, in some instances, to direct resources away from slack activities toward critical activities in the hope of either avoiding a delay in project completion or shortening the critical path. Although it may increase total project costs, managers can also seek additional resources that can be applied to critical activities in the interests of shortening the critical path and overall project completion time.

When deciding to shorten the length of the critical path by directing additional resources at one or more

Table 1 Start and finish times for activities on the PERT diagram

Activity	LS	ES	LF	EF	Slack	Critical activity?
A	3	0	9	6	3	No
B	9	6	11	8	3	No
C	11	11	14	14	0	Yes
D	14	14	15	15	0	Yes
E	0	0	2	2	0	Yes
F	2	2	5	5	0	Yes
G	5	5	11	11	0	Yes

Note: LS, latest start; ES, earliest start; LF, latest finish; EF, earliest finish; Slack, LF-EF.

critical activities, a process known as “crashing,” a manager must determine how to achieve this result at the lowest cost. Thus, project managers are guided by concerns for efficiency when determining which activities to expedite. Critical path method dictates that shortening the critical path, and, relatedly, the length of the project, is rational only to the extent that doing so either reduces total project costs or increases project costs to an acceptable limit. For example, if a delay in the project completion date is deemed less costly than a crashing strategy, no effort at expediting an activity should be undertaken. It may be that crashing one activity yields a shortened project duration and lower total project costs, but the cost of crashing a second or third activity produces no reduction in project length that yields a cost savings. Determining whether to initiate a crashing scheme, which activities to expedite, and the order in which activities should be expedited are the principal goals of CPM.

In its most basic form, CPM involves a search for the critical activity that can be shortened at the lowest cost. In fact, the costs and benefits of crashing all other critical activities should be ranked, beginning with the activity that can be crashed with the most efficient result. Critical path method dictates that this process continues as long as other critical activities can be crashed to produce savings in project costs.^b

CONCLUSION

Originating in the late 1950s, PERT and CPM remain two of the most important techniques for rationalizing decision-making processes in the public and private sectors. As a result, instruction in each method is a standard feature of courses in project management. Yet, PERT and CPM are little addressed in current scholarship; both techniques are part of the conventional stock of knowledge in management science. Neither method is a source of new theoretical development in the public administration field.

REFERENCES

1. Levin, R.I.; Kirkpatrick, C.A. *Planning and Control with PERT/CPM*; McGraw-Hill Book Company: New York, 1966; pp. 1, 2, 28, 29, 40–45, 121–145.
2. Moder, J.J.; Phillips, C.R. *Project Management with CPM and PERT*; Reinhold Publishing Company: New York. For example, see 1964 p. 204–209.
3. Starling, G. *Managing the Public Sector*; Harcourt Brace and Company: Orlando, FL, 1998; 420–423.
4. Wiest, J.D.; Levy, F.K. *A Management Guide to PERT/CPM*; Prentice-Hall, Incorporated: Englewood Cliffs, NJ, 1969; pp. 26–30, 37–40.
5. Hamdy, T.A. *Operations Research: An Introduction*; Mac-millan Publishing Company: New York, 1987; 474–476.

^bFor a detailed explanation of this process see Ref.^[1].

Public Administration as Management

Lawrence F. Keller

*Public Administration Program, Cleveland State University,
Cleveland, Ohio, U.S.A.*

INTRODUCTION

Public administration began, and the mainstream continues, as a system of management for complex public organizations. From its founding around the turn of the twentieth century during the Progressive era, the field if not profession has focused on the rational running of the public sector. Philosophically, the rational approach is derived from the Enlightenment, a philosophical movement aimed at replacing tradition and similar non-reasoned-based foundations for social action.^[1] The approach underlies most of modern life and, although heavily criticized in some quarters,^[2] shows little sign of abating.

LIMITS TO A MANAGEMENT PERSPECTIVE

The result is a powerful and highly efficient system for operating public organizations. The evolution of this system is the focus of this article. However, it may first be useful to note what has not been part of, or has been difficult to reconcile with, management qua public administration. This highlights both the nature of management in public administration and the limits of the field from a managerial perspective.

Political philosophy has been difficult to reconcile with what Max Weber termed legal-rational organizations.^[3] As a result, public administration as management fostered organizations that were internally authoritarian and externally divorced from politics.^[4,5] They reflected more a belief in the power of a rational management than implementation of a political philosophy.

In addition, public administration as management tended to reduce law to strict rules for conducting daily business and establishing managerial ethics. The result was an emphasis on the contribution of law to rationality rather than as part of the political structure of a democratic republic. Leonard White underscored this in the first public administration textbook in 1927, arguing that public administration should be based on management principles and not law.^[6]

Finally, management was considered a rational activity in the context of an organization. This made it difficult for public administration to handle relations

among organizations and within networks. The focus continues to be on the development and implementation of rational methods for efficiently managing organizations vested with public purposes. History illuminates how this focus came to dominate the field.

DEVELOPMENT OF PUBLIC ADMINISTRATION AS MANAGEMENT

After the Civil War, the United States underwent dramatic changes. The development was quite uneven, with the Northeast and Midwest entering the Industrial Age. This section underwent dramatic urban development, with New York City, for example, adding 1 million people every decade between 1880 and 1920. In the same years, Cleveland rose from a city of 50,000 to a metropolis of 720,000. Many citizens of these cities and often a majority of the laborers in the factories were immigrants.

Simultaneously, the South turned into a racially divided Third World environment, and the West was the scene of battles between the Native Americans and the Army. Much of the urban and industrial development occurred in and after the 1880s, when Reconstruction of the South was abandoned. The abandonment freed the Republican party to pursue industrial policies but imperiled race relations.^[7] Such relations would not be addressed seriously for almost 100 years, until the tumultuous 1960s.

A primary need during this rapid development was a method or methods for running large-scale production. With the advent of the railroad, production could be scheduled yearround. Simultaneously, large amounts of supplies could be shipped to, and a large amount of output shipped out of, a plant. Initial organization of plants was around a foremen system (and it was only a male world until World War II), with foremen responsible for recruiting and directing teams of employees. The foremen system connected work with ethnic neighborhoods as foremen recruited from their neighborhood. However, it complicated coordination of activities and was soon replaced with scientific management, a system in which a staff of planners determined the work process.^[8,9] Scientific management rationalized the work process, stressing

maximum efficiency. Interestingly, breaks and other activities were instituted as they promoted production. The system was a highly rational one, with little or no account of the social or psychological side of human activity. (This outcome was delightfully captured in the movie, “Cheaper by the Dozen,” and detailed in the Hawthorne Plant research.)

Public administration during the Progressive era was largely an urban practice. Initially, concepts of management reflected perspectives on community.^[10] For many, it was an attempt to shape American cities along the lines of the ancient polis, such as Athens or Rome. For others, it was a process for improving the morals of urban life. Thus, many cities established juvenile facilities, usually in the country side, where the separation from adult criminals and the beneficial environment of rural life could save otherwise lost young souls. Finally, for many reformers it was a battle against political machines and what they called “privilege.” Many advocated public ownership of utilities, such as electricity and water, as a way to avoid political privileges granted to a few, that is, those who owned such facilities. However, by the end of the era, the start of the 1920s, reform began focusing on managing the urban organization. Attempts at improving the community were jettisoned from the field of public administration because the concern was with managing the increasingly complex public organization.

Thus, the initial thrust of management in both the public and private sectors was on rationalizing production. Although often not successful in implementing the full version of scientific management, illustrated by Taylor’s testimony to a curious Congress, the rationalization of production became a permanent part of both public and private management. From this perspective, design of work processes was a conscious maximization of efficiency.

By the 1920s, production had ceased to be the primary concern. The management of large-scale organizations became the main issue. In fact, economic development was promoted by the rapid consolidation of industries, leading to monopolies with names such as the United States Steel, National Biscuit Company, and Standard Oil.^[11] Antitrust was partially successful in reducing some monopolies, but the resulting companies were still quite large, often involving multiple plants and needing more effective management.

In the 1930s, the national government developed rapidly. The development was a crisis-driven process that seemed to be chaotic to most reformers. As political opponents often called attention to the chaotic process, President Franklin D. Roosevelt enlisted reformers from public administration to address the

issue. The Brownlow Commission^[12] established the foundation for American public administration with its study of and answers to the organization of the executive branch of the federal government. As part of the Brownlow Commission effort, public administration scholars of the time were recruited by Luther Gulick and Lyndall Urwick to write on the various dimensions of public administration as management. The resulting *Notes on the Theory of Organization*^[13] created the most famous of many management acronyms, POSDCORB (Planning, Organizing, Staffing, Directing, Coordinating, Reporting, and Budgeting). In essence, public administration as management was based on perceived universal principles that constituted effective management of any organization in any culture.

In public administration, management became the sole concern of the executive branch, which developed several key staff agencies with broad mandates and direct access to the Office of President. For example, the Bureau of the Budget, a reform created more than a decade earlier, was moved from the Treasury Department to the newly created Executive Office of the President. Later, the name was changed to the Office of Management and Budget in recognition of the larger roles such staff agencies play.

This general view of public administration was clearly depicted in an earlier novel, *Philip Dru, Administrator*, circulated in unpublished form.^[14] Later, the work was attributed to Woodrow Wilson’s Chief of Staff, Colonel House. In the book, Mr. Dru is appointed Administrator of the United States after a second civil war. After successfully solving American public problems by the end of his 6-year term, he sails to the Soviet Union to become their administrator. The universality reflected, as did the earlier scientific management view of production, the application of science to public administration. For these intrepid practitioners and creators of science, no problem could escape solution if the proper technique were applied. Science would disclose, in unambiguous results, the proper technique. Thus, a robust science of public administration would establish a rational system of solving social problems through the agency of rational management.

At this time, the late 1930s, the American Society for Public Administration (ASPA) was founded. The ASPA deliberately used the preposition “for” in their name to call attention to the organization as a catalyst for action.^[15] The ASPA published the *Public Administration Review*, which became, and has remained, the leading journal in the field. Most of the published articles examine some aspect of management, reflecting the dominant view of the field as public management.

It is indeed a review of public administration as management.

Neoclassical Management

After World War II, a new group of academics began investigating the field. Many challenged the managerial focus—some because of the manner in which management was studied and delineated, others because of the exclusion of politics and philosophy. Dwight Waldo is the most notable example of the latter researchers. His *Administrative State*^[16] illuminated the political roots and consequences of public administration as management, often implicit in much of the literature, and created the phrase that for most captures the resulting polity. This vast literature is beyond this article, which focuses on the former researchers, those who attempted to improve public administration as management.

The most noted of these researchers, and a Nobel Prize winner for his efforts, was Herbert Simon. Simon did a frontal assault on traditional public management, arguing that the so-called principles were nothing but “homilies.” He based his research on logical positivism, the first explicit scientific development of management in the field.^[17] He focused on decision making, distinguishing, in line with his positivistic approach, facts from values. Management for him created organizational systems that efficiently provide factual and value premises to the proper decision maker at the appropriate time. In the first edition of his classic, he distinguished a science of administration from what he called “sociology” of administration. In later editions, he avoided such a stark distinction, but clearly perceived the management of an organization to be a more rational process than other organizational activities.

In line with his scientific approach, he created a more realistic portrait of the administrator. Although earlier writers talked about science, they were actually engaged in what Merton calls technology.^[18] They did not engage in science, but rather used experience and formal logic in a less than rigorous fashion. Thus, Simon’s critique was devastating to the scientific validity of much of the previous management literature. In the classic work titled simply, *Organizations*, March and Simon summarized the empirical findings on management.^[19] The book is an extensive, nearly endless compilation of findings. It illustrates well the results of a careful, truly scientific approach to management.

Although Simon criticized the earlier management literature, he was in turn criticized by Argyris and others concerned with the human variable of organizational life.^[20] Simon and Argyris participated in a

classic dialog in *Public Administration Review*,^[21,22] in which Argyris maintained that efficient organizations were inhospitable to healthy individuals. The scientific base of management selected by Simon did not address the psychological needs of workers in a unified manner. In line with the rational approach of public administration as management, Simon, especially in Simon and March, noted the psychological needs more from a management perspective than from a concern with the health of organizational members. In essence, Simon strengthened the scientific base of management, but did not enlarge its scope or fundamentally shift it from a rationalistic approach.

CONTEMPORARY PUBLIC MANAGEMENT

A final strand of public administration as management emerged in the 1970s and 1980s. During this era, perceptions of government changed. In contrast to the images of public agencies as effective agents of beneficial social change, an image existing since the 1930s, many challenged the ability, or even the legitimacy, of government. The challenge was in part a response to calls for a more active public service, one centered on social equity. Starting with a conference of young administrators and academics held at Minnowbrook, a retreat for Syracuse University, the movement called for administrators to be more politically oriented in terms of the effects of public policies on vulnerable groups.^[23] Although the effort engaged many scholars in extensive debates, it had little effect on the nature of public management. However, it did reflect the general belief in public administration as an agency for social change.

The challengers to an active public service addressing social equity often painted government agencies as ineffective and out of touch with the needs and desires of the citizens. More extreme critiques perceived public agencies as oppressive and pursuing elitist, non-democratic agendas. Some wanted very limited government, recalling the political philosophy of Thomas Jefferson. Others wanted to reform public administration, emphasizing responsiveness to the service demands of citizens. A former city manager and an organizational consultant, Ted Gaebler and David Osborne, personified the latter group of management reformers. They called for the reinvention of government by creating an organizational culture centered on managerial empowerment and entrepreneurial policy making.^[24]

Their call and perspective became the centerpiece of President Clinton’s reform of the federal government. Vice President Gore headed a task force that produced a blueprint for a more responsive federal government, one whose civil servants were empowered to meet

service demands without the stultifying restraints of bureaucracy.^[25] President Clinton stated the goals of the National Performance Review, later the National Partnership for Reinventing Government, in the following terms:

Our goal is to make the entire federal government less expensive and more efficient, and to change the culture of our national bureaucracy away from complacency and entitlement toward initiative and empowerment.

The effort continued throughout the Clinton administration and was the most visible reinventing project. Although the results were debated, both in terms of what was produced and what should have been produced, they were part of a new perspective on public management. This perspective embraced notions of a market and customers, and called on public organizations to create and sustain organizational cultures responsive to the needs of their customers. In the field, the movement became known as the “new public management.” The emphasis was on empowering front-line managers—those who interacted with customers—reducing the bureaucratic overhead of public organizations and decentralizing government generally. An OECD report in 1997 stated, “a new paradigm for public management has emerged, aiming at fostering a performance-oriented culture in a less centralised public sector.”

A researcher identified nine characteristics of the new public management:^[26]

1. Devolving authority, providing flexibility
2. Ensuring performance, control, [and] accountability
3. Developing competition and choice, market-type mechanisms
4. Providing responsive services, client orientation
5. Improving the management of human resources
6. Optimizing information technology
7. Improving the quality of regulation
8. Strengthening the steering functions at the center
9. Private sector-style management

This new thrust, in contrast to previous foci on internal production and organizational control, centered on the interface between the organization and those who were putatively served. The criteria for agency performance were generally taken from economics, emphasizing meeting the needs of those who were to be served. Front-line managers, those who either meet or manage those who meet the clientele, were to be empowered, particularly in the creation of procedural policies. Human resource management

was a concern as the effort required the ability to recruit and reward responsive managers. The human resource thrust continued the policies of civil service reform, conceptualizing management along the lines of and rewarding performance in accordance with private, for-profit management.

In addition, many called for measurement of agency performance, documenting the extent to which agencies fulfilled their mandates and satisfied their customers. These techniques were aided by the extensive development of financial management, culminating in the generally accepted accounting principles (GAAP).^a In contrast to the management in general in public administration, financial management has developed some robust technologies that have been generally implemented. This paved the way for implementation of GAAP, providing effective financial assessment of government.

Human resource management has also created some powerful technologies. Perhaps not as uniformly developed or accepted as is the case with financial management, public administration as management has incorporated many lessons from other fields such as psychology into managing human capital. However, such advances have not been fully integrated into general notions of management in public administration, which despite considerable criticism and research to the contrary still relies on decades old principles.

Although the new public management was part of the reform of government, aimed at reducing the bureaucratic restraints and more conveniently meeting the needs of agency customers, it continued the efficiency criterion of classical management theory. The streamlining of procedures and the rewarding of responsiveness also followed rational models of organization. As a result, the core of public administration as management was continued although the politics of the field were greatly changed.

CONTEMPORARY PUBLIC ADMINISTRATION AS MANAGEMENT

Public administration as management has developed some powerful technologies that greatly contribute to modern governance. This is particularly true with regard to financial management, where professional management has been extensively practiced and

^aFor information on the Generally Accepted Accounting Principles, see the *Yellow Book*, actually titled *Government Auditing Standards*, published by the General Accounting Office, which contains information on both traditional accounting as well as performance auditing.

generally accepted. Even in governments that are generally less than professional, the financial arm meets many professional standards. In large part, this is a consequence of extensive funding of state and local government by the federal government with a consequent demand for accurate accounting of the funds. The application of professional accounting and auditing standards has also contributed. It has been facilitated by creation of techniques such as program budgets and performance audits. Agencies such as the General Accounting Office (GAO) at the federal level have successfully used such techniques.

Similar advances were made in performance measurement. These were to become a significant part of the new public management. One of the earliest efforts was by Herbert Simon, then working with the International Association of City/County Managers.^[27] Some sophisticated technologies permit monitoring achievement of objectives. The rapid advance of the computer and associated information technologies greatly aided such efforts. Conceptually, Simon's frame of management as decision making provided a method for designing information systems. In fact, Simon turned to researching such systems, as well as artificial intelligence, abandoning the field of general management.

To a lesser extent, advances were made in HR management. Incorporating findings and frames from psychology, sociology, and business administration, as well as indigenously in public administration research, robust technologies were created for improving human relations within organizations. For example, the Myers-Briggs assessment of personal activity preference is often used to illustrate the different ways in which people work. This not only promotes tolerance, but also increases understanding about, and the ability of, diverse people to improve work efficiency greatly.

In contrast to the advances in these subfields, public administration as management relies in great part on the early principles. Despite the penetrating criticisms from Simon to Terry, public management has not changed greatly in nearly 100 years.^[28] One suspects the staying power reflects its general utility, both for running complex organizations and for coping with the increasingly political environment of public administration. Similarly, there is no concept more criticized yet more often invoked than the politics administration dichotomy. For the near future, it appears that public administration as management will continue to hone the technologies of its subfields, while implementing classical management in relatively traditional patterns.

In the long term, public administration as management will continue to seek integration into public administration as governance. Such integration

will most probably be difficult, if not impossible. Conceptually, such a quest may parallel the hope for a unified theory in physics. In physics, micro- and macrolevels have resisted explanation via a general theory. Politically, the difficulty of integrating rational management and political governance may reflect a Madisonian political culture and system. In such a system, unification of management and politics may be perceived as creating potentially dangerous authority; in fact, such a unified authority could endanger the system itself. Perhaps much of the criticism of public administration as management is a political critique of its tendency to use powerful technologies through a centralized chief executive divorced from a meaningful political philosophy.

Historically, such a concern is definitely justifiable. The suppression of individual rights and national mobilization during World War I, culminating in the Red Scare in 1920, illustrates dramatically how centralized authority driven by fierce implementation of managerial technologies can endanger the system itself. Eric Foner writes that to the Progressives, "the war offered the possibility of rationalizing American society, instilling a sense of national unity and self-sacrifice, and imposing justice in labor relations. That American power could disseminate Progressive values around the globe heightened the war's appeal." Foner quotes John Dewey, who wrote in the *New Republic*, a journal started by the Progressives and illustrating well what they thought of the political context, urging "Progressives to recognize the 'social possibilities of war.' The crisis, offered the prospect of 'the more conscious and extensive use of science for communal purposes'." With few exceptions, the American Progressives agreed with Dewey's assessment and were willing to sacrifice some rights for a "more effective freedom."^[29]

In his famous work on the Progressive era, Robert Wiebe wrote of *The Search for Order* in what the Progressives called the "New Republic." In a chapter entitled, "The Illusion of Fulfillment," he wrote about the migration of Progressive focus from its initial concern with the modern city to the nation:^[30]

Success had bred confidence among urban and state reformers, and rather than undertake more of the same, they moved outward. Progressivism contained an inherent expansive thrust, partly from its need for ever-broader legislation, partly from its all-conquering optimism, but even more from its faith in method. If the right technique guaranteed the right results, no problems, whatever the size and scope, could withstand its magic.... They would join these millions of Americans [in all manner of movements] in a single crusade against inefficiency and injustice of special privilege, social ignorance, and habitual indifference.

... Such men as Frederic Howe and Walter Rauschenbusch not only encouraged local progressives to think of themselves as part of one grand drive, but they also supplied the rationale for integration: the skeleton of stages to an inevitable triumph, fleshed out with a new national intelligence; and at its heart the wonders of rational management.

In light of the possible mischief in the application of rationalism, and even more so its embodiment in powerful technologies at the disposal of highly partisan and political chief executives enacting political ambition, the field may well content itself with a muddling-through approach.^[31] This would implement powerful technologies in selected areas, while recognizing and respecting core political values and individual rights. Public administration as management must begin to put political philosophy on par with rational methods even though this may be somewhat inefficient. In essence, public administration as management must recognize it is part of politics and not a crusade of science driven by experts with visions of a one and true society.^[32] The creation of powerful financial tools demonstrates that public administration as management can play meaningful political roles through professional standards. The danger is in extending rational tools into a rational system. Both history and philosophy caution against such an extension, and can serve as excellent guides to a future of professional and effective public management in a constitutional political system in which public administration also performs critical governance responsibilities.

REFERENCES

- Keller, L.F. Public administration, city management and the american enlightenment. *Int. J. Public Adm.* **1989**, XX (2), 213–249.
- McSwite, O.C. Post modernism, public administration and the public interest. In *Refounding Democratic Public Administration: Modern Paradoxes, Postmodern Challenges*; Wamsley, G., Wolf, J., Eds.; Sage Publications: Thousand Oaks, CA, 1996; 198–224.
- Gerth, H.H.; Mills, C.W. *From Max Weber: Essays in Sociology*; Oxford University Press: New York, 1946.
- Denhardt, R. *In the Shadow of Organization*; Regents Press of Kansas: Lawrence, KS, 1981.
- Hummel, R. *The Bureaucratic Experience*; St. Martin's Press: New York, 1977.
- White, L. *Introduction to Public Administration*; MacMillan: New York, 1926.
- Benedict, M.L. Reform republicans and the retreat from reconstruction. In *The Facts of Reconstruction: Essays in Honor of John Hope Franklin*; Anderson, E., Moss, A.A., Jr., Eds.; Louisiana State University Press: Baton Rouge, 1991; 53–77.
- Bodnar, J.E.; Simon, R.; Weber, M.P. *Lives of Their Own: Blacks, Italians and Poles in Pittsburgh, 1900–1960*; University of Illinois Press: Urbana, IL, 1982.
- Taylor, F. *The Principles of Scientific Management*; Harper: New York, 1911.
- Hoffman, M. City republic, civil religion, and the single tax: The progressive-era founding of public administration in cleveland, 1901–1915. In *Dissertation, Public Administration Program*; Cleveland State University, 1998.
- Chandler, A.D. *The Visible Hand: The Managerial Revolution in America*; Belknap: Cambridge, MA, 1977.
- Brownlow, L. *Administrative Management in the Government of the United States/The President's Committee on Administrative Management*; Government Printing Office: Washington, DC, 1937; 1–47.
- Gulick, L., Urwick, L., Eds.; *Papers on the Science of Administration*; Institute of Public Administration, Columbia University: New York, 1937.
- House, E.M. *Philip Dru, Administrator*; Literature House: Upper Saddle River, NJ, 1969.
- Hamilton, M. The American society for public administration and the field of public administration. In *Handbook of Public Management Practice and Reform*; Liou, K.T., Ed.; Marcel Dekker: New York, 2001; 675–712.
- Waldo, D. *Administrative State: A Study of the Political Theory of American Public Administration*; Ronald Press Company: New York, 1948.
- Simon, H. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*; MacMillan: New York, 1947.
- Piotr, S., Ed.; *On Social Structure and Science/Robert K. Merton*; University of Chicago Press: Chicago, 1996.
- Simon, H.; March, J.; Guetzkow, H. *Organizations*; Wiley: New York, 1958.
- Argyris, C. *Personality and Organization: The Conflict Between System and the Individual*; Harper: New York, 1957.
- Simon, H. Organizational man: Rational or self-actualizing? *Public Adm. Rev.* **1957**, 33 (4), 346–353.
- Argyris, C. Organizational man: Rational and self-actualizing? *Public Adm. Rev.* **1957**, 33 (4), 354–357.
- Marini, F., Ed.; *Toward a New Public Administration: The Minnowbrook Perspective*; Chandler Publishing Company: San Francisco, 1971.
- Gaebler, T.; Osborne, D. *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*; Addison-Wesley: Reading, MA, 1992.
- Gore, A. *Creating a Government That Works Better and Costs Less: Report of the National Performance Review*; Government Printing Office: Washington, DC, 1993; 1–168.

26. Kasemets, A. In *Implications of New Public Management Theory in the Research Services*, Paper Presented at the 66th International Federation of Library Associations and Institutions, Jerusalem, August, 13–18, 2000.
27. Ridley, C.; Simon, H. *Measuring Municipal Activities, a Survey of Suggested Criteria for Appraising Administration*; International City Manager's Association: Chicago, 1943.
28. Terry, L.D. *Leadership of Public Bureaucracies: The Administrator as Conservator*; Sage Publications: Thousand Oaks, CA, 1995.
29. Foner, E. *Story of American Freedom*; W. W. Norton & Co.: New York, 1998.
30. Wiebe, R. *The Search for Order 1877–1920*; Hill and Wang: New York, 1967.
31. Lindblom, C. *The Science of "Muddling Through,"*; 1957; 79–88. Bobbs Merrill Reprints in the Social Sciences; PS 109, Indianapolis.
32. Lasch, C. *The True and Only Heaven: Progress and Its Critics*; Norton: New York, 1991.

BIBLIOGRAPHY

1. Perry, J., Ed.; *Handbook of Public Administration*, 2nd Ed.; Jossey-Bass: San Francisco, 1996.
2. Rabin, J., Hildreth, W.B., Miller, G.J., Eds.; *Handbook of Public Administration*, 2nd Ed.; Marcel Dekker: New York, 1998.

Public Administration, A Twenty-First Century Lens of Understanding

Thomas D. Lynch

Public Administrations Institute, Louisiana State University, Baton Rouge, Louisiana, U.S.A.

Cynthia E. Lynch

Nelson Mandela School of Public Policy and Urban Affairs, Southern University, Baton Rouge, Louisiana, U.S.A.

Peter L. Cruise

Department of Health and Community Services, California State University, Chico, California, U.S.A.

INTRODUCTION

This article argues that public administration uses philosophy as a lens of understanding and that the current lenses available to us are inadequate for the twenty-first century. Possibly the metaphor of eyeglasses that we use to enhance our eyesight is the best means to explain the argument in this article. When we are young, we may be lucky enough not to need eyeglasses, but many of us soon have to get them to bring our eyesight up to a standard that permits us to function in our environment. As we grow older, our prescription often needs to change, as we change. This article argues that humankind has also grown older. As that aging process occurred due largely to technological change, humankind needed to and did change its prescription to adapt to its changed circumstances. The twenty-first century is an extension of that evolutionary process of humankind and the various prescriptions, called philosophy, are again inadequate, especially in the area of ethics and morality in public administration. We need an enhanced or improved prescription to help us in the twenty-first century with its technologically dependent society.

Joseph Rost argued that the world is experiencing a radical transformation, which futurists claim change the basic values of the present era.^[1] Charles Garofalo predicted that the rational, male, technocrat, quantitative, goal-dominated, cost-benefit driven, hierarchical, short-term, pragmatic, materialistic model must give way to a new kind of leadership based on different assumptions and values.^[2] Garofalo also asserted that a pervasive sense exists that our fundamental perspectives on life are radically different from previous times, and any new values or perspectives built on the past paradigms are no longer adequate for the next century. According to him, “our civilization is

unbalanced, with our material culture far ahead of our ethical and moral independence.”^[2]

This article makes the case for a fundamental change in perspective, or a “lens change” in the vocabulary of our metaphor, in public administration. This article also explains the basics of that desired change. The first section after this introduction explains why we need to rethink philosophy, especially its role in the public sector. The next section critiques modernist and postmodernist philosophy, and makes the case for not using the fundamental assumptions that underlie those philosophies. Last, this article proposes a more useful, twenty-first century approach to rethinking the philosophical lens for public administration.

RETHINKING PHILOSOPHY

The Role of Philosophy

In thinking about the role of philosophy in the public sector, some consider philosophy as a means to explain phenomena and answer larger questions, such as the nature of understanding, the basis to judge good and evil, or the proper role of government in society. This article views philosophy as a conceptual lens through which we understand the thoughts and phenomena about those important fundamental matters. From those understandings, we then proceed to act and manifest behavior that our philosophy thus largely creates.

Although a strong influence on humankind, philosophy is also a dependent variable that society shapes. However, before that, key technologies shape society. Philosophy is a product of linear influence or, more properly in many cases, a synergistic influence with society. Alvin Tofler’s *Future Shock* dramatically presented the importance and influence of technology on

society.^[3] In that book, the driving force changing society was not philosophy, but rather technology. For example, the invention and eventual use of gunpowder revolutionized the art of war by ending the importance of the defensive castle that was key to the decentralized power relationships of the baron to the king in feudalism. Gunpowder strengthened the role of the crown and greatly strengthened the power of what eventually became the nation state. Another example was the invention of the precise clock, which meant that accurate navigation by longitude became possible. In turn, this meant that an economic philosophy built on mercantilism was not only possible, but also a superior means than war to increase the wealth of the emerging nation states such as England.

A major thesis of this article is that certain key technologies in various eras drive and shape society so that eventually the old lens becomes increasingly less functional, and that we need a new lens to better conduct ourselves in that revised society. New technologies will always advantage the people in society who are most able to adapt and think in new ways. They quickly comprehend their advantages and adapt their lens of understanding accordingly. As a result, in the evolved society, they tend to be more successful than others in terms of wealth, power, or both. Others may be slower to recognize the changing opportunities, but eventually they learn by experience and the demonstrated successes of those that adapted earlier. As their success becomes more obvious, others follow the example of success. In the past 200-year period, the two defining technologies are industrialization and information. In their respective eras, each had broad and remarkable impacts on almost every aspect of society.

As society changes, intellectuals cast about for new lenses to use to understand the conditions that are changing their lives. New philosophies are born as the older lenses become inadequate to explain the shifting paradigm brought on by technological change. For example, industrialization resulted in concentration of wealth and political power in the hands of a few economic barons. In turn, this situation meant that government could take over or control those limited power centers with an administrative state. The lenses to accomplish that end were socialism and progressivism that both accepted and championed the primary role of the state in society.

Philosophy acts as a mental filter for the more intellectual among humankind, and it is also the bases for humankind to act and make judgments in ways defined by the particular philosophy. Because philosophy has influence, it advantages and disadvantages various groups in society. Over time, if the philosophy is a good lens on society, the advantaged groups prosper and gain increasing power. Commonly, even the disadvantaged groups begin to adopt the philosophy

of the successful groups if the new ruling elite permits them. A good example is the invention of modernism. In Western Europe, modernism fit the new industrial society particularly well because the so-called objective empirical inquiry using the so-called scientific method tended to create functional knowledge for the advantaged elite groups. Those industrial and government groups tended to prosper and other societies around the world saw the remarkable successes, and began to emulate them and adopt modernism.

Modernism and Public Administration Theory

Contemporary public administration can be thought of in terms of modernist thinking and, to a much lesser degree, various counter perspectives. Less directly, but nevertheless significant, are the premoderns: Plato and Jesus. Few would argue that Plato did not influence Western thought, but with the American and eventually the Western tradition of secularization, intellectuals rarely speak of their faith traditions except within the walls of houses of worship.

Nevertheless, faith and especially Christianity radically influenced Western thought and particularly the views of the premodernists. For example, the faith traditions of Judaism and Christianity with its stress on the 10 commandments and the golden rule still define today what people accept as fundamental rights and wrongs in society. In the premodern society, interpretations of religious texts heavily influenced the very conduct of what intellectuals considered proper inquiry that advanced human knowledge. For example, interpretations of religious texts in the Middle Ages brought intellectuals to think that the center of the universe was the earth and thus leaders punished scholars that dared to be politically incorrect thinkers. In defining the modernist, Rene Descartes and Francis Bacon are important, but so are Jeremy Bentham, John Locke, and Adam Smith, who also significantly refined the modernist "lens." These philosophers were secular thinkers that focused on the good of the people as defined by rigorous rational thought associated with the scientific method. Many of them defined nineteenth-century liberalism with its distrust of government as a social instrument. They had great faith in the rational thinking capacity of mankind to discover, articulate, and apply knowledge. In contemporary language, the term "liberal" shifted in meaning, primarily due to progressives in America and socialists in Europe, who embraced and envisioned government as a social instrument unlike the earlier nineteenth-century liberals. In both the nineteenth and the twentieth centuries, the hallmark of modernists is their faith in human reason to discover truth and use it to improve the human condition. Modern science is a product of that faith.^[4,5]

One could easily stop with the modernists as their influence is so significant on Western thought, but there are other views that were becoming increasingly heard and influential. Two philosophers, David Hume and Edmund Burke, questioned the capability of human reason to seek out and find knowledge that could guide our civilization with sufficient certainty. Later modernist opponents were Jean-Paul Sartre, John Rawls, and the school of thought known as phenomenology. Each builds on earlier philosophers and challenges the fundamental core of modernist thought.^[6-8]

Modernism, such as developed and used by John Locke and Jeremy Bentham, biases human thinking toward left-brain analytical thinking and away from religious right brain thinking that dominated the very being of human behavior in the previous centuries prior to the arrival of modernism. With modernism, religious thinking became marginalized and even dysfunctional. Modernism advantaged gifted analytical left-brain thinkers, who enjoyed great success in society. However, right brain thinking did not disappear, but those who thought in that manner were less central to power and economic success. Especially among the world's intellectuals, power became secularized and religious thought became decreasingly important even in such matters as values and ethics.

Postmodernism and Public Administration Theory

Possibly because of later modernist opposition due to partly the popular discontent with government policy in the developed world, there was an increasing direct challenge to modernist thinking. Postmodernism arose first with Friedrich Nietzsche, but gained much of its current direction from Ludwig Wittgenstein, who had abandoned his earlier version of modernism called logical positivism. Most cite Friedrich Nietzsche as the first postmodern philosopher, and many would agree that Ludwig Wittgenstein was the most influential in the more contemporary group of such thinkers. If one had to cite an area that developed the philosophy the most, clearly France is where this philosophy found its most fertile ground to grow.^[9]

Various concepts and vocabulary populate the postmodern universe. For example, postmodern thought is the rejection of universalism, essentialism, ontological realism, and metanarratives. In other words, postmodern thought rejects any absolute ahistorical and universal truths such as God or a universal knowledge based on science. They even reject the quest for such truths.

The lens of postmodernism sees multiple paradigms in which one paradigm believer cannot logically

dispute the correctness of another paradigm believer. However, within a paradigm or localized logic, we can use language games to at least rule out some non-sensical reasoning. For postmodernists, the self is not subjectively determined, but largely influenced by the inherited language games of the time and culture. There is no centered unified self, but rather we are split between our conscious and unconscious. Knowledge is merely institutional rules that guide our discourse and our actions. Truth is merely vocabulary that arbitrarily defines itself as truth.

To postmodernists, no definition is possible for fundamental concepts such as "being." Words are only weak, but necessary, abstractions of abstractions and nothing more. Thus, language, the vehicle of communication, limits full and complete interconnective understandings among humankind because each person must speak from their *relative* position in the universe. According to its adherents, postmodern thought does contribute to public administration theory. For example, it teaches us the foolishness of most so-called policy decisions that are meant to be universal or broad in implication. We also learn from postmodernism that organization structure is in itself a system of power. Last, we learn that reality is not important, but rather what is important is the measure that we use to indicate the condition of reality.^[10,11]

Philosophy does influence the way humans see, understand, and judge. Philosophy does not advantage all humans in exactly the same way because other factors such as heritage and social conditions are also important. Nevertheless, philosophy is a remarkably powerful influence on humankind in general and on some people in particular.

CRITIQUE OF MODERNIST AND POSTMODERNIST PHILOSOPHY

Certainly, humankind is at its historical zenith and modernist philosophy, in particular, can take some of the credit. Today, modernism leads humankind to believe that it can eventually know everything and with knowledge there are no limits. Postmodernism is an active minority voice of realism that questions the optimism of the modernist and takes some of the authority out of what could be an intellectual tyranny. Because of modernism, humankind concentrated on developing left-brain analytical capabilities. In contrast, postmodernism reminds humankind that the left brain is insufficient. Thus, the two contesting yin and yang positions present humankind with a balance that not only is inconsistent, but also dynamic and functional. Nevertheless, this article argues that both are inadequate for the twenty-first century.

Modernism and postmodernism thinking place belief in an ultimate truth or absolute as either beside the point or simply not relevant. We do not need to debate the role of doubt in philosophy again as did Descartes, but this article needs to discuss the issue of the absolute's existence, as it is central to our theme. Postmodernism adopts a relativist perspective that only accepts judgments of truth within the confines of a single paradigm. Modernists adopt a temporary version of truth that is subject to continual revision. In a pragmatic sense, neither accepts a fundamental or universal permanent truth, especially the existence of an absolute. By different paths, both produce the same result—both say essentially that truth is relative, as are ultimate values including ethics.

What are the consequences of relative values and ethics? Relative values and relative ethics are a certain formula for human conflict in which a few can terrorize the many because they are so certain of their beliefs and so dismissive of the beliefs of others. With modernism and postmodernism, consensus or capitulation is the only means to resolve conflict between parties. With modernism and postmodernism, there is no ultimate means to settle a dispute by appealing logically to ethics or morality.

With those two lenses of understanding, we only consider people who oppose us as not being consistent with our understanding; they can never be correct because no such condition exists. Our lenses focus our understanding on the temporary nature of truth or the impossibility of thinking that truth exists between people who hold different paradigms of understanding. Such a reasoning process not only leads to openness, but also creates a cynicism among the premodernists that we commonly call fundamentalists in religious traditions. Such premodernists do not see the value of consensus and what the other groups claim to know. That cynicism can provide the premodernist the rationale to continually bring terror to the many.

Is there an escape from this philosophic cul de sac? Yes, but we must first understand that we can conceptualize the altruistic/materialistic, rational, and government capability assumptions used by Western philosophers in a simple matrix of three sets of two choices each. Most Western philosophers make fundamental assumptions in their "lens" that we can grasp in terms of this simple matrix. For example, some philosophers like Thomas Hobbes argue that mankind is essentially materialistic and in need of a strong leader to address their needs. Some, like Jean-Jacques Rousseau argue that mankind over time is essentially concerned with the needs of all. Some, like Jeremy Bentham, argue that humankind can reason through its general welfare by using rational thought, including analysis. Others, like Edmund Burke, argue that such large matters are beyond rational human inquiry and

that such decisions need to rest on appeals to historical traditions. Twentieth-century liberals and nineteenth-century conservatives argued that government could be a positive instrument to direct society. Still others, like twentieth-century conservatives and nineteenth-century liberals, argued that government cannot be a positive instrument for society.

THINKING OUTSIDE THE BOX

James Madison: Reconceptualizing the Box in the Public Sector

Although the three dimensions constitute a "box" that limits the choices of some philosophers and significantly contributes to their lenses, we need not be so limited. We can reconceptualize the matrix or box in a more sophisticated way or even think outside the box. James Madison thought within that box but he also reconceptualized the box. He decided each dimension was not a matter of one or the other, but really a mix of the two opposite view points.

In other words, Madison, as a primary author of the U.S. Constitution, assumed that in part humankind was both altruistic and materialistic. To him, the goal of the government process was to maximize the likelihood that leaders in society would make the altruistic choice. In addition, as he considered the matrix, Madison did not believe that leaders could always be either rational or incremental in their decision making, but instead he assumed that major public policy decisions were often made in some combination of the two means of decision making. In a similar manner, he assumed that government was not necessarily the sole proper or improper agent to solve society's problems, but he assumed again that a mix was often the proper solution to the challenges of living together as a society and that the mix would like alter over time.^[12]

Another option for confronting the matrix assumption of philosophers is to transcend the box entirely. For example, instead of assuming that people are either altruistically or materialistically driven, or even a mix of the two, we can assume that individual free choice can allow people to move from being materialistically driven to a higher plane of altruistic motivation. Thus, for example, people can morally improve with education and training. In this way, a whole new set of policies becomes possible and even likely with a new prescription in our lens.

Under this assumption, the concept of free choice is critical because humankind is not bound to any motivation, but can transcend to a higher plane or even fall back to a lower plane of moral development. With such an assumption, the role of philosophy is not to interfere with free choice, but rather to help

those interested to transcend to a higher plane of altruistic choice. Those that make this assumption are somewhat similar to James Madison in that they also recognize that many, and possibly most, people will still select the materialistic-driven choice in spite of their freedom to select the higher plane of moral thought. They are dissimilar to James Madison in that making people aware of that choice and helping those who want to develop morally are central beliefs of this new lens. To those thinkers, aspiration ethics, rather than conformist rulebound ethics, is critical to the development of a more advanced civilization that can better cope with the challenges of the information era.

Toward a Global Ethic

The box looks at the rational dimension of the matrix in a different way. It agrees with the nineteenth-century conservatives that complete rationality is impossible, but then also assumes, with God's help, that humankind can transcend to a higher thinking capability that somehow combines left- and right-brain thinking. For example, in this view, thinkers believe that humankind is not capable of reaching a viable global ethic with mere rational thought. However, they believe that humankind can combine rational thought with the feelings of the heart associated with the transcendent. When this is done, they believe that a significant part of humankind can accept a global ethic based on the spiritual wisdom of many of the religions of the world. Such a decision-making process can establish a universal ethic and encourage decision making using the full range of human capability. The best example of this type of reasoning is found in the work of the second Parliament of Religions and Hans Kung, in particular.^[13]

To transcend the box, one must consider the philosophic assumption of the appropriate role of government in society. In the box, there is an either/or assumption. To transcend the box in terms of the proper role of government in society, we simply assume that this decision is really not important or salient to the real purpose of life. This philosophy argues that in the larger scheme of things, the role of government in society is not really all that significant and not one of the major appropriate questions for philosophy to address. Transcendent thinkers maintain that what is salient is the care each person and each group of persons takes in others and in nature. Both government and private action can and should manifest that caring. They argue that the key is human caring and not the instrument used to manifest that caring. What is the proper mix is a function of culture and the programmatic contextual situation that each set of government leaders must decide in their own time and circumstances.

RETHINKING PHILOSOPHY

The Twenty-First Century Lens

Philosophy is a collection of lenses of understanding that helps people to realize the fundamental solution for society's problems is primarily a matter of getting people to think of themselves as interconnected. When that is accomplished, government is sometimes the solution and sometimes the problem, depending on contextual situations that change with the times and the place. To the transcendental philosophers, the answer to the government dimension is not a constant "yes" or "no," but rather "it depends." That answer transcends the box.

Is a philosophy that transcends the box possible? The answer is "yes," and it existed in its fundamental form since the beginning of recorded history. The common spiritual wisdom literature (that is found among other places in the Holy Scriptures of the Hindu, Jewish, Buddhist, Christian, and Islamic traditions) provides the basis for developing a twenty-first century philosophy. Humankind need only understand and apply that literature to develop a lens of understanding suitable for the new century. For example, the *Koran* uses the metaphor "blind" and says, "Momentous signs have come to you from your Lord. He that sees them shall have much to gain, but he who is blind to them shall lose much indeed."^[14] In other words, we are continually given signs and messages (lessons) from God. Unless we learn those lessons with our hearts, we will remain blind and unable to update the lens by which we understand life. The scriptures of all traditions teach us that government is not the answer to society's problems, but rather the answer is individually and collectively within each of us.

Moving from the abstract to the practical application is always the most difficult aspect of developing a new philosophy. What would such a philosophy mean to public administration? To the maximum extent possible, society would need to minimize judgmental behavior and maximize positive supportive altruistic behavior. Certainly, materialistic and ego-driven behavior would exist. However, under this new philosophy, when such behavior caused harm to other individuals and nature, leaders and society would either discourage or when necessary curb such negative behavior to protect people and nature. This new philosophy creates social goals to minimize divisiveness among groups, and to maximize supportive behavior of individuals and groups that supported the whole of society. However, it also paradoxically encourages individual self-development and expression.

Making decisions using this philosophy would be significantly different. This new philosophy would encourage individual use of their total mental capabilities,

including their intuitive sense of righteousness as manifested in virtue ethics. Most significantly, this new philosophy holds that improving the total person must include both the right and the left brain. Under such a new philosophy, possibly the following best defines the importance of the total mind: "Everything has mind in the lead, has mind in the forefront, is made of mind. If one speaks or acts with a corrupt mind, misery will follow, as the wheel of a cart follows the foot of the ox. Everything has mind in the lead, has mind in the forefront, and is made by mind. If one speaks or acts with a pure mind, happiness will follow, like a shadow that never leaves."^[15] Following this philosophy, part of professional training and education is global ethics applied to the work context so that administrators are well equipped to exercise their professional judgments through their responsibilities.

Under this philosophy, government is merely an instrument that is ultimately of secondary importance. Policy makers and public managers would need to consider themselves in their governmental roles as radically less significant than their roles as humans. Their primary purpose in life is their own spiritual development and their service to their fellow human-kind. This philosophy does not stress status or power of the administrator, but rather the central role of their total growth as a person in service to others. With such thinking, corruption does not make any sense and decisions that are meant to benefit the whole rather than one group makes all the sense in the world.

Applying philosophy in this manner is certainly not a call for a theocracy, but merely a call to look and apply the spiritual wisdom that informs all our religions. The scriptures of all traditions warn us that religious leaders are not immune to materialistic ego-driven decisions. Clearly, their claims to being religious enable them to easily abuse the special regard afforded to them by others. This philosophy is not a call to select leaders by religious tests, but rather a call to define a new philosophy that places that inner worth of the individual as being centrally important. This philosophy defines worth not in terms of religious doctrinal zeal or intolerance, but in terms of accumulated daily actions that reflect a person's continuing concern for the interests of the whole rather than the causes of one or any given group in society.

REFERENCES

1. Rost, J. *Leadership for the Twenty-First Century*; Praeger Press: New York, 1991.
2. Garofalo, C. In *Leadership, Ethics and Change*, Paper Presented at the Annual Meeting of the American Society for Public Administration (ASPA), San Antonio, TX, April, 4-6, 1995.
3. Toffler, A. *Future Shock*; Bantam Books: New York, 1970.
4. Griffith, M. John Locke's influence on American government and public administration. In *Handbook of Organization Theory and Management: The Philosophical Approach*; Lynch, T., Dicker, T., Eds.; Marcel-Dekker: New York, 1998; 125-140.
5. Martin, L. Jeremy Bentham: utilitarianism, public policy and the administrative state. In *Handbook of Organization Theory and Management: The Philosophical Approach*; Lynch, T., Dicker, T., Eds.; Marcel-Dekker: New York, 1998; 111-123.
6. Spicer, M. David Hume and public administration: empiricism, skepticism and constitutionalism. In *Handbook of Organization Theory and Management: The Philosophical Approach*; Lynch, T., Dicker, T., Eds.; Marcel-Dekker: New York, 1998; 167-180.
7. Haque, A. Edmund Burke: the role of public administration in a constitutional order. In *Handbook of Organization Theory and Management: The Philosophical Approach*; Lynch, T., Dicker, T., Eds.; Marcel-Dekker: New York, 1998; 181-202.
8. Waugh, W. Phenomenology. In *Handbook of Organization Theory and Management: The Philosophical Approach*; Lynch, T., Dicker, T., Eds.; Marcel-Dekker: New York, 1998; 305-324.
9. Watson, R. On the language of bureaucracy: postmodernism, plain english and Wittgenstein. In *Handbook of Organization Theory and Management: The Philosophical Approach*; Lynch, T., Dicker, T., Eds.; Marcel-Dekker: New York, 1998; 389-413.
10. Fox, C.; Miller, H. Postmodern philosophy, postmodernity, and public organization theory. In *Handbook of Organization Theory and Management: The Philosophical Approach*; Lynch, T., Dicker, T., Eds.; Marcel-Dekker: New York, 1998; 45-437.
11. Dobuzinskis, L. From positivism to postmodernism: an unfinished journey. In *Handbook of Organization Theory and Management: The Philosophical Approach*; Lynch, T., Dicker, T., Eds.; Marcel-Dekker: New York, 1998; 359-387.
12. Lynch, T.; Cruise, P. Federal budgeting prognosis: Madison's nightmare? *Int. J. Public Adm.* **1992**, *15* (5), 1053-1066.
13. Lynch, T.; Lynch, C.; Cruise, P. Is a global ethics possible? *Int. Rev. Public Adm.* **2001**, *6* (1), 59-69.
14. Koran 6: 104. In *The Wisdom Literature*; O'Connor, K., Ed.; The Liturgical Press: Minnesota, 1990.
15. Dhammapada 1:1-2. In *The Wisdom Literature*; O'Connor, K., Ed.; The Liturgical Press: Minnesota, 1990.

Public Authorities

Suzanne Leland

*Department of Political Science, University of North Carolina at Charlotte,
Charlotte, North Carolina, U.S.A.*

Gary A. Johnson

University of North Carolina, Charlotte, Charlotte, North Carolina, U.S.A.

INTRODUCTION

Public authorities are often labeled oddities because of their quasigovernment approach to delivering services. They function outside the normal framework of government and are sometimes not counted as units of government by the census bureau.^[1] These ad hoc agencies are hybrids of public, nonprofit, and private organizations. They typically share some characteristics of general purpose governments (such as cities and counties), yet stand unique and separate in their own respect because they are utilized to finance, develop, and operate revenue-producing facilities and services. Public authorities establish their own personnel systems, adopt their own corporate name, make bylaws, have the right to sue and be sued, possess the power of eminent domain, exercise the discretion to establish rates and charges, are exempt from property tax, enter into agreements with other public and private organizations, and have the ability to utilize private money markets. In general, public authorities are a form of special purpose governments. That is, they are typically formed to deliver a specific service as opposed to multiple services. Like special districts, public authorities come in all shapes and sizes. Their characteristics vary depending on the services they deliver and the federal or state legislation from which they were created.

Public authorities have become influential in the United States, operating at the national, state, and local level and performing almost every public service imaginable. They borrow more in the tax-exempt bond market than state or local governments. The authorities have the ability to bypass debt and taxation limits that state and local governments are often subjected to.^[2] They are powerful economic and political institutions that control trade, shipping, waste disposal, and transportation in most major cities. Increasingly, authorities are utilized as conduits for intergovernmental grants and aids. They have been justified as a mechanism to rescue ailing industries, promote regional economic development, and subsidize private industry.^[3] To some, public authorities are viewed as undemocratic institutions that are insulated from the

electorate. Others criticize them as being heavy borrowers created to provide better access to capital markets.^[4]

SCOPE

For the most part, public authorities are major governmental players frequently ignored by public policy and public administration research. However, such “peculiar animals,” in the words of Doig, are more common than one would think.^[5] According to a study by Leigland, there are over 7000 public authorities in the United States.^[6] In 1995 alone, they issued \$75.3 billion in new long-term debt.^[7]

Estimating the number of public authorities in the United States is difficult for several reasons. First, state and local governments do not necessarily track the development and number of authorities. They are continuously created and dissolved. Second, the U.S. census bureau does not separate public authorities from special districts. Some special districts use the name “authority” in their title but have few characteristics of public authorities. Likewise, some special districts do not use the word “authority” in their title but have characteristics of an authority.

Some scholars argue that a public authority is by definition a corporate entity.^[8] Dimock and Dimock suggest that most state and local government corporations are called authorities.^[9] A number of other scholars do not find corporate status to be a key characteristic of public authorities.^[10] The matter is further complicated when organizations, such as those by the Council of State Governments, use such terms as “public authority” and sometimes “special district” generically to apply to classes of entities with similar characteristics, regardless of particular names given to these entities in enabling legislation. The generic use of such terms results in confusion when what a state government calls a “special district” is labeled a public authority by a social scientist. For example, Walsh, Doig, and Gerwig all exclude, by definition, from their studies of public authorities the quasi-government organizations that have taxing power.^[5,8,11]

Yet, by 1982, 23 states allowed some type of entities formally identified in state statutes as authorities to levy taxes. The argument is often made that districts and authorities can be distinguished from one another on the basis of financing methods, that districts levy taxes (and sometimes issue general obligation bonds) and authorities issue revenue bonds and are prohibited from levying taxes. Gerwig claims that the power to issue revenue bonds is a “major distinguishing characteristic” of public authorities.^[11] Walsh maintains that most special districts do not borrow heavily. This generic use of terminology is awkward because it means that a majority of states have somehow applied the name “special district” to entities that are in reality public authorities.

HISTORY

The British Parliament created the first “modern” authority in 1909 when London’s ports were consolidated into a financially sufficient organization. The Port Authority became known for its financial security and the continual development of facilities.^[2] The progressive movement inspired the use of public authorities in the United States as a way to “depoliticize” administrative operations for the management of bridges, parks, ports, and other infrastructure. Ideally, they were constructed to separate politics from administration just like the city manager form of government. They were advocated over traditional governments because their structure incorporated business practices and expertise and have their roots in the managerial approach to public administration.^[12] The first authority established in the United States was the Port of New York Authority in 1921. Later, the Port Authority was renamed the Port Authority of New York and New Jersey. This spurred the development of other port authorities along the East Coast. A second type of authority emerged with the creation of the Tennessee Valley Authority during the Great Depression. Public authorities flourished after this. They include policy areas such as economic development, education, utilities, environmental protection, health, housing, marine services, public facilities, and transportation. There are also multipurpose authorities that implement two or more public policies. They can behave similar to general purpose governments. The majority of these are located in the northeast, particularly in Pennsylvania.^[2]

PUBLIC AUTHORITIES AND STATE ENABLING LEGISLATION

As pointed out in the previous section, public authorities are defined differently depending on whom one

asks. However, since most public authorities are a form of local special purpose governments, they are typically enabled by state law. Therefore, state enabling legislation plays the key role in defining the parameters of public authorities. Quite simply, without legal permission to form, public authorities cannot exist. So what one state defines as an “authority” may differ from another, and therefore it is difficult to construct a uniform workable definition throughout the United States. Because there are 50 different states, there are as many different sets of law governing local authorities.

For example, in North Carolina, public authorities are defined by the Local Government Budget and Fiscal Control Act (LGBFCA). All monies received or spent by local governmental agencies must be budgeted, disbursed, and accounted for in accordance with the Act. The LGBFCA defines two types of public authorities. The first type has, like a unit of local government, three distinguishing characteristics: 1) it is a municipal corporation; 2) it is not part of the state’s budgeting system; and 3) it does not have the power to levy taxes. The second type has five distinguishing characteristics: 1) it is not a municipal corporation; 2) it is not part of the state’s budgeting system; 3) it has no power to levy taxes; 4) it operates on an area, regional, or multiunit basis; and 5) it is not a part of the budgeting and accounting system of a unit of local government.^[13]

ASSESSING PUBLIC AUTHORITIES AS A FORM OF GOVERNMENT

Public authorities can be considered a subset of special districts, the fast-growing form of government in the United States. For this reason, they can be viewed as one of the largest contributors of governmental fragmentation since the Great Depression.^[14] They have flourished as a form of local government, particularly in the South, because they have the fiscal advantage of shifting fiscal responsibilities for a costly function(s) to alleviate or stabilize pressures on the finances of other local governments. They are also a way for local governments to expand their revenue sources, particularly in urban areas. Cities and counties can also be constrained by the failure of state governments to equip local general purpose governments with the fiscal powers that allow them to cope with urbanization. The creation of public authorities allows local governments to get around state constitutional and statutory restrictions on taxing and borrowing powers.^[15]

Governmental fragmentation in the United States has both benefits and costs. It produces allocational efficiency, allowing residents to vote with their feet and choose their tax price. As mentioned, efficiency arguments can also be rooted in the progressive belief

that separation of politics and administration promotes efficiency that other governments do not possess. Public authorities have the ability to operate more like a business and can ignore electoral politics of other types of special districts and cities and counties. Other efficiency advantages for area-wide authorities include the achievement of economies of scale.^[16]

Governmental fragmentation has also been criticized for producing inequities among localities. However, area-wide public authorities can be a more equitable system of financing a particular service and may provide for more even service provision in a particular region than the traditional city or county. A public authority is often needed to solve region-wide problems, such as pollution or traffic congestion where a lack of cooperation exists among municipalities. They can be created for a single purpose and encompass an area in which cities and counties cannot stretch beyond their traditional boundaries to solve a problem, and political barriers exist to develop an inter-local government agreement on providing and financing a particular service.

In terms of accountability, adding another special purpose government in a metropolitan area further complicates matters by adding yet another governmental entity. Citizens often do not understand who is accountable for the services authorities provide. As public authorities have similar powers to state and local governments, but are typically politically insulated from elections, they are seen as undemocratic, inequitable, and unresponsive to citizens.

However, critics claim that public authorities enjoy too many freedoms from the rules and regulations affecting general purpose governments and that they should be controlled or effectively managed by parent governments or oversight bodies. Over the years, various types of these entities have been accused of inherent policy biases.^[8] Others have accused authorities of subverting the economic preferences of voters because they are not controlled directly by the electorate.^[17] Several public authorities have also come under fire for financial mismanagement and inefficiency.^[15]

CONCLUSIONS

Public authorities as a type of special purpose government remain one of the fastest growing forms of government in the United States. They share the features of both governments and businesses, and are responsible for providing and financing a myriad of services such as providing public transportation, regional economic development, and public housing. Their unique hybrid structure insulates and exempts them from the normal restrictions imposed on other

local governments and firms. This gives them the ability to operate in a more flexible and often desirable manner. Yet they are not without their critics. Public authorities are often viewed as undemocratic, unresponsive, and beholden to special interests.

After reviewing the definitions, costs, and benefits of public authorities as a public policy institution, a pressing question remains. How can outside accountability be increased without compromising the flexibility, independence, and efficiency required to conduct a business-like organization? The answers to this question are important and require more research because public authorities play an integral role in the policy process in the United States. They are powerful quasigovernment institutions that influence the implementation of economic and social policy. Public authorities affect how millions of Americans receive services such as mass transit, housing, and provide power. In the future, we can expect public authorities to continue to expand in numbers, increase the amount of money they borrow, and expand into new policy areas such as the justice system and welfare system. Yet there are few studies in the area of public authorities and policy implementation. There is also little emphasis on education and training of public authority managers in the field of public administration. In fact, we do not know much about these organizations and who manages them. As Mitchell reminds us: "Research on how to effectively control authorities is important for improving the legislative oversight process."^[2] This is critical to understanding accountability of these quasigovernmental institutions.

REFERENCES

1. Smith, R.G. The web of actors in authority policy implementation. *Policy Studies J.* **1990**, 18 (4).
2. Mitchell, J. The policy activities of public authorities. *Policy Studies J.* **1990**, 18 (4).
3. Axelrod, D. *Shadow Government*; John Wiley and Sons, Inc.: New York, 1992.
4. Frant, H. Reconsidering the determinants of public authority use. *J. Public. Adm. Rev. Theory* **1997**, 74.
5. Doig, J.W. If I see a murderous fellow sharpening a knife cleverly...the Wilsonian dichotomy and the public authority tradition. *Public Adm. Rev.* **1983**, 43, 292–304.
6. Leigland, J. Public authorities and the determinants of their use by state and local governments. *J. Public. Adm. Rev. Theory* **1994**, 4, 521–544.
7. Sbragia, A.M. *Debt Wish: Entrepreneurial Cities, U.S. Federalism and Economic Development*; University of Pittsburgh Press, 1996.
8. Walsh, A. *The Public's Business: The Politics and Practices of Government Corporations*; MIT Press: Cambridge, MA, 1978.

9. Dimock, M.; Dimock, G. *Public Administration*; Holt, Reinhart and Winston: New York, 1969.
10. Smith, R.G. *Ad Hoc Governments*; Sage Publications: London, 1974.
11. Gerwig, R. Public authorities: legislative panacea? *J. Public Law* **1956**, 5, 387–407.
12. Rosenbloom, D.; Kravchuk, R. *Public Administration: Understanding Management, Politics and Law in the Public Sector*, 5th Ed.; McGraw Hill: New York, 2002.
13. Lawrence, D. *Local Government Finance in North Carolina*, 2nd Ed.; Institute of Government: Chapel Hill, NC, 1990; 102–103.
14. Advisory Commission on Intergovernmental Relations. In *State and Local Roles in the Federal System*; U.S. Government Printing Office: Washington, DC, 1982.
15. MacManus, S.A. Special district governments: a note on their use as property tax mechanisms in the 1970s. *The Journal of Politics* **1981**, 43 (4), 1207–1214.
16. Henriques, D. *The Machinery of Greed: Public Authority Abuse and What to Do about it*; Lexington Books: Lexington, MA, 1986.
17. Bennett, J.T.; DiLorenzo, T. How the government evades taxes. *Policy Rev.* **1982**, 6 (Winter), 71–89.

Public Budgeting, Politics in

Irene Rubin

Division of Public Administration, Northern Illinois University, DeKalb, Illinois, U.S.A.

INTRODUCTION

Public budgeting is intrinsically political. Budgeting is necessarily political for numerous reasons. The demand for public projects and programs always exceeds the supply of dollars, so there is always choice involved in any budget. Someone's request will be honored; someone else's will be cut back, ignored, or delayed into future budget years. Some degree of prioritization must take place, even when revenues are relatively plentiful, which means a process must be in place to make those allocation decisions, criteria must be used to make those decisions, and decision makers are going to have influence on those decisions. Similarly, taxes and fees must be collected to pay for public services, which means that decisions must be made about how that burden will be distributed, that is, on whom the burden will fall. Decisions on spending and taxing can be made more openly, with more public participation, or more secretly, between governmental and non-governmental elites with little accountability to the public. In a democracy, the budget is a prime tool for public accountability. In that role, the budget process may function well or poorly.

Budgetary politics results in series of tensions between technical neutrality and political policy, between ideology and pragmatism, between concerns for institutional processes and requirements and short-term partisan gains, between across-the-board and targeted allocations, and between openness to all and openness to a few wealthy individuals and interest groups. How these tensions can be resolved, whether the right blend can be achieved, and whether, if achieved, it can be maintained over time, are all highly problematic. Each budget cycle, each step in the budget process, reveals a continuing struggle.

A satisfactory balance between these tensions helps to implement democratic processes, fosters accountability, and efficiently allocates public money to the solution of collective problems. If the balance shifts too far in one direction, the result can be gridlock, technocracy, or domination by wealthy interest groups, with massive citizen distrust of the public sector. Budgetary processes may encourage movement toward the extremes, with increasing negative consequences over time. Those negative outcomes eventually force

the process back toward the middle and may restructure the budget process itself.

MEANING OF POLITICS IN BUDGETING

The word "political" is used in different ways. Political sometimes means partisan, loyalty to political parties in their efforts to win elections and determine policy. In this sense, politics suggests credit taking and blame shifting across parties. It may secondarily imply party platforms, campaign contributions, and paybacks for those contributions. Partisanship is most obvious at the national and state levels, but also functions at the local level, although often detached from national level politics and sometimes not tied closely to Democratic or Republican party politics but to local versions of parties and platforms.

In another use of the word, political refers to the policy goals of elected officials. It refers to the issues of interest to the elected officials, the problems they focus on and try to address, symbolically or practically. In the sense of policy goals, politics implies a widely discussed dichotomy between policy formation and policy implementation, and the division of responsibility between elected and appointed officials in coming up with policy proposals and shaping them during implementation.

Politics is also used to refer more broadly to the process of articulating and resolving the conflicts between priorities of different segments of society. Among the decision makers are ideologues, both appointed and elected, who have strong opinions about particular solutions or class outcomes, and pragmatists and compromisers, who try to keep the level of conflict down, providing something for everyone, and working to keep all parties feeling like they have an effective voice in government. The process of interest articulation and satisfaction maybe more open, with widespread and easy access, or it may be relatively closed, with a markedly uneven playing field that in no way approximates democratic participation. Open conflict may be encouraged or suppressed.

These three views of politics all come into play in public budgeting. They manifest themselves in five

main tensions. The first is between political policies and technical and professional neutrality. The second is between political partisanship and institutional requirements, the processes necessary to keep the institutions of government working. The third is between ideology, class politics, and symbolic politics, on the one hand, and pragmatic approaches and conflict reduction, on the other hand. The fourth is between incrementalist or across-the-board solutions and more targeted approaches, implying prioritizing and trade-offs. The fifth is between more democratic or open processes and more elitist or closed decision making. This last one emphasizes the degree of access of citizens, interest groups, elected officials, and professional staff and bureaucracies.

Policy Versus Professional Neutrality

The policy goals of elected officials sometimes clash with proposals and analyses coming from professional staff in budget offices and department or program heads. With respect to budget offices, elected officials have policy preferences they want to see implemented in the budget and want the budget office to include those preferences in the budget proposals and, if necessary, reduce or reject proposals that do not conform to these policies. For their part, budget staff want to present unbiased analyses and recommend funding for the legitimate needs of agencies or programs, even if they sometimes do not mesh with the policy priorities of the elected officials. The clash between departments and elected officials may take the form of politicians micromanaging departments or departments framing policy issues in such a way that elected officials have little say.

Compromise would suggest that major policy priorities be accommodated, but that important requests coming up from departments, including new initiatives, also be given serious consideration. Sometimes either the departments dominate to the exclusion of political priorities or political priorities dominate to the exclusion of departmental needs. Most of the time, the balance between bottom-up and top-down budgeting is closer to the middle, but may shift over time.

In recent years, the top-down policy orientation of the federal Office of Management and Budget has dominated over a more bottom-up approach. To the extent that the legitimacy and respect granted to budget examiners comes from their detailed program knowledge and ability to recognize when a proposal is necessary and appropriate, when they lose that expertise, the budget examiners become mere implementers of policies determined by elected officials.^[1]

Although elected officials may want the budget office to be responsive only or primarily to their policy

concerns, the office must maintain some degree of neutrality to function effectively and credibly. It must not become so allied with one elected official or faction that it ceases to be acceptable to a successor, or in the legislature where several parties may be present simultaneously, acceptable to both at the same time. The entire budget process hangs on the quality of revenue estimates, and numerous policy issues, such as the size of tax reductions or tax increases, hang on the estimates of the size of deficits or surpluses. If these numbers are seen to be distorted to suit policy preferences of one elected official or group, the entire process may break down.

The result is a tension between pressure from elected officials to come up with particular policy analysis and recommendations, and counterpressure from the budget and other staff offices to maintain a necessary level of professional integrity. The budget and staff offices may yield at some times to extreme pressures, and at other times, revert to more professional modes of analysis. Politicians may intensify pressures at some times and back off at others, or “go shopping.” Shopping means looking for the budget or policy shop that will come up with the desired analysis or recommendations. As a result, much budget information is slanted, putting a premium on budget analysis that is not slanted or is minimally slanted toward given policy outcomes. Gaining and keeping a reputation for such neutrality can be a major strategy for heads of budget and policy research shops.

Three examples illustrate the pressures involved, and how some budget and data provision agencies have responded. One involves the General Accounting Office (GAO), a second the Congressional Budget Office (CBO), and the third the Bureau of Labor Statistics (BLS).

The GAO is a watchdog group for Congress. Over the years, it has generated some of its own projects and responded also to congressional requests for studies. In response to intense antagonism on the part of many in Congress to the agency and deep budget and staffing cuts, the GAO cut back its self-generated studies to almost nothing, working only on Congressional requests. Because the GAO responds more to the committee chairs than to the ranking minority members, it disproportionately serves the agenda of the majority party in each house. Thus, it has positioned itself closer to the end of the continuum of policy responsiveness than professional independence. However, it has also devised for itself a prioritization scheme so that it can still act independently on issues it deems crucial. The history of the agency suggests that the proportion of self-generated studies has varied, and will continue to vary in the future, but it was least able to do independent studies during a period of intense threat to its staffing and resources.^[2]

The CBO performs various functions for Congress, including estimating the costs of bills, the impact of bills on state and local government, compliance with budgeting rules, and estimates of the growth of the economy. It is also responsible for estimating and tracking the size of the deficit or surplus. Many Republicans in Congress endorsed a policy of large tax reductions and wanted the CBO to determine estimates of the deficit or surplus that would support such a proposal. To tweak the estimates of the growth of the economy necessary to push up estimates of the surplus, they pressed the CBO to adopt some form of dynamic estimates, that is, estimates of the positive effects of tax cuts on the growth of the economy and hence the growth of revenue. Modest assessments of such effects are included in the president's budget proposal and in Congress's budget resolution, but these legislators wanted to exaggerate the effect because they wanted to adopt a particular policy. They pressured CBO in various ways to get a larger number. These pressures included appointing a head of the office who would support dynamic estimates and legislating a requirement that CBO offer Congress dynamic estimates. Although there was no agreed-on methodology for estimating dynamic effects, the CBO complied by increasing the estimates of the fiscal dividend, the economic stimulus that was supposed to result from a balanced budget. These numbers kept going up, and then began to come back down. The CBO put the estimates into its reports as Congress demanded, but also put its own estimates into its reports, arguing that its own figures represented a more likely outcome. Congress ended up deciding on a process called directed scoring, taking the estimates of the CBO or the OMB, whichever it liked better. The CBO thus maintained its reputation for integrity but lost some of its influence over decisions, as Congress took whichever analysis fit its policy preferences most closely.^[2]

The BLS is located in the Department of Labor. Its responsibilities include the collection and reporting of various statistics on business and the economy. One of its tasks is to regularly update the Consumer Price Index. Because many of the pensions that the federal government owes, including social security, are indexed to the Consumer Price Index, a slower rate of growth in the index would reduce spending and help to balance the budget. During the years of massive efforts to balance the federal budget, this strategy seemed like a politician's dream, a way of cutting expenses without taking any responsibility for doing so. The problem was that the BLS was jealous of its reputation for professional integrity and refused to amend the index without the appropriate level of research to justify the changes. Some members of Congress wanted the changes quickly, regardless of the underlying research, and were willing to use a range

of pressureful tactics to get their way. Those tactics included convening an "independent" commission (the Boskin commission) to estimate the "true" rate of inflation, threatening to take the decision out of the hands of the agency and to cut their budget.

Responding to Federal Reserve Chairman Alan Greenspan's observation that reducing the index would make it easier to balance the federal budget, Speaker of the House Newt Gingrich tried to force quick changes on the BLS. "This weekend, when asked about [Alan] Greenspan's comments, the Speaker of the House said that he would give the Bureau of Labor Statistics people '30 days to get it right'^[13] or he would fire them and give the job to the Fed." Gingrich's position was that if BLS did not revise the index downward quickly—presumably without technical study—they would lose the responsibility for making the changes. He would give the job to the agency or organization that agreed with his policy position. Mr. Gingrich reportedly threatened to cut the BLS out of the budget if it did not make the changes in 30 days (Congressional Record, Senate, January 18, 1995, Page S1125).^[2] In the end, the BLS did gradually revise the index downward, not by as much as the Boskin commission recommended, but substantially. The agency continued to do the underlying research to adjust the index, asking for more money to do the research, but it was not clear that the agency was examining sources of increased inflation as well as decreases.

The history of the BLS indicates that the agency has been subject to political pressure, but that those periods have given the agency a determination to resist that pressure, stay out of the political limelight, and focus on neutral analysis. In this case, however, the BLS was unable to avoid a policy controversy. It did what it could to maintain its professional status, but probably did not come out completely unscathed. Relative to the level of pressure it was enduring, it remained fairly strong and independent, helped by a structure that granted its commissioner a long-term insulation from political pressures.

These three cases indicate that maintaining professional neutrality can be a struggle and agencies succeed more or less at different times. However, gaining a reputation for independent analysis and not following policy dictates gives an agency an extra importance. Some agencies find that there is little to be gained by yielding to political pressures and that they do best by standing tall. If they are taken out of the decision-making loop, at least they can come back another day with their integrity intact. However, if it looks likely that elected officials will make good on some of their more extreme threats to the agency, it may have little choice but to cave in on that issue, for the time being.

Although the relations between budget or data offices and political officials suggests that agencies and the whole decision-making apparatus are better off if the staff agencies are allowed to maintain their professional independence, the situation may be different in the relationship between elected officials and line department or program offices. Sometimes the requests coming from the programs and departments are technical, but sometimes they are policy related or geared to the survival and growth of the agency. On the one hand, the Defense Department bitterly resents the intervention of legislators in their business, dictating particular weapons systems or particular companies or locations for acquisitions, but, on the other hand, it may itself propose weapons and contractors with the idea of winning support from legislators who are from particular states or regions affected by the contracts. The elected representatives may engage in legitimate policy making or less legitimate micro-management, whereas the agencies may engage in technical implementation or independent policy making, thus bucking congressional priorities. Absence of a reasonable balance is likely to generate conflict until a better balance can be achieved.

In Kansas City, Missouri, the budget office and departments were relatively strong with respect to the elected officials. Their position was generally that they were operating with marginal and inadequate budgets, doing the best they could, and that any further guidance or cuts from the political officials were unacceptable. They did not trust the elected officials to make the right decisions or to see matters as they did. They reportedly made most of the policy decisions themselves, without informing the elected officials, informing them after the fact, or offering them alternatives that represented a narrow range of options, including the one they wanted and some other completely unacceptable choices. Elected officials, led by the mayor, resented being excluded from major budgeting and allocation decisions, and thus ordered a study of the process and requested recommendations for changes that would include them in more meaningful ways. The Kansas case reinforces the idea that, if the situation becomes too unbalanced, pressures will develop to reestablish a more satisfactory distribution of policy-making power.

Partisanship Versus Institutional Loyalty

Partisanship means loyalty to party or faction, as opposed to loyalty to processes or the institution as a whole. This division applies to both staff and elected officials. Partisanship means efforts to win short-term gains to look good to the voters and to make the other side look bad. It is often associated with symbolic

stances that appeal to a particular faction of voters. These symbolic stances may be extreme, be resistant to compromises, and represent unworkable solutions to problems.

Snow and Willoughby discussed the distinction between partisan staff and staff who work for the whole organization at the state legislative level.^[3] They argued that the staff who work for the whole organization deal with issues that affect the whole institution and its prerogatives. The example they give is of a governor who is proposing to merge numerous line items. Such a merger is likely to enhance the decision-making power of the governor vis-à-vis the legislature. The institutionally oriented budget staffers would keep abreast of this issue and call it to the attention of the legislators.

Partisanship in Congress is sometimes in direct opposition to the requirements for the institution to function smoothly. One extreme example of partisanship, as opposed to institutional requirements, occurred in the federal government in 1995/1996 when neither political party was willing to vote for a budget compromise, shutting down the federal government, which is not allowed to spend money unless there has been money appropriated for it. Institutional requirements allow for compromises and passing the budget in a timely way each year; partisan concerns may seriously interrupt that flow. Other examples have occurred when legislation was included in bills by reference, resulting in huge bills with major content that legislators did not have time to read and which were described in the text for examination and discussion.

Joe White illustrated the tension between partisan concerns and the requirements for routine passage of legislation in his discussion of earmarking in the appropriations committees in Congress. He argued that normally appropriations committees take into consideration the interests of the minority as well as the majority party in the distribution of pork or earmarked spending, but that the leadership, which is intensely partisan, often presses for more exclusive distribution of pork and earmarks for the majority party members in key districts for key elections. Appropriations committee members know that the earmarks will not survive unless there is a coalition on the floor to pass them, and hence pass around the earmarks to create that majority, a requirement that generates a certain amount of civility. If they award pork more exclusively along party lines, they risk losing their own pork by riling up the opposition who have nothing to lose from attacking.^[4]

In recent years, partisanship has often created a kind of brinkmanship that prevents timely passage of the 13 money bills that are supposed to pass by September 30 of each year and encourages budget

tricks that thwart the intent of the rules. The credibility of the process suffers as a result.

in recent years, Congress has been unable to complete all of them, resulting in last-minute deals that evade, break and ignore spending caps called for in the budget resolution by resorting to various budgetary devices, such as designating large amounts of spending as “emergency” spending.

“Once you go down the road of winking and nodding through the process, you erode the credibility of the process,” says Richard E. May, a principal at Davidson & Co. and a former Republican staff director of the House Budget Committee. The last-minute haggling results in helter-skelter lawmaking, one analyst argues. “Lately, the closing days of the session have deteriorated into a very costly and unstatesman-like cross between a food fight and a game of budgetary chicken in which the aim of each side seems to be to inflict maximum political embarrassment on the other while getting as much as possible for one’s own spending or tax priorities,” Martha Phillips, former executive director of the Concord Coalition, a group that advocates fiscal restraint, told the House Rules Committee earlier this year.^[5]

Budget staffer Bill Dauster called attention to various procedural rules violations in the mid-1990s, after a major influx of ideologically motivated Republicans. He argued that such violations occurred because the Republicans sought short-term, partisan advantages. His list of rules violations included removing the standing rule to limit conference committees to same subject sent to the conference.

To secure last-minute passage of a version of the Federal Aviation Authorization Act that included an extraneous provision of special interest to the Federal Express Corporation, the Senate voted 56 to 39 to overturn the Chair and nullify the subject-matter rule. Now small groups of Senators, like the five men who sat on the conference committee on the FAA bill, can add any new matter they please. Now, Senators have no recourse if the conference committee exceeds the scope of what the Senate committed to it. In Senator Kennedy’s words: “They will acquire the power to legislate in a privileged, unreviewable fashion on virtually any subject. They will be able to completely bypass the deliberative process of the Senate.”^[6]

Dauster listed numerous other serious procedural changes engaged in for short-term partisan gain. Then he quoted Senator Kennedy’s warning about the change in conference reports:

This . . . is a vote about whether this body is going to be governed by a neutral set of rules that protect the

rights of all Members, and by extension, the rights of all Americans. If the rules of the Senate can be twisted and broke and overridden to achieve a momentary legislative goal we will have diminished the institution itself.^[6]

Dauster concluded:

This past year, the Republican Senate majority has thus repeatedly arrogated greater power to the majority to use fast-moving legislative trains to move unrelated matters of its choosing. This majority has thus repeatedly torn down the protections that the Senate affords minorities. This majority has shown a disturbing willingness to cast aside long-held precedents and institutions to serve immediate policy ends. That, then, is the unifying theme that runs through the procedural actions that the Republicans took last year and that they seek to take this year. The Republican Senate stands ready to sell its birthright for a mess of pottage. It shows contempt for the institution. It focuses solely on the immediate political gain.^[6]

An article in *Govexec* underscored several recent budget gimmicks used to circumvent legitimate and necessary budget rules. Some of them were evasions of budget caps probably agreed to by both parties, but some seem to have been used for partisan gain.

Back in 1995, Republicans tried to insist that the Clinton administration submit a budget that would be certified as balanced by the Congressional Budget Office. The administration argued that a balanced budget certified by the Office of Management and Budget should have been good enough. Now, OMB appears good enough for Republicans—sometimes. Using “directed scoring,” budgeteers are able to choose which office scores certain programs. This allows Congress to choose either CBO or OMB estimates—whichever office says the program costs the least. House Democrats charged earlier this year that Republicans used at least \$14 billion of OMB scoring in determining the costs of programs. “They’re using two sets of books, but they’re not telling us which set of books they’re using on what occasion,” Obey charges, adding that, “One day it’s the OMB. One day it’s the CBO. One day it’s the Wizard of Oz.”^[7]

Another procedure absolutely necessary for the intelligibility of the process is the use of baselines. Baselines allow participants to see whether budget allocations are going up or down.

There are disagreements as to how baselines should be calculated, but as long as there was technical disagreement over relatively similar baselines, the results were easily interpretable. But the argument became partisan, with some advocating baselines that included inflation and others baselines that excluded inflation,

avored agencies' baselines were calculated differently than other agencies, and the result became muddled to say the least (refer to Ref.^[8]).

Some analysts argue that the partisan and short-term gains will always win out when pitted against the longer-term institutional and process requirements, but the negative consequences of eroding comity, delaying decisions, and continuous finger pointing sometimes have the effect of shifting the balance in the other direction. Many of the procedures that Dauster saw falling by the wayside lasted many decades, providing evidence that process requirements have a kind of vitality of their own. Partisan fights may dominate for periods at a time, and then die down.

Ideology Versus Pragmatism

Partisanship does not necessarily imply ideological or symbolic stances. Advancing party interests may involve passage of legislation and compromises, as well as credit taking, blame casting, and brinkmanship. But in budgeting, as in politics more broadly, elected officials are sometimes interested in ideological positions and/or symbolic stances rather than compromises and legislative solutions to problems. Such stances are likely to exacerbate conflicts. Although partisan squabbles seem like street fighting, ideological battles mobilize followers in ways that may seem more like wars.

In issues of taxation, those who are taxed make demands on politicians to reduce their taxation. That can be done either across the board, for everyone, for the wealthy at the expense of the poor, the poor at the expense of the wealthy, or for selected smaller groups or individuals who have negotiated political access. Some of these proposals may take on an ideological tone. Tax cuts across the board are expensive, reducing the revenue available for expenditures, and thereby increasing the size of deficits or shrinking the scope of government. The latter has been a goal of conservatives for some time. Democrats try to shift the burden of taxation more to businesses and the wealthy, Republicans more away from the wealthy onto the middle and lower classes. Ad hoc sets of exemptions can add up to considerable revenue losses, and because achieved through the progressive tax structure at the federal level, often provide disproportionate benefits to the well to do.

Citizens for Tax Justice, a respected, left-of-center advocacy and research group, charged that "profitable corporations that will receive a total of \$7.4 billion in immediate Alternative Minimum Tax (AMT) rebates if the economic 'stimulus' bill recently passed by the House of Representatives becomes law are also major campaign contributors, giving \$45.7 million to federal elections since 1991." The organization charged that

"these companies and others use campaign money to manipulate the political process to win huge tax breaks unavailable to ordinary Americans."^[9] Citizens for Tax Justice tried to make a case that these tax breaks were class based and not available to individuals.

Some political candidates and elected officials have argued for a flat tax at the federal level, ending the graduation that results in wealthier people paying a relatively higher tax burden than poorer people. In the United States, some have graduated incomes taxes, whereas others have flat taxes, showing a distinctly different profile of class politics.

Ideological and symbolic positions have been important in campaigns to balance the federal budget, as some elected officials supported proposals for a constitutional balanced budget amendment. Many of the supporters of such an action wanted to combine it with revenue limits in an effort to reduce the scope of government. This position can be considered an ideological one, rather than a more technical one requiring budgets to balance. When the budget was finally balanced in 1998, ideologues had to shift to buying down the debt to keep the pressure on government to reduce spending and the scope of government.

The line between pragmatic approaches and ideological ones can be overdrawn, there is a point at which an accumulation of smaller, less symbolic legislation may well add up to an ideological platform, but there are differences. Ideological stances need to be pure and clear, and hence avoid many of the compromises necessary for passage. Pragmatic approaches may be so compromised or so counterbalanced by broad distribution of benefits that their ideological content is lost.

Incrementalism Versus Tradeoffs

Incrementalism refers to a kind of across-the-board allocation, in which each department or program gets a proportionate share of additional revenues available to be spent. Incremental approaches avoid conflict, eschew policy concerns, and are not really about uncovering or solving collective problems. They work best when there is a steady but moderate increase in revenues and the existing allocations reasonably match public expectations and concerns. They function less well in more constrained revenue situations, where trade-offs are necessary, and some criteria for making the trade-offs is often desirable.

Because revenue is constrained, the stress between the relative comfort of incremental allocations and the relative pain but increased efficiency of targeted cuts is often apparent. A common form of budgeting at state and local levels is called target-based budgeting. In target-based budgeting, the central budget

office and the chief executive officer give a target (a maximum) for the request from each department. These requests are set in such a way that the combined total of the requests will be lower than expected revenues, creating a pool of resources, within tight constraints, for reallocation. The process of making the cuts in each program requires some kind of prioritization, and the allocation of the savings also requires some kind of prioritization. The possibility exists of cutting one department a bit and giving more to another, changing the budget priorities to match policy shifts, but this opportunity is often not taken seriously. The power of the departments puts enormous pressure on the central budget office to continue to support them much as they always have. There is some prioritization within departments, but not usually much reallocation between departments. Prioritization may take place in some circumstances, but the fallback position, the more comfortable position, is a more across-the-board approach.^[10] Some who use target-based budgeting have suggested that there is a constant battle to prevent it from becoming just another incremental budget allocation technique.

Trade-offs—increasing spending on one department or program and reducing it elsewhere—occur in every budget to some extent, but such trade-offs become more visible when there are tight, widely publicized ceilings on expenditures. Trade-offs may occur without much prioritization, however. At the national level, when the federal budget was constrained by the spending caps on the discretionary side of the budget (from 1990 to the period of budget balance beginning in 1998) trade-offs were made frequently, but they were framed by the committee structure and by departmental programs. Secretaries were told that, if one expenditure was increased, then they had to make reductions somewhere else to stay under the caps. Appropriations subcommittee members could not raise the spending of one department under their jurisdiction without reducing the spending of other departments in their jurisdiction. What became relevant was the jurisdictions of each subcommittee, which in some cases did not represent a logical set of comparisons. Moreover, the caps called for limits and hence offsetting reductions, but did not require any criteria for prioritization or any formal prioritization process. Some agencies experienced budget cuts because they were less popular than the other programs in a subcommittee's jurisdiction. If they been compared with other departments or programs, they likely would have fared better.

That is not to say that budgeting never involves formal or informal prioritization, only to say that it is likely to occur only at intervals or at the margins. Some communities do a form of zero-based budgeting that incorporates a list of priorities and compares proposals with that list. Others require council or board members

to state for each program how important it is to them to expand it, contract it, or keep it the same.

Open or Closed Decision Making

The process of budgeting itself is laden with politics, as decision makers jockey for power to determine increases or decreases. One key element of that process is its openness to individual citizens and interest groups. Ideally, the process is open to citizens who can express their priorities and see them incorporated into the budget. They should also get to see the results of their spending in accounting reports and performance reporting. Because individuals may not have the time or knowledge to participate directly and because the federal government is removed in space from most citizens, this citizen involvement may be indirect, through various interest and advocacy groups. However, openness to such organized groups may or may not add up to democratic participation, depending on the openness of the decision making to all interests or only some. If interest groups primarily buy their influence, then poorer groups are likely to be left out. Elected officials are interested in votes as well as money, so they may remain open to smaller and poorer groups if they represent a major constituency group, but those with more money may at times be granted more access.

From the point of view of political power, the question has to do with the levelness of the playing field—does each group have the same kind of access to budgetary decision making? One attempt to make the playing field less level was called the Istook amendment (there have also been numerous variations of the Istook amendment). After much opposition from the not-for-profit charities who would have been hurt by it, the measure failed. Its goal was to prevent not-for-profit groups and charities, which had received federal funds lobbying government or influencing programs. Because not-for-profits have to apply for tax-free status and show that they are not lobbyists, they have to record their lobbying efforts and keep them below a fixed level. Those requirements can be made more burdensome and confusing or less so. The more burdensome, the more difficult for smaller organizations to have any input on the legislation that designs and funds the programs they run, normally at considerable savings to the public.

Advocates for the non-profits described how this proposed legislation would affect them:

Nearly every charity pointed out that service delivery and advocacy go hand-in-hand. The Catholic Community Services in Salt Lake City, Utah, feeds over 500 people a day in their homeless shelters, receives federal grants, and views itself an advocate for homeless

families. The Vermont Public Transportation Association provided over 420,000 trips to doctors' offices and hospitals, receives federal grants, and views itself an advocate for seniors and those with disabilities. The Big Brothers/Big Sisters in Delaware receives a combination of federal and state grants for a variety of projects including one to establish a mentor relationship between volunteers and children from single-parent families. Yet they, too, are active in advocating on youth issues, including child abuse.

But because the Istook Amendment is so technical and confusing, they would be afraid to do any advocacy. One senior citizens group in Michigan, run by an 85 year old man, said he could not even understand one provision in the Istook Amendment that dealt with the definition of an "affiliated organization"—yet he knew he was covered. And with the "bounty hunter" provision that allows anyone to sue a federal grantee for up to 10 years for alleged non-compliance, most charities are fearful they will be sued—even when trying to faithfully comply.^[11]

In the case of the Istook amendment, there was an intentional effort to make it more difficult for those providing services to the poor to advocate for their clienteles or programs. Although Istook was followed by numerous others, all the efforts failed. The results are less optimistic though when looking at tax policy, where it seems clearer that big contributors get more than their share of tax breaks, and get far more back in tax dollars than they spent in campaign donations. In this area, it seems that the playing field is not level. Supporters of this system would probably argue that they are the ones being taxed, so naturally they are the ones who need the tax relief. Still, because the benefits are tailored to big contributors, there are others who would want and maybe need tax relief who do not get it.

Joe White presented this tension between representing everyone in an open fashion and representing only organized interests with money in his discussion of interest groups and the appropriations committees. He favors the side of openness when discussing earmarks or pork allocations:

A former House subcommittee chair expressed this preference well when he commented that, "I think there is a sense of accomplishment on the part of a member who is approached by constituents—generally it is local government types, the town council, the school board, the airport authority—who have a problem and they think federal money will help... If it's wasteful, if it's not wanted, they'll hear about it... if you do an \$800,000 innovative housing project, it won't change the face of that town. But it tells the town, 'we think what you're doing has merit, it might be seen as a model.'" So the member and the

community both feel pretty good. They had a specific need, brought it to Congress, and it got fulfilled. It says the system works, their voice is heard in a faraway place called Washington, and it got done. And that's representative government.^[4]

White's interviewees argued that it was not only constituents who stood to gain jobs who caught legislators' attention, but also the local elites. Further, White argued that legislators want to feel good about a project, feel that they have done something good. White concluded that "passion can be a relevant resource for interest groups: the intensity of demand can affect the legislator's sense that meeting a demand is doing good." Based on White's interviews with legislators, budget processes and decisions are open at the level of earmarking of funds on appropriations committees. At the national level, one would expect decision making to be less open on decisions about priorities and overall program funding.

At the state and local level, much effort has gone into including the public in budget planning activities and goal formation, but this responsiveness appears to be cyclical. In Phoenix, the city manager reported that the city had alternated between domination by business interest groups and broader public participation. He offered the explanation that the business domination, when it occurred, was too extreme and set off a reaction among the public who felt unrepresented.^[12]

CONCLUSION

Politics in budgeting can best be viewed as a series of tensions, which may be reasonably balanced or become seriously unbalanced, with major consequences. Elected officials may be denied their legitimate right to determine policy; monied interest groups may buy their way into less fair distribution of government benefits, thwarting democratic processes; budgets may be slow to adapt to changing priorities; and short-term partisan advantages may disrupt processes and institutions, leading to massive public distrust. Political scientists need to pay more attention to what leads to imbalances, which budget processes encourage excesses and which curtail them, and how to reestablish balance once it has dissolved into acrimony and exclusiveness.

REFERENCES

1. Tomkin, S. *Inside OMB*; M. E. Sharpe: Armonk, NY, 1998.
2. Rubin, I. *Balancing the Federal Budget: Eating the Seed Corn or Trimming the Herds*; Chatham House Publishers: New York City, 2002; forthcoming.

3. Snow, D.; Willoughby, K. *Legislative Fiscal Staff Influence in Legislatures: A Model of Decision Making*; Association for Budgeting and Financial Management: Washington, DC, January 18, 2002.
4. White, J. Making connections to the appropriations process. In *The Interest Group Connection*, 2nd Ed.; Herrnson, P., Shaiko, R., Wilcox, C., Eds.; Washington, DC: CQ Press, 2004.
5. Baumann, D.; Cohen, R.E. Budget breakdown. Govexec **2000** (October 1), (<http://www.govexec.com>, accessed January 28, 2002).
6. Dauster, B. Stupid budget laws. Talk presented to the American Association of Law Schools. January 5, **1996**. (Posted on OMBwatch, <http://www.ombwatch.org/ombwatch.html>, accessed 1999).
7. Baumann, D. A guide to the gimmicks. Govexec **1999** (September 1), (<http://www.govexec.com>, accessed January 29, 2002).
8. Schick, A. *The Federal Budget: Politics, Policy, and Process*; Brookings: Washington, DC, 1995.
9. Center for Tax Justice; Institute on Taxation and Economic Policy; the Public Campaign. Buy now, pay later: How generous corporate campaign donors save billions in taxes. **2001**. (<http://www.ctj.org>, accessed January 28, 2002).
10. Rubin, I.S. Budgeting for our times: target based budgeting. PBF **1991**, *11* (3), 5–14.
11. Let America speak. Handcuffing America's charities. **1995** (<http://www.ombwatch.org/las/1995/handcuff>, accessed January 28, 2002).
12. Rubin, I. *Class, Tax, and Power: Municipal Budgeting in the United States*; Chatham House: Chatham, NJ, 1998.
13. Abraham, K. *Statistics in the Spotlight: Improving the Consumer Price Index*, Presented at a meeting of the American Statistical Association, Chicago, IL, August, 6, 1996.

Public Choice

Darrene L. Hackler

*Department of Public and International Affairs, George Mason University,
Fairfax, Virginia, U.S.A.*

INTRODUCTION

For those who agree or disagree over the assumptions and policy implications of the public choice approach, one item remains non-controversial—public choice uses the economics methodology to study politics. The theoretical perspective of public choice uses economic principles to understand political organization and practices, covering matters of constitutional design, rational actions of politician and citizen, voting mechanisms, resulting policy outcomes that affect the provision of public goods, and the bureaucrat's role in government. Public choice theorists examine the political market as the public sector's equivalent institution for the private market. How governments allocate resources to provide goods and services to its citizens is a collective decision-making process. The shaping of this collective decision-making process is where public choice theorists focus their attention. How institutions are involved at each level of this process is the primary concern of public choice theory. Although the focus is on institutions, at the center of public choice theory is the individual; from government to political parties, every decision-making unit is composed of individuals.^[1]

ROLE OF THE POLITICAL MARKET

The political market is necessary to provide goods that the private market cannot produce because of the nature of the good. Public goods hold special characteristics of non-exclusion and non-rivalness. The first characteristic refers to the impossibility of excluding the consumption of a particular good by others, leading to excess demand. For example, a lighthouse benefits all boats, and the owner of that lighthouse cannot exclude one boat from its beam. Similarly, a non-rival good is one that, in the process of consuming or using it, does not change or alter the ability of others to consume or use it. Thus, free riders will accompany a non-rival good because a person can consume this good without ever paying for it. An example of a non-rival good is a web site where many people can view the content at the same time, as long as the server does not

become congested from too much traffic. These two characteristics present problems for private market production, and because a failure in production or consumption exists in the private market, public goods result from this failure.^[1]

Public choice theorists refer to these private market failures as externalities and consider externalities to be the major reason for government intervention into the private marketplace. An externality occurs when a decision or action of one person benefits or harms another such that the market price does not reflect the true costs of these actions. Public choice examines why this externality exists and attempts to address the problem through the analysis of producer's and consumer's property rights. Many problems arise during such analysis: property rights are often not clearly defined; the liability for the costs and benefits is unclear and does not affect one set of decision makers; and the externality often affects a large number of entities. Thus, public choice suggests that it is the government's role to define and enforce property rights when it is feasible, and where some resource may become extinct. The policy instruments used for definition and enforcement, from auctions to taxes and subsidies, are debated in regard to the costs that government intervention imposes on the market. However, although most economists identify private market failure and the existence of externalities as a reason for government intervention, Tullock cautioned that government involvement and the instrument used to combat an externality also produce externalities. "We are choosing between two imperfect instrumentalities. Unfortunately, that choice [among instruments] is a government decision, and [this decision] will also generate externalities."^[2]

The political market not only affects the government's role in externality cases, but it also affects the production of public goods. This production occurs within numerous governmental institutions, which in turn impacts how government will know what public good to produce—from quantity to quality. The influence of each decision and how that decision is made can result in inefficiencies. Public choice uses the principles of private market efficiency to guide the development of political market theory that addresses matters and problems such as public good production, externalities, and the absence of property rights.

HISTORY

Although many recognize Buchanan and Tullock's *Calculus of Consent*^[3] as the beginning of public choice, the genealogy can be traced to Adam Smith's *Wealth of Nations*.^[4] Smith focused on the nature and causes of the wealth of nations, specifically the efficiency of government expenditures.^[1] This was the old political economy. Throughout the centuries, economics and political science developed separately until the mid-twentieth century. The works of Black, Arrow, and Downs refocused economists on political issues.^[5–7] However, *Calculus of Consent* was the first work to explicitly integrate two economic elements—economic modeling of utility-maximizing behavior for political choices and the understanding of politics as an exchange similar to the private market.^[8] Buchanan and Tullock's book led to the formation of a journal entitled *Non-Market Decision Making*, which later evolved into *Public Choice*.

In the ensuing 40 years, two main schools of study have evolved: constitutional political economy and the theory of bureaucrats. Buchanan remains tightly bound to the former, and this course of study is often referred to as the Virginia School. This school emphasizes classical political economy and examines the effects of institutions and structures. The latter area of study examines both the normative and empirical aspects of bureaucratic theory. The Chicago School is most associated with the empirical aspect, focusing on three distinct ideas. First, bureaucrats, like all other individuals, are indeed utility maximizers such that they will support policies that convey the greatest individual benefit. Second, citizens are also rational utility maximizers and seek the greatest personal gain, denoted as methodological individualism. Third, institutional service arrangements are more efficient if multiple, smaller scale units exist. The first and the second ideas are in direct contradiction of public administration theory, which presents the bureaucrat as serving the public interest best through Wilsonian-influenced, centralized, service delivery units.^[9] Through public choice's two major schools, public choice theorists have addressed numerous political issues at all levels of government. The remaining sections examine some of the most prevalent issues.

CONSTITUTIONAL ECONOMICS AND GOVERNMENT INSTITUTIONS

In the vein of Buchanan's work, public choice from a constitutional political economy perspective questions the outcomes of the political market by examining the origins of government. The argument is that the choice of constitutional rules will define the available

choices within a constitution—operational decisions such as the size of the education or defense budget.^[1] Thus, constitutional rules are the institutions that shape a government's response to society and the efficiency of that response.

A constitution's framework also affects the composition of government and determines how politicians can represent and address their constituent's concerns. Politicians use political party platforms to appeal to voters, and individual desires form an aggregate impact on politicians through voting mechanisms and the actions of pressure groups. Finally, within government, politicians work for legislation that is beneficial to them. In the eyes of a public choice theorist, these events do not proceed without private interests outweighing public interests. Each event entails decision making in the political market. The voting rules, enforced to permit the mixture of majority and minority preferences, create further problems and biases. For example, government may gain greater discretionary power because of the incentives for rational ignorance—that it is in an individual's best interest to remain ignorant of particular issues because the costs of gaining information is greater than the benefits received. Such political market failures result because the institutions that exist are unable to allow citizens to express the intensity of their preferences for public goods.

With government as the producer of public goods and citizens as the consumers, rational actors represent both sides of the political market. Just as for bureaucrats, public choice holds that citizens have preferences for the provision of public goods, demanding a specific quantity, quality, and desired mix of these public goods. The central problem in the political market is how citizen preferences are realized and expressed. Public choice holds that only efficient government institutions will enable the conveyance of this information such that government will produce the appropriate amount and mix of public services. Thus, the institutions and mechanisms established to gain information from citizens are the intellectual fodder for public choice theorists, and especially for those constitutional political economy theorists. A general contention is that the mechanisms—from voting, representation, referendum, and opinion polling to political party platforms—do not enable individual realization of preferences, nor do they efficiently reflect individual preferences in a collective way. The political market is also at the mercy of impediments that also affect the private market. Information is not perfect or free, so individuals rationally abstain from trying to understand political debates and candidate positions. Coupling this with political parties in which one candidate represents multiple political issues and inefficient voting systems, public choice theorists

believe that the mechanisms must provide lower costs to the voting citizen.

The concern for efficient mechanisms leads to another prominent issue within public choice theory—rent seeking. Within economics, a rent is the amount paid for a resource that is fixed in supply in both the short and long term. Wealth is accumulated from this resource because its supply is restricted, particularly if market conditions are not competitive. Rent seekers desire government regulation that promotes barriers to entry, and they will engage in activities that use numerous resources inefficiently. Rent seekers employ the time and other resources of law, marketing, and financial experts to lobby for favorable government legislation. Public choice desires better constitutional rules and institutions to minimize rent-seeking activity. Consequently, as in the case of externalities, public choice theorists desire the most efficient mechanism, acknowledging that all mechanisms will entail other inefficiencies.

THEORY OF THE BUREAUCRAT

Constitutionally defined governmental institutions do not alone influence policy creation. In fact, a large part of public choice theory examines the bureaucratic implementation of policy within government and its agencies. Although public administration and political science prescribes that bureaucrats seek the common good or public interest as civil servants, public choice holds that bureaucrats are rational individuals that will pursue policies that are personally and professionally beneficial. As utility maximizers, bureaucrats will use their power to enlarge agency budgets, giving them larger perquisites as a complement to the static, civil servant income. These bureaucratic actions lead to resource waste and rent seeking.^[3,10] Johnson stated that “the source of bureaucratic problems lies not in the bureaucrats but in the nature of public goods and the incentives, controls, and institutions existing in the bureaucracy.”^[1] He identified several primary causes of bureaucratic inefficiencies. The first is the loss of administrative control because the legislature intervenes between the bureaucracy that produces the goods and the citizens who consume the goods. The other causes result in the bureaucracy producing more public goods than demanded. This excess supply is a result of bureaucracies facing no competition in providing public goods. The political market gives bureaucracies few incentives to be efficient, so the tendency is toward budget maximization, not minimization.^[1,10]

Given this view of bureaucrats, how public goods should be produced is of great concern to public choice theorists. Focusing on local governments, the perspective calls for multiple, smaller-scale operations rather

than the Wilsonian-dictated, centralized, consolidated delivery mechanisms. This conclusion is a direct result of public choice criticisms of Wilsonian ideals. Public choice theorists believe that market competition results in more efficient distribution of the preferred mix of services because public service provision is labor intensive and lacks economies of scale. In addition, natural coordination among service units is superior to formal structure.^[9] Public choice theorists have conducted numerous empirical studies examining the outcome of jurisdictional, service delivery consolidations to determine if evidence supports this predicted outcome; the results are mixed as Lovrich and Neiman’s work suggests.^[9]

Although public choice is contentious to those in political science and public administration, many recognize the importance of questioning bureaucratic activity. Jackson stressed the importance of questioning “whose interests are served by the policy, which interests dominate the process and why.”^[11] Inefficient government institutions lead to bureaucratic failures, and public choice addresses these failures by proposing changes to the rules and incentive structures in bureaucracies.

GOVERNMENT ACTION

The application of private market principles to the political market and combining that with economic methodology has provided a rigorous framework to analyze the government’s role in society. Public choice theory seeks to improve government decision making through the minimization of bureaucratic tendencies and rent seeking. Public choice recognizes that every action has multiple costs associated with it, and understanding the political market consequences of the selected instrument associated with such government action is central to making decisions about whether particular goods should be publicly provided; and if public, which instruments best minimize the costs and externalities in the political market.

Not surprisingly, viewing government through the lens of public choice naturally leads to conclusions that government is often too large, too wasteful, and too inefficient. Such conclusions are often more in line with politically conservative positions and policies. However, those who would characterize public choice as “antigovernment” often miss the point. It is not the goal of public choice theorist to limit government for the sake of ideology. Rather their point is that the choice for government contains its own set of externalities, and the institutions and mechanisms chosen can be selected to minimize these externalities. One must always recognize that interventions may be more costly than the externalities themselves.

REFERENCES

1. Johnson, D.B. *Public Choice: An Introduction to the New Political Economy*; BristleCone Books: Mountain View, 1991.
2. Tullock, G. Externalities and government. *Public Choice* **1998**, 96 (3/4), 411–415.
3. Buchanan, J.; Tullock, G. *The Calculus of Consent: Logical Foundations of Constitutional Democracy*; University of Michigan Press: Ann Arbor, 1962.
4. Smith, A. *An Inquiry into the Nature of Causes of the Wealth of Nations*; The Modern Library: New York, 1937.
5. Black, D. *The Theory of Committees and Elections*; Cambridge University Press: Cambridge, 1958.
6. Arrow, K. *Social Choice and Individual Values*; John Wiley and Sons: New York, 1951.
7. Downs, A. *An Economic Analysis of Democracy*; Harper and Row: New York, 1957.
8. Van den Hauwe, L. Public choice, constitutional political economy and law and economics. In *Encyclopedia of Law and Economics: The History and Methodology of Law and Economics*; Bouckaert, B., De Geest, G., Eds.; Edward Elgar: Cheltenham, 2000; Vol. I, 604–659.
9. Lovrich, N.P.; Neiman, M. *Public Choice Theory in Public Administration*; Garland Publishing: New York, 1984.
10. Borchering, T.E. *Budgets and Bureaucrats: The Sources of Government Growth*; Duke University Press: Durham, 1977.
11. Jackson, P.M. Economics of public services: public choice and public sector. *Public Money Manag.* **1988**, 10, 13–20 (Winter).

Public Choice Theory

Thomas S. Nesselain

Urban and Regional Studies, Department of Economics, University of Wisconsin-Green Bay,
Green Bay, Wisconsin, U.S.A.

INTRODUCTION

In the eighteenth and nineteenth centuries, the study of politics and political economy overlapped in many important ways.^a By the early twentieth century, however, the disciplines of political science and economics had become quite distinct. The two disciplines differed in their focus of inquiry, the assumptions they made about individual motivation and in the research methods employed. Political scientists studied actors in the political arena, and economists studied actors in the marketplace. Moreover, political scientists tended to assume that political actors pursued the public interest, and their concerns were strongly normative. In distinct contrast, the fundamental behavioral assumption of neoclassical economic theory is that economic actors vigorously pursue their self-interest and the focus of inquiry is positive. Finally, the two disciplines were sharply divided with respect to methodology. Economics became increasingly deductive with heavy emphasis on mathematical modeling. In sharp contrast, political science retained its tradition of verbal argument and inductive empirical methods.

Beginning in the late 1940s and greatly accelerating in the 1960s, the chasm dividing the disciplines of economics and political science began to close in significant ways with the emergence of the subfield within the discipline of economics called *public choice*.^[1]

The subject matter of public choice is the same as that of political science: the theory of the state, voting rules, voter behavior, party politics, the bureaucracy, and so

on. The methodology of public choice is that of economics, however. The basic behavioral postulate of public choice, as for economics, is that man is an egoistic, rational, utility maximizer.

In brief, public choice economists use neoclassical theory “to study voters as rational maximizers, politicians as entrepreneurs, and bureaucrats as suppliers in a market-like process of consumption, production, and exchange.”^[2]

In retrospect, the emergence of public choice theory can be seen as a natural progression of social science. The competitive market economy is one basic kind of social choice mechanism relating to the allocation of scarce resources through various market exchanges. By the mid-twentieth century, economists had developed good knowledge of how the decentralized decision making, information, and incentive structures of a competitive market economy worked. Under certain conditions, albeit highly restrictive, an efficient allocation of resources would result. That is, the dollar-votes of consumers for alternative private goods and services would be efficiently channeled to producers and made effective. Competitive pressures would force those producers failing to produce in line with consumer demand, and at least cost, out of the marketplace. Relatedly, various defects or failures of market allocation also were increasingly recognized and gave rise to a large literature on the theory of market failure focused on such topics as imperfect competition, public goods, and externalities.

Moreover, by the early 1950s, it was increasingly recognized that the political process should be viewed as a social choice mechanism that in important ways complemented the market mechanism. Indeed, the failure of market allocation to achieve various social goals, equity, as well as efficiency, provides a natural explanation for the emergence of political institutions.

KEY TOPICS OF PUBLIC CHOICE

Nobel laureate Kenneth Arrow’s pathbreaking work *Social Choice and Individual Values* (1951) gave rise to a large literature examining the aggregation of individual preferences to maximize a defined set of

^aThis article is largely based on two sources: Mueller, D.C. *Public Choice II*; University Press: Cambridge, 1989; and Miller, G.J. The impact of economics on contemporary political science. *J. Econ. Lit.* 1997, 35 (4), 1173–1204.^[1-2] More basic sources include: Peacock, A.T. *Public Choice Analysis in Historical Perspective*; Cambridge University Press, 1992; Mueller, D.C. *The Public Choice Approach to Politics*; Edgar Elgar: Brookfield, VT, USA, 1993; and Tullock, G. *Government: Whose Obedient Servant?: A Primer on Public Choice*; Institute of Economic Affairs: London, 2000. Two strong critiques of public choice theory are the following: Green, D.; Shapiro, I. *Pathologies of Rational Choice Theory*; Yale University Press: New Haven, CT, USA, 1994; and Cullis, J.; Jones, P. Public choice: outrageous theory and enraging tests? In *Public Choice Analysis of Economic Policy*; Chrystal, K.A., Pennant-Rhea, R., Eds.; St. Martin’s Press: New York, 2000; 59–90.

normative criteria. Important topics of Arrow's work include the so-called "voting paradox," when majority rule is used to decide issues.

A concise example of Arrow's voting paradox is as follows:^[3] Assume a community of three individuals A, B, and C who are considering three alternative levels of public expenditure: \$10 million, \$5 million, and \$3 million. What Arrow showed is that even though each individual voter's preferences for alternative spending levels may be consistent, the community's preferences may not be. Specifically, in a series of paired elections to determine the socially preferred level of expenditure, an election between the \$10 million level and the \$5 million level, the \$10 million level may win by a 2 to 1 vote. Likewise, in an election between the \$5 million level and the \$3 million level, the \$5 million level may be chosen by a 2 to 1 vote. Finally, in an election between the \$10 million level and the \$3 million level, it may happen that the \$3 million level wins by a 2 to 1 margin. Such a majority rule outcome clearly is inconsistent. The first election suggests that the community prefers the \$10 million level of expenditure to the \$5 million level. The second election suggests that the \$5 million level is preferred to the \$3 million level. Standard notions of consistency thus imply that the community prefers the \$10 million level to the \$3 million level. The third election, however, suggests that the community prefers the \$3 million level to the \$10 million level. It must be stressed here that majority rule need not necessarily lead to inconsistent decisions by the community. The possibility clearly exists, however, especially when the issues decided collectively are multidimensional. That is, the choices do not correspond to more or less of a single characteristic, such as alternative spending levels. And when majority rule fails to result in consistent decisions, such problems as agenda manipulation and voting cycles emerge (i.e., paired voting can go on forever without a determinate outcome).

Arrow's work stimulated much research on real-world procedures for aggregating the preferences of individual voters under alternative voting schemes. Consequently, a significant literature exists on public choice in a direct democracy, that is, where social choices are decided directly via voter referenda. Key issues of this literature are the general question of the choice of voting rule, and investigations of the properties of specific voting rules such as the unanimity rule, the simple majority rule, and various alternatives to the simple majority rule. Closely related is the literature on local government finance and expenditure examining the so-called Tiebout hypothesis.^[4]

In the mid-1950s, Charles Tiebout advanced an important hypothesis with respect to the provision of local public goods such as education and refuse collection. He hypothesized that in urban areas with

a large number of independent political jurisdictions, households would vote with their feet and choose to reside in the jurisdiction with the bundle of local public goods and related taxes that they most preferred. Under a set of restrictive assumptions, Tiebout postulated that an efficient public goods and tax price equilibrium would emerge. Tiebout's provocative hypothesis has stimulated a considerable theoretical and empirical literature on the financing and provision of local public goods.

An even more important area of analysis is that of public choice in a representative democracy. The vast literature in this respect has been stimulated by Anthony Downs' classic work, *An Economic Theory of Democracy* (1957), probably the single most influential public choice work on the discipline of political science. Key concerns of this literature include the positive and normative aspects of two-party and multiparty competition as well as the effects of wealth-maximizing special-interest groups on political outcomes. An important line of research in this respect is the so-called "rent-seeking literature" that focuses on how special-interest groups (e.g., farmers, trade unions, industry associations as well as various professional groups such as physicians and lawyers) attempt to use government to transfer income to themselves. One well-known means is the erection of various international trade barriers to protect domestic industries and their workers from foreign competition.^[5]

Another major subject of public choice analysis is the supply of government output by government agencies and enterprises. While many social scientists have analyzed government bureaucracies, Niskanen was the first to use neoclassical economic theory to rigorously model the behavior of a government bureau.^[6] A key implication of Niskanen's model is that government bureaus will attempt to maximize their budgets, and in doing so, will produce an inefficiently large output. Niskanen's analysis is controversial and has been difficult to test empirically. Some evidence, however, from the peculiar Oregon school budget referenda process, supports the budget-maximizing hypothesis.^[1]

Public choice modeling also has been applied to the analysis of politics on key macroeconomic outcomes as well as the size and growth of government. Key topics include the effects of politics on short-run business fluctuations, the so-called political business cycle. Relatedly, the influence of special-interest groups and their redistributive activities on long-run economic growth has been investigated. Finally, a significant public choice literature exists, examining alternative hypotheses concerning the rapid growth of governments in the twentieth century, in particular, in the high-income welfare states of northern Europe.

PUBLIC CHOICE: A BRIEF OVERVIEW OF THE EMPIRICAL EVIDENCE

Are key predictions of public choice models supported by credible empirical research? A single, clear-cut answer cannot be given in this respect. Certain predictions are more strongly supported than others. For example, broad and rigorous evidence now exists that, in general, government suppliers are less efficient than private enterprise (see also *Enterprise: Public versus Private* in this Encyclopedia). For example, Mueller presented a listing of some 50 studies examining comparative enterprise efficiency.^[1] In only two studies were government enterprises found to be more efficient. In over 40 studies, government firms were found to be higher-cost producers, often some 40–60% (or more) higher.

A second important public choice hypothesis is that advanced by Mancur Olson in *The Rise and Decline of Nations* (1982). Olson argued that most interest group activity is focused on distributional goals that raise the satisfaction of their members. A key macroeconomic implication is that “distributional coalitions slow down a society’s capacity to adopt new techniques and to reallocate resources in response to changing conditions and thereby reduce the rate of economic growth.”^[1] Mueller concisely examined several studies that provided reasonably strong empirical support for Olsen’s theory.^[1]

Finally, key empirical facts to be explained by a credible theory of government are the large cross-national differences in the size of government as well as the rapid growth of government in the twentieth century, in particular, in the welfare states of northern Europe. Several supply-side and demand-side hypotheses have been put forward to explain the size and growth of government. A major study testing the relative strength of alternative hypotheses is the study of Switzerland by Pommerehne and Schneider.^[7] Commenting on this study, Mueller noted:^[1]

Pommerehne and Schneider’s results forcefully demonstrate what may be regarded as the single most important message that public choice has to teach—namely, the rules of the game *do* affect the outcomes of the game. Institutions matter. In Switzerland, the more direct the citizen influence on political outcomes, the smaller is the scale of government. Among the developed countries, citizens of Switzerland are able to exercise control over government more effectively than anywhere else. Only Switzerland makes much use of direct democracy and the referendum, and it has the strongest federalist system in the world. It also has the smallest public sector among the developed countries. The results of Pommerehne and Schneider suggest that these facts are related.

LIMITATIONS OF PUBLIC CHOICE MODELS

Clearly many public choice hypotheses are supported in some measure by existing empirical research as well as real-world experience. One must stress, however, that the straightforward application of neoclassical economic theory to the political arena has been only partial successful. Many important predictions emanating from public choice models have not been well-supported empirically. Indeed, in a major recent article on public choice theory, Miller argued that the greatest value of the incursion of public choice economists on political science is that it has stimulated the development of the subfield of “positive political theory” that seeks to build equally rigorous models that better explain real-world politics.^[2] Miller wrote:

[T]he most unambiguously beneficial effect of economics on contemporary political science has been methodological... Downs, Arrow, Olson, and others not only brought a new rigor to a theory-starved discipline, they had the prestige of micro-economics behind it. After all, the rationality assumptions worked to explain markets; why should political actors be less rational than other humans? It was necessary to come up with an equally logical deductive argument from equally reasonable first assumptions that led to empirical predictions that were somewhat closer to the world as political scientists could measure it. The casual, verbal argument, supported by selected case evidence, was no longer sufficient to carry the day....

Substantively, the contributions of economists have also been significant. Contrary, to the original intent of Downs, though, the contribution of economics was not simply to reveal the mechanistic, market-like underpinnings of the political world. Modern political economy is not simply an adjunct to market analysis, in which smooth exchanges are made between producer/politicians and consumer/voters via market-like competition and some political substitute for a price mechanism....

On the contrary, the most powerful impact has always been from the puzzling paradoxes revealed first by economic analysis, rather than from any predictable patterns revealed.

Miller developed his argument at length by focusing on four central topics of contemporary political economy: individual voting and electoral participation; the aggregation of voter preferences; the aggregation of decentralized voter information in a world of imperfect and costly information; and the influence of institutional constraints on political actors. To illustrate the contribution of positive political theory, we turn to a brief discussion of the theory of public goods, or more broadly, the theory of collective action. Based on the parsimonious assumptions of neoclassical economic theory, one would predict severe impediments to the provision of a public good (e.g., national

defense) or the attainment of a group goal requiring member contributions. If all group members are assumed to be rational egoists, then a basic incentive problem, the free-rider problem, would seem to arise invariably. Each person acting in their self-interest will not voluntarily contribute to the provision of a public good (or group goal), anticipating that others will contribute, and that the public good will be provided to their benefit but without their contribution. If all members of the group, however, are rational egoists, then none will contribute, and collective action will fail.

This basic prediction flowing from neoclassical economic theory, however, is not well supported by the huge number of public goods experiments that have been undertaken in recent years as well as much real-world experience. For example, Ostrom outlined key findings strongly supported by the existing body of experimental research and which are at considerable variance with the zero contribution thesis outlined above. Key findings here include the following:^[8]

1. Participants in public goods experiments are found to contribute 40–60% of the resource endowment given them in support of collective action to provide the public good.
2. When participants believe that others will cooperate, they are more likely to cooperate.
3. As participants learn the game better, they tend to cooperate more, not less, as rational egoists would be predicted to do.
4. Face-to-face communication tends to increase cooperation.
5. How much group members contribute in a public goods experiment is influenced by contextual factors (e.g., allowing communication among members, assigning sanctions against noncontributors, etc.).

On the basis of the above findings, Ostrom argued that a revised theory of collective action consistent with the empirical evidence is one that assumes two kinds of “norm-using” participants in addition to rational egoists. The first she labels “conditional cooperators,” that is, individuals willing to initiate cooperative action when they believe others will reciprocate. The second kind of “norm-using” participant she labels as “willing punishers,” that is, individuals who will use permitted sanctions to punish noncontributors to the public good (or group goal).

A less radical approach to explaining collective action is to retain the assumption that individuals are rational egoists but assume that an individual's satisfaction depends, at least in part, on an individual's acceptance and standing as defined by a relevant group relationship. Miller illustrated this point using the

example of participation in the U.S. civil rights movement of the 1960s:^[2]

[M]ost protestors in the civil rights movement did not decide to participate as individuals, but as members of local churches with highly structured systems of social incentives that linked good standing in the group with participation in the civil rights movement. And a common denominator of the research on voting, interest groups, and public goods is that much of the variation in individual behavior has to do with how an individuals relevant group has defined the social costs and incentives associated with those acts of political participation.

It also should be stressed at this point that the discipline of economics and public choice theory have had a major impact on political science via the emergence and rapid growth since the mid-1970s of the partially related subfield called the “new institutional economics.” The two central propositions of this new subfield are, first, that “institutions do matter” and, second, “the determinants of institutions are susceptible to analysis by the tools of economic theory.”^[9] In particular, concepts such as “transaction costs” and “credible commitment” have been used by economists and positive political theorists to explain key institutional features of real-world political systems and why they change over time.^[10] In this context, Miller noted:^[2]

[T]he anomalous and paradoxical behavior revealed by [the application of economic theory to politics] forced political scientists to pay attention to the structured institutional constraints on individual maximizing behavior. Understanding political participation indicated that individuals did not make atomistic choices in a structureless environment, but responded to blandishments shaped by the social structures of which they were a part.... It was only my focusing analysis on the institutional determinants of voting, legislative, and regulatory behavior that anomalous behavior began to be understood....

As the force of political institutions was revealed, the emphasis in economic modeling shifted subtly from the determinants of the self-interested choice, to *restraints* on self-interested choice. Increasingly, institutional analysis focused on what came to be known as the “commitment” issue—settings in which long-term gains could be achieved by limiting, rather than expanding, the scope of individual choice.

CONCLUSION

The application of neoclassical economic theory to politics has clearly contributed in a substantial way to an improved understanding of political institutions and outcomes. It is also the case, however, that many

key predictions, in particular those fundamentally emanating solely from the rational egoist assumptions of neoclassical economic theory, have not been well supported by empirical research. A credible case can be made that the primary long-run impact of neoclassical economics on the study of political institutions and behavior is the emergence of the new field of positive political theory that has led to a richer, more varied understanding of politics:

[C]ontrary to the early economists in political science the most powerful determinants of outcomes are not the atomistic maximizing choices of individuals, but the groups and institutions that guide and constrain self-interested individual choice. Our understanding of the anomalies generated by the first generation of political economic analysis has propelled political science (and perhaps economics!) to understand the rich and compelling social norms and institutional contracts that govern political exchange in its diversity.^[2]

REFERENCES

1. Mueller, D.C. *Public Choice II*; Cambridge University Press: New York, U.S.A., 1989; 1–2, 259–266, 311–317, and 346.
2. Miller, G.J. The impact of economics on contemporary political science. *J. Econ. Lit.* **1997**, 35 (4), 1173–1204.
3. Rosen, H. *Public Finance*; Irwin: Chicago, 1995; 122–125.
4. Tiebout, C.M. A pure theory of local public expenditure. *J. Polit. Econ.* **1956**, 64, 416–424.
5. Tullock, G. The welfare costs of tariffs, monopolies, and theft. *Western Econ. J.* **1967**, 5 (June), 224–232; Krueger, A. The political economy of the rent-seeking society. *Am. Econ. Rev.* **1974**, 64, 291–303; and Tollison, R.D. Rent-seeking: a survey. *Kyklos* **1982**, 35 (4), 575–602.
6. Niskanen, W.A., Jr. *Bureaucracy and Representative Government*; Aldine-Atherton: Chicago, IL, U.S.A., 1971.
7. Pommerehne, W.; Schneider, F. Unbalanced growth between public and private sectors: an empirical examination. In *Public Finance and Public Employment*; Haveman, R., Ed.; Wayne State University Press: Detroit, MI, 1982; 309–326.
8. Ostrom, E. Collective action and the evolution of social norms. *J. Econ. Perspect.* **2000**, 14 (3), 141–145.
9. Williamson, O.E. The new institutional economics: taking stock, looking ahead. *J. Econ. Lit.* **2000**, 38 (3), 595–613.
10. Dixit, A. *The Making of Economic Policy: A Transaction Cost Politics Perspective*; Cambridge University Press: Cambridge, 1996.

Public Funds Investment Strategies

William G. Albrecht

Department of Political Science and Public Administration, University of North Carolina at Pembroke, Pembroke, North Carolina, U.S.A.

INTRODUCTION

In 1939, John Donner, Manager-Secretary for Los Angeles California's Department of Pensions, published an article titled "Investment of Pension and Other Trust Funds" in which he detailed two personal observations. First, he indicated that the question of investing monies with both safety and income in any type of financial organization was a problem challenging the ingenuity of the best minds. Second, he noted that the prime requisite being followed by government trust funds was safety of principal; the extent of income was considered secondary. In terms of management, Donner's conclusion was that investment techniques were incapable of substituting for the deductive ability and experience of a seasoned administrator.^[1]

Retrospectively speaking, Donner's early 20th century commentary on public sector investment practices resembles the adage "the more things change the more they stay the same." While the field has evolved in terms of constraints, goals, and objectives, the persistent constant is that the question of investing monies with both safety and income remains a 21st century problem challenging the ingenuity of the best minds. This statement is supported with the acknowledgement that those working in the discipline continue to encounter increasingly sophisticated techniques and unorthodox proposals for handling the nation's finances. In recognition of these facts, the essay presented here broadly defines and illustrates public sector investment strategies while outlining relevant factors for public organizations operating in the modern era.

INVESTING PUBLIC FUNDS

Any contemporary discussion concerning public sector investment practices must begin by realizing that public funds are collective monies segregated by a mechanism known as fund accounting. The intent of a fund accounting system is to control the purpose of financial resources while illuminating sources and uses of funds. Although a complete delineation of fund structures varies by organization, the most relevant aspect pertaining to the investment of public monies

is the proximity of the relationship between a fund's liquidity requirements and operating budget decisions.

Future operating budget decisions are important factors in determining disbursements and the need for liquidity or "cash" at a particular point in time.^[2] Furthermore, the results of investment decisions contribute to underlying patterns in cash flows. Funds susceptible to operating budget decisions require a predictable short-term investment focus and generally fall under the purview of cash management. A government's general fund is an example of a public fund with cash management characteristics. Funds less sensitive to these same decisions afford administrators the prospect of pursuing investment opportunities requiring a longer term focus. Trusts are the most straightforward illustration of these types of funds.

The fundamental justification for investing public funds is the economic notion that money has a "time value." The premise of the time value concept is that an amount of money is worth more in the present than some future period if the money is invested to produce a "yield." Yield^a is referred to with a variety of labels including "interest" and "rate of return." Whatever the designation, the common element is that each term refers to what an investor receives over and above the principal amount invested.^[3] Given that public organizations operate in a world of scarcity a higher yield on invested monies is preferred to a lower yield.

A corollary to the time value of money concept is that there are opportunity costs in the form of foregone revenues associated with an organization's failure to invest idle funds.^[3] Sustained increases in average price levels are also capable of eroding the purchasing power of monies not employed to keep up with inflation. In both cases the economic consequence is that an organization's capacity to disburse benefits, internalize safety margins, replace existing assets, expand services, or reduce reliance on future contributors is, in all probability, compromised.

The counterpoint to an attempt at capitalizing on the time value of money notion is that such activities introduce a degree of uncertainty in relation to the

^aTechnically, the specific definition of "yield" is dependent upon the type of investment and underlying monetary flows.

future value of the original or “principal” amount invested. In finance uncertainty is commonly called “risk.” Like yield the term risk is referred to with a variety of labels including default risk, market risk, and liquidity risk.^[2] Whatever the designation, the common element is that each expression refers to the possibility of a loss of investment income or principal amount.^b

Compared to yield the existence of scarce resources implies that organizations prefer less risk to more risk. Combining this preference with that for a higher yield suggests that managers of public funds face the task of generating a yield on invested monies while reducing uncertainty. Consequentially, investment strategies often focus on making decisions within the confines of these two constructs. However, every administrative choice of approach must necessarily be tempered by two realities. The first is that achieving a higher yield generally entails the acceptance of increased risk. The second is any course of action involving public funds must consider nuances underlying the public sector.

DELIMITING PUBLIC FUNDS INVESTMENT STRATEGIES

Drawing on the discussion above, a public sector investment strategy may be defined as a public organization’s plan for using idle money in a way that increases financial holdings while protecting the principal amount invested. A subjective assessment of the literature suggests that researchers and practitioners tend to refer to these plans as well as their formulation and implementation in one of three ways: asset allocation decisions, investment process decisions, and decisions concerning combinations of asset allocations and investment processes.

Asset allocation tactics involve choices as to the placement of funds in capital market instruments or location. While the universe of possible instruments is vast most can be classified as either debt or equity. Debt is essentially a loan and includes short-term money market instruments such as U.S. Treasury bills, bankers acceptances, and commercial paper. In addition, debt investments consist of long-term instruments of the U.S. government, banks, and the bonds of various private sector businesses both foreign and domestic. As an “asset class,” equity instruments refer to stock ownership in domestic or international firms.

By definition any public organization’s choice of a single instrument constitutes a basic asset allocation

strategy as the decision amounts to an assumed level of risk and expected yield. Fig. 1 illustrates this point. As shown, Treasury bills are positioned at the low end of the risk–yield range while government and corporate bonds are located in the middle spectrum of these criteria. By comparison, domestic and international equities are at the high end of the risk–yield range. The slope of the capital market line reveals a positive relationship indicating that a higher yield is achieved at the expense of increased risk.

Collective combinations of instruments are called a portfolio. And, portfolio decisions culminate in an investment strategy with characteristics that are apart from a single instrument. The divergence stems from the concept of “diversification,” which suggests that risk can be reduced by investing in financial assets that exhibit less than perfectly positive correlation.^[4] While a complete discussion on the benefits of diversification is complicated, an intuitive feel can be achieved by recognizing that the yields of various instruments are unlikely to consistently move in directional tandem. This analogy also applies to economic sectors and economies. Thus, risk reduction is the *raison d’être* for investing public funds in a portfolio with a variety of instruments from various locations.

Compared to asset allocation strategies decisions concerning investment processes involve procedures that a public organization may choose to utilize during the implementation of investment activities. Again, while the universe of prospective processes is substantial most involve engaging in some type of purposive action. For example, in the cash management arena time period “hedging” strategies entail the *processes* of buying or selling contracts called “futures” for the *purpose* of locking in future or current yields, respectively.^[5] Futures contracts are legally binding obligations between buyers and sellers for the transfer of

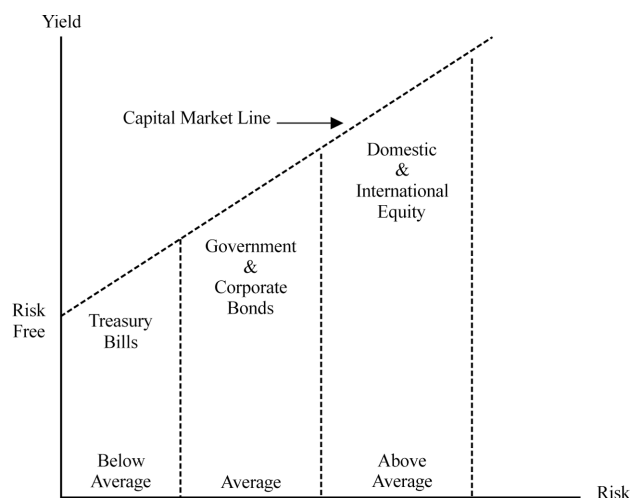


Fig. 1 Relationship between risk and yield.

^bMiller^[2] summarizes default, liquidity, and market risks as uncertainty surrounding full recovery of principal value, the ability to convert to cash, and negative changes in financial markets, respectively.

U.S. securities, certificates of deposits of banks, and state and local government bonds. Options contracts are an additional type of tactic for hedging against undesirable interest rate movements. However, unlike futures contracts, options do not represent an obligation but the reservation of a right by the holder to take action at a specified time. An interesting point to note is that a combination of these hedging processes or “options on futures contracts” is also a possibility.^[5]

Another process frequently used by public trusts is that of passive indexing. When engaging in this strategy an investment manager purposively attempts to duplicate the risk–yield characteristics of a target portfolio by mimicking its construction with identical proportionate holdings of matching instruments. The term indexing also denotes the simpler procedure of investing in a “mutual fund” company that has an established portfolio resembling that of a target index. Regardless of the approach, passive indexing implies that no attempt is made to accrue risk or yield levels in excess of the market. Popular indexes used by investment managers include the well known “S & P 500” for equity investments and the “Lehman Aggregate” for debt holdings.^[6]

While an obvious relationship exists between the imitative processing strategy of indexing and asset allocation tactics there is no reason to assume that an organization pursuing an indexing strategy will do so exclusively. Instead, some may opt to engage in both indexing and active asset allocation tactics, which are attempts to achieve yields above that of market indices. Strategic decisions of this nature are based on a view of inefficient markets where additional information, insight, or skill can lead to yields that are superior to that of a passive target portfolio.^[7] Market views aside this is one more illustration of how asset allocation and process decisions can be combined to create an overall investment strategy out of subsets of strategies.

Inputs to Modern Public Funds Investment Strategy Decisions

While Donner’s conclusion regarding the non-substitutability of investment techniques for seasoned administrators with deductive ability may or may not be deemed as accurate for his time, a case can be made that they are now complementary inputs to public investment strategy decisions. A notable reason for this assertion is that escalating levels of fiscal stress and increasingly volatile markets have combined with the legislative trend of liberalizing investment statutes to create a complexity that is essentially a barrier to unaided rationalism. The result is that the administrative ability to predict has at least joined if not

superseded the ability to deduce in terms of prominence. This normative shift has facilitated the development and use of techniques designed to aid administrators in formulating competent investment decisions.

To illustrate, various formulas such as the Miller-Orr and Economic Order Quantity models can assist managers in determining optimal cash balances so that extra treasury funds may be invested in a marketable securities portfolio.^[8] Graphical procedure such as yield curves are also available for examining the relationship between the yield on various financial securities and time. Other techniques include optimization procedures that allocate trust monies to debt and equity investments within a portfolio that are “efficient.” Efficiency in this sense means that no other possible combination of instruments can provide a lower level of probable risk or a higher rate of yield.^[9] While a number of other techniques are also available the shared facet among most is that each is a purposive process culminating in a decision rule for formulating and implementing a rational investment strategy. Thus, techniques are arguably a complementary input to modern strategy decisions if not actually part of a strategy.

Public Sector Distinctions

A discussion concerning public sector investment practices would be incomplete if no specific mention were made concerning the purpose of such activities in relation to public organizations and the environment in which they are implemented. The public setting relinquishes both explicit and implied “governance” constraints that are in addition to that of fund limitations or efficiency considerations.^[10,11]

For example, public organizations exist to fulfill missions that relate to achieving social goals; the accumulation of money is not considered an end. Instead, money is deemed as a means to a mission oriented end. Investing public funds, like all financial management activities, is intended to reduce the likelihood that financial concerns come to dominate the normatively superior goal of achieving organizational effectiveness.^[12] In this sense, the mission orientation of public organizations is an explicit governance constraint on investment objectives and the formulation and implementation of strategies designed to achieve them.

Another illustration of an environmental constraint can be seen in relation to the historical evolution of legalities pertaining to the field. While the liberalization of public sector investment statutes appears to be a persistent trend the inclination is far from unfettered. Public money managers often remain bound by internal investment policies or constitutional restraints

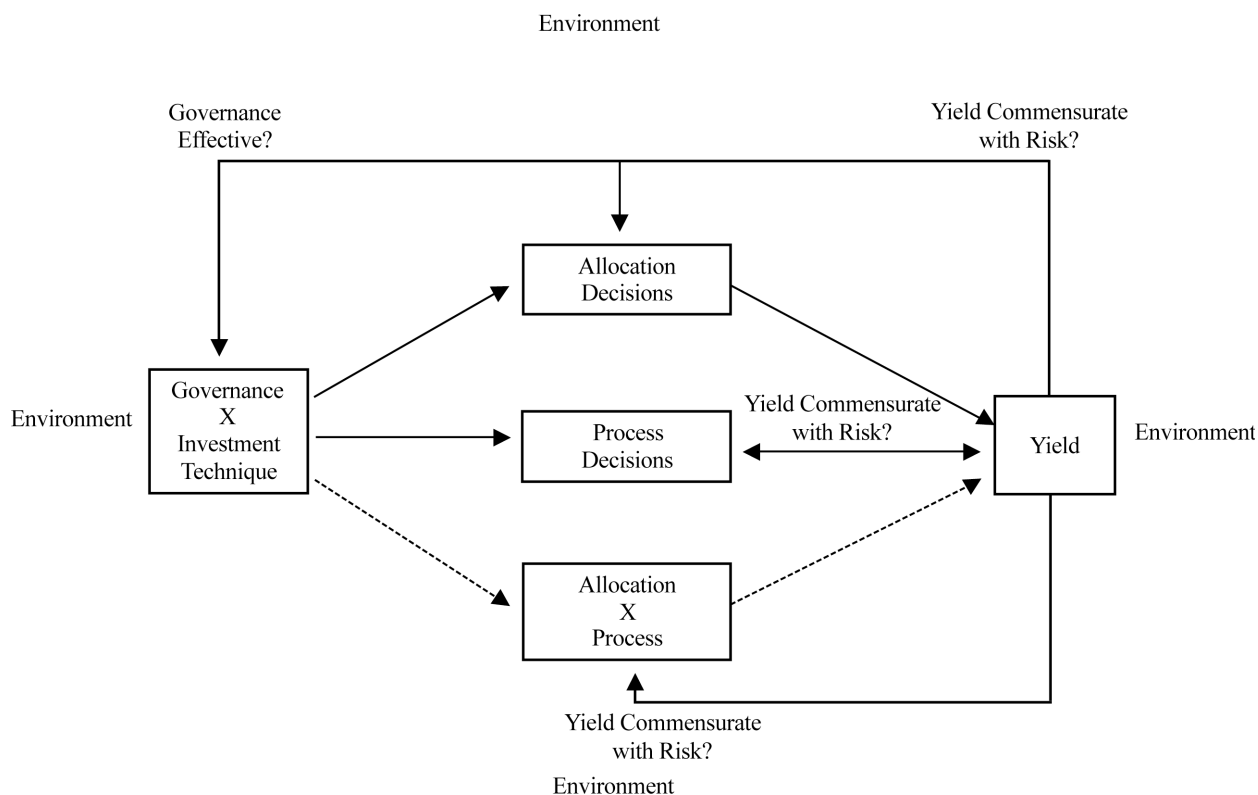


Fig. 2 Investment strategies as intervening variables.

that limit investments to low-risk instruments in specific geographic areas.^[13] In some cases, trusts, such as pension funds, are directed to consider social or economically targeted investments.^[14] These political decisions limit the investment strategy opportunity set and confound the discharge of rational investment tactics. When these circumstances occur, public managers working in a liberal democracy must adjust strategies and techniques to account for prescribed social purposes.

An implicit constraint also exists in relation to public money managers themselves. Empirical research consistently indicates that most seek to protect investment resources against any type of loss.^[2,15] In addition, theoretical work on the acceptance of financial management techniques by public administrators suggests that managers' full definitions of risk and yield may incorporate qualitative characteristics as well as quantitative measures.^[15,16] Normative considerations aside, the reality of such occurrences is that administrator behavior potentially limits the accumulation of fund surpluses via investment strategy decisions.

CONCLUSIONS

This essay has broadly defined and illustrated public sector investment strategies and outlined relevant

factors for organizations operating in the modern era. The emerging picture is that investment strategy decisions are intervening variables between an organization's governance and resulting yield. Fig. 2 portrays this "mediating" relationship.

As shown, governance and techniques are complementary inputs to strategy choices.^c These decisions, either in isolation or combination, are then predicted to determine yield. The potential for combining asset allocation tactics with process decisions is denoted by the dashed path. In the case of such an occurrence a process strategy would *potentially* qualify as a moderating factor for asset allocation decisions. This moderating hypothesis is supported empirically with studies finding that more than 90% of the variability in a fund's yield is the result of asset allocation tactics, and that managers perceive asset allocation decisions as being the most important investment strategy. The feedback loop designates the figure as a systems model akin to Lynch's approach for analyzing budget accountability.^[17] An interesting and unanswered empirical question is how realized yield determines these precursors.^[10] The (political) environment is

^cHere governance is a generic term referring to the multifaceted complex of behaviors, rules, or practices that influence investment decisions.

depicted as an exogenous force surrounding organizations and their decisions.

An important point to note is that factors other than those mentioned in this limited space may influence an organization's choice of investment strategy. For example, fund size and professional sophistication are practical determinants for active participation in money markets.^[2] These requisites often preclude the eligibility of organizations that are unable to incur the higher transaction costs in reducing their effect. Of course, investment management practices have evolved toward diminishing such barriers through processes such as pooling and the production of professionals who specialize in modern techniques. This confirms that the question of investing monies with both safety and income remains a 21st century problem challenging the ingenuity of the best minds.

REFERENCES

1. Donner, J. Investment of pension and other trust funds. *Municipal Finance* **1939**, 12 (2), 10–16.
2. Miller, G. *Investing Public Funds*; Government Finance Officers Association: Chicago, IL, 1998.
3. Reed, B.J.; Swain, J.W. *Public Finance Administration*, 2nd Ed.; Sage Publications, Inc.: Thousand Oaks, CA, 1997.
4. Markowitz, H.M. *Portfolio Selection: Efficient Diversification of Investments*; Yale University Press: New Haven, CT, 1959.
5. Miller, G.J. *Government Financial Management Theory*; Marcel Dekker, Inc.: New York, 1991.
6. Johnson, T.B. Overview of the equity index fund marketplace. In *Indexing for Maximum Investment Results*; Neubert, A.S., Ed.; Glenlake Publishing Company, Ltd.: Chicago, IL, 1998; 211–220.
7. Haugen, R.A. *Modern Investment Theory*, 4th Ed.; Prentice Hall: Upper Saddle River, NJ, 1997.
8. Schwartz, E. Inventory and cash management. In *Management Policies in Local Government Finance*; Aronson, J.R., Schwartz, E., Eds.; International City/Council Management Association: Washington, DC, 1996; 389–410.
9. Lecompte, R.L.B. Portfolio theory. In *Encyclopedia of Public Administration and Public Policy*; Rabin, J., Ed.; Marcel Dekker, Inc.: New York, 2003; 2, 947–951.
10. Useem, M.; Mitchell, O.S. Holders of the purse strings: governance and performance of public retirement systems. *Social Sci. Q.* **2000**, 81 (2), 489–506.
11. Albrecht, W.G.; Hingorani, V. Effects of governance and investment strategies on state and local government pension fund financial performance. *Int. J. Public Admin.* **2004**, 27 (8–9), 673–700.
12. Finkler, S.A. *Financial Management for Public, Health, and Not-For-Profit Organizations*; Prentice Hall: Upper Saddle River, NJ, 2001.
13. Collins, B.K.; Khan, A. Information asymmetry in public investment management. In *Financial Management Theory in the Public Sector*; Khan, A., Hildreth, W.B., Eds.; Praeger Publishers: Westport, CT, 2004; 25–54.
14. Nofsinger, J.R. Why investment targeting does not make sense! *Financial Manage.* **1998**, 27 (3), 87–96.
15. McCue, C.P. The risk–return paradox in local government investing. *Public Budgeting Finance* **2000**, Fall, 80–99.
16. Albrecht, W.G.; Lynch, T.D. Assessing the likely acceptance of financial management techniques in the public sector: a characteristic approach. In *Financial Management Theory in the Public Sector*; Khan, A., Hildreth, W.B., Eds.; Praeger Publishers: Westport, CT, 2004; 141–163.
17. Lynch, T.D. *Public Budgeting in America*, 4th Ed.; Prentice Hall: Upper Saddle River, NJ, 1995.

Public Goods

Cristina E. Ciocirlan

*Department of Public Administration, Pennsylvania State University at Harrisburg,
Middletown, Pennsylvania, U.S.A.*

INTRODUCTION

Public, or collective, goods are characterized by non-excludability in use, non-rivalry in consumption, or both these characteristics combined. A good is non-excludable if it is costly and counterproductive to exclude individuals from using it. It is non-rivalrous if one person's consumption cannot reduce or impede the ability of others to consume the good. When individuals do not take into account the marginal social cost of consumption, the public good becomes congested. Generally, a public good is also indivisible; one cannot meaningfully speak of units of a public good (e.g., ambient air).

When public goods are provided through collective action settings, there is a high likelihood of conflict between self-interest and group interest. This conflict leads to a prisoner's dilemma type of outcome, which corresponds to a Pareto-inferior^a allocation of goods in economy.

The inherent nature of public goods and the collective action problems can lead to the failure of markets to achieve Pareto efficiency.^b Thus, there is a justifiable role for government intervention. Legislators are faced with the problem of pricing public goods and providing the value-maximizing quantity of public goods for society.

CHARACTERISTICS OF PUBLIC GOODS

Public, or collective, goods are non-excludable in use, non-rivalrous in consumption, or may exhibit both these characteristics in varying degrees. A good is non-excludable if it is impossible to exclude other individuals from simultaneously consuming the same good. It is non-rivalrous if the consumption of one person of the public good does not impede another person's consumption of the same good at the same time. Both the consumption and the production of public goods are indivisible; public goods cannot be divided up and

sold.^[2] Public goods can often be congested or non-congested. Congestion occurs when the marginal social cost of consumption exceeds the marginal private cost of consumption.^c The divergence between these costs represents an externality of consumption. When non-excludable and non-rivalrous public goods are also uncongested, they are called "pure public goods." Lighthouses and national defense are classic examples. When these goods are congested, they are called "ambient, or impure, public goods" with consumption externalities.^d Ambient air is an example. When consumption of public goods is limited to residents of the political jurisdiction in which these goods are provided, they are called "local public goods."^[2]

In contrast, the defining characteristics of private goods are rivalry in consumption, excludability in ownership and use, and divisibility. Rivalrous consumption means that a private good is consumed exclusively by the person who buys it. If person X wears a pair of shoes, person Y cannot wear the same shoes at the same time. Excludability in ownership and use means that one has control over use of the good. One can undoubtedly decide who is allowed to drive their car. Private goods are divisible; they can be divided into units and assigned a unit price.

The public good-private good dichotomy recognizes a continuum between the two extremes. Any good exhibits one or more of the public good characteristics. When public good characteristics are found in a commodity, both the need for consumers to pay for it and the incentive of suppliers to provide it are challenged. Thus, markets tend to produce those commodities with public good characteristics in inefficiently small quantities.^e

^cThe marginal social cost of consumption is defined as the sum of the marginal costs that consumers of a public good impose on each other. When the good is uncongested, the marginal social cost is zero. The marginal private cost of consumption is the cost of consumption generated by an additional unit of consumption supported by one individual (Ref.^[1]).

^dThese goods were labeled as such by Ref.^[1].

^eThe two basic questions that could indicate whether markets could effectively provide commodities with public good characteristics are whether consumers would pay for something that would be available anyway and whether suppliers could collect from buyers the value of their services (Ref.^[2]).

^aA Pareto-inferior allocation of goods is not a socially ideal outcome because it is possible to find a different allocation that makes some individuals better off without making anyone worse off (Ref.^[1]).

^bA Pareto-efficient allocation of goods is an optimal allocation; it is impossible to find an alternative allocation that makes some individuals better off without making anyone worse off (Ref.^[1]).

When the costs of exclusion are small, but the good is non-rivalrous, private provision is possible. For instance, a private supplier could build a bridge and charge a fee for the usage of it. There are two problems with this setting, however. First, the producer might not set the price equal to the marginal cost of consumption, which is zero.^f Second, the private supplier may not provide the optimal resource size to maximize social surplus. In the presence of congestion, the marginal social cost of congestion^g becomes positive. A private supplier must charge users a price equal to the cost imposed on all other users during congested times. However, private firms would charge consumers a higher cost than the marginal social cost of congestion to generate sufficient revenues to cover the costs of construction and maintenance. Thus, the market fails because it excludes users who would obtain higher marginal benefits than the marginal social costs that they impose.

When public goods are non-excludable but rivalrous, they confront problems of overconsumption. Non-excludability translates into the failure of the market to efficiently provide open access resources. When choosing the optimal consumption quantity, individuals consider their marginal private costs instead of marginal social costs; hence, they overconsume these resources and underinvest in their preservation.^[1]

THE COLLECTIVE ACTION PROBLEM

A collective action problem arises when individuals, acting out of pure self-interest, are unable to coordinate their efforts to produce desirable public goods. In a collective action setting, there is a high likelihood of conflict between the self-interest and the common interest of members involved.^[3] Each individual attempts to free ride on others' efforts. Free riding occurs when one individual reaps the benefits of a certain action or product without paying for it. Because everyone thinks alike, no public good will be produced and everyone is worse off than they would have been had each individual contributed toward the provision of the public good. In short, collective action problems arise when individuals arise at Pareto-inferior outcomes in the pursuit of their self-interest.

The collective action problem can be illustrated with the game-theoretic terminology of a prisoner's dilemma

setting (see Fig. 1). Suppose a community confronts serious air pollution problems caused by inefficient car driving. Ambient air is the closest to the purest definition of public goods: once cleaned, it is not possible to exclude individuals from breathing it, and one individual's breathing of clean air does not reduce the ability of others to breathe it at the same time. The community decides to institute a system by which cars are driven less and only when it is necessary. It is assumed that cycling and walking are viable alternatives.

Given the perfect excludability and non-rivalry in consumption and use, an individual has a strong incentive to free ride on others' efforts to keep the air clean. Thus, the individual has two strategies available: to cooperate and drive less, or to defect and rely on others' efforts to reduce driving in the community. Community members can also pursue these strategies, depending on their valuation of clean air and the extent to which they pursue self-interest.

In Fig. 1, if the individual decides to cooperate, the payoffs are either clean air with the expense of less driving both for individual and community (when community cooperates), or dirty air at the price of less driving for individual only (when community defects). If the individual decides to defect, the payoffs are either clean air at no cost to the individual (when community cooperates) or dirty air (when community defects, too). Mathematically, if the individual decides to cooperate, regardless of what community members do, the payoff is -2 , which is the sum of individual payoffs from cells A and B. If the individual decides to defect, the sum of payoffs is 0 (the sum of individual payoffs from cells C and D); thus, the individual will choose to defect.

		Community	
		Cooperate	Defect
Individual	Cooperate	A. Clean air; less driving for both individual and community Payoff = $(-1, -1)$	B. Dirty air; less driving for individual only Payoff = $(-1, 0)$
	Defect	C. Clean air; less driving for community only Payoff = $(0, -1)$	D. Dirty air; driving stays the same Payoff = $(0, 0)$

Fig. 1 Collective action problems as prisoner's dilemma games. *Note:* Prisoner's dilemma games have been used for the first time as a conceptual framework in Ref.^[4]. The first number in parentheses represents the payoff to the individual; the second number represents the payoff to the community. Costs of less convenience stemming from reduced driving are indicated by -1 , the absence of these costs is indicated by 0 .

^fIn the absence of congestion, the cost of an additional car passing that bridge (i.e., the marginal cost of consumption) is approximately zero.

^gThe marginal social cost of congestion is the cost imposed to society by an additional unit of public good consumed. The marginal private cost is the cost of consuming one additional unit supported by one individual.

Likewise, if community decides to cooperate, the sum of its payoffs for cooperation will be -2 (community payoffs from cells A and C). If the community decides to defect, the sum of its payoffs is 0 (the sum of community payoffs from cells B and D); thus, the community chooses to defect, too. The dominant strategy is, therefore, that both individual and community defect, and the public good will not be provided.

Nevertheless, the view that collective action is undermined by private interest can be challenged on several grounds. First, in large groups, the pursuit of one individual's self-interest yields an insignificant benefit. Thus, the outcome of the larger groups may exhibit public good characteristics.^[4] Second, the formation of coalitions within organizations, facing attractive ratios of benefits to costs, could lead to optimal collective solutions. Third, altruism, gaining utility when someone else gains utility from an outcome, could lead to optimal cooperative solutions. Billions of dollars are donated every year to charitable and religious organizations.^h These organizations work toward a collective good, the achievement of which does not selectively benefit organization members. The collective outcome is a public good. Fourth, cooperation tends to be possible if social incentives—such as desire to gain or sustain friendship, maintain one's social standing, or concerns about reputation reliability and dependability—are present.^[6]

THE ROLE OF GOVERNMENT

The presence of non-rivalry, non-excludability, or congestion can lead to the failure of markets to efficiently provide public goods.ⁱ For example, if national defense were to be provided privately, all individuals in society could benefit from it, and it would not be possible to exclude or charge individuals for national defense. If everybody free rides and nobody contributes to the provision of a public good, the public good will not be supplied and everyone will be worse off.

The rationale for government intervention stems from its power of coercion and taxation. Government has the power to force consumers to contribute to the provision of the public good. Public good provision through government can be achieved by direct and indirect means. Direct provision could be performed

through government bureaus and special districts, whereas indirect provision could be done by independent agencies and contracting out government functions to private sector.^[1]

Economic theory stipulates that various levels of government produce different public goods efficiently. Local governments should produce the optimal quantity of a local public good demanded by their local citizens, whereas the federal government must supply those goods that generate benefits for people living outside the local jurisdiction. In reality, however, cities and counties are multiproduct jurisdictions; they often produce goods that spill over to other jurisdictions.

Just as markets sometimes fail to produce social value, so too can government. Political economy theories of regulation demonstrate that politicians often fail to employ the best policy instruments that would achieve Pareto-efficient allocations. This may occur either because of their capture by special interests^[7] or simply because of problems inherent in direct democracy.^[1]

CONCLUSION

The provision of public goods through private markets and collective action is inefficient because of free riding and prisoner's dilemma types of games. The intervention of government in economy is thus justified. Through its power of coercion and taxation, government is theoretically able to supply an efficient quantity of public goods. However, just as markets fail, governments can also fail. Optimal economic instruments are often not used because of the capture of politicians by special interest groups or simply because of problems inherent in direct democracy. Depending on the public good in question, partnerships between government and private firms sometimes leads to efficient outcomes.

Although theoretically it has been demonstrated that the provision of public goods is more efficiently provided by the government, the relative efficiency of government versus markets in practice is questionable. In recent years, there has been an increasing trend toward contracting out to the private sector and public-private partnerships. The reasoning behind this evolution rests with the assumption that the ownership of a public good should lie with the party that values its benefits more.

ACKNOWLEDGMENT

The author would like to thank Dr. Steven A. Peterson, Pennsylvania State Harrisburg and Gabi Huber, Clemson University, for suggestions to an earlier draft.

^h\$168 billion was donated by the public to charitable organizations in 1999 (Ref.^[5]).

ⁱMarket failures arise when the decentralized production of goods yield Pareto-inefficient outcomes. Pareto efficiency rests on the assumption that the free markets are able to yield an optimal distribution of goods in society. This means that it is impossible to find another allocation that makes someone better off without making anyone worse off (Ref.^[1]).

REFERENCES

1. Weimer, D.L.; Vining, A.R. Rationales for public policy: market failures. In *Policy Analysis: Concepts and Practice*, 3rd Ed.; Prentice Hall: Upper Saddle River, NJ, 1999.
2. Stevens, J.B. *The Economics of Collective Action*; Westview Press: Boulder, CO, 1993.
3. Olson, M. *The Logic of Collective Action*; Harvard University Press: Cambridge, MA, 1965.
4. Hardin, R. Collective action as an agreeable n-Prisoner's Dilemma. *Behav. Sci.* **1971**, *16*, 472–481.
5. National Center for Charitable Statistics, at <http://nccs.urban.org/factsht.htm>.
6. Chong, D. *Collective Action and the Civil Rights Movement*; The University of Chicago Press: Chicago, 1991.
7. Stigler, G. The theory of economic regulation. *Bell J. Econ. Manage. Sci.* **1971**, *2*, 3–21.

Public Policy

Steven A. Peterson

*School of Public Affairs, Pennsylvania State University at Harrisburg,
Middletown, Pennsylvania, U.S.A.*

INTRODUCTION

Public policy is normally conceived as government action to address some problem. The central focus is, as the old phrase has it, who gets what when how. Policy is the end result of a complex decision-making system, encompassing many actors. This article explores standard definitions, a typology of policy types, and some debate over how well one can draw a boundary between “public” policy and “private” policy.

DEFINITION

Public policy, in its most simple definition, is that which government does. James Anderson puts it more formally:^[1] “A relatively stable, purposive course of action followed by an actor or set of actors in dealing with a problem or matter of concern.” Another standard view is by Peters:^[2] “public policy is the sum of government activities, whether acting directly or through agents, as it has an influence on the lives of citizens.”

POLICY TYPES

Over time, numerous policy experts have suggested that there are different types of policies, and that the political struggle over public policy depends on the type of policy at issue.^[3,4] One illustrative example is a typology created by James Q. Wilson.^[5] Table 1 indicates the two factors that define the typology—the extent to which benefits are concentrated or distributed, and the extent to which costs are concentrated or distributed among groups. The combination of costs and benefits, according to Wilson, helps to shape political discourse.

Majoritarian politics involves policy issues that have distributed (widespread) benefits and distributed costs. That is, many individuals and groups will pay for the policy to be implemented and services delivered, and many will benefit. This type of policy issue is visible and depends on development of a winning majority coalition in support of a policy choice. If a

widespread coalition determines that the issue calls for government involvement, then the policy is apt to be approved. This type of policy engenders societywide discussion and debate. A classic example would be social security, in which a vast number of citizens pay into the fund and from which large numbers of citizens at retirement will be able to receive payments.

Entrepreneurial politics is the result of a policy entrepreneur (or policy entrepreneurs) rallying support to a cause on behalf of a significant segment of the public and overcoming the opposition of a relatively small interest. In this case, costs are concentrated to a relatively small segment of society and benefits are widely distributed. An example would be Ralph Nader serving as a policy entrepreneur with respect to auto safety. The automobile industry had to invest in greater safety, to the benefit of the larger public.

Client politics occurs when a narrow interest group lobbies government successfully for some benefit without arousing mass opposition; the many pay for the benefit of the few. That is, society at large may pay for the benefits accruing to a few, as the recent farm bill, with subsidies to agricultural sectors coming out of general tax revenues, illustrates. A small sector—agricultural business—benefits, while the public pays for these benefits through taxes.

Interest group politics is a contest between two narrowly defined groups. This type of policy conflict is often invisible to the larger public because the contest is between two relatively narrowly defined groups. One group wins and another loses. An example might be National Labor Relations Board decisions, when one side will win and one will lose, with the debate taking place largely out of public scrutiny.

THE BLURRED BOUNDARIES BETWEEN PUBLIC AND PRIVATE POLICY

Public policy, for some analysts, is more broadly defined than the textbook definitions at the outset of this entry. Some have argued that public policy speaks to decisions made by powerful actors that affect the lives of many people. These decisions have to be intentional and they must bind those people affected by the decisions; that is, once the decision is made, affected

Table 1 James Q. Wilson's matrix: interest groups and policy types

Perceived benefits	Perceived costs	
	Distributed	Concentrated
Distributed	Majoritarian politics	Entrepreneurial politics
Concentrated	Client politics	Interest group politics

people must go along or suffer consequences. By this definition, decisions made by large and powerful corporations are as much public policy as decisions made by the United States, the state of New York, or the city of Chicago.^[6,7]

Normally, those who study public policy restrict the definition to decisions made by government, but the distinction between private and public sectors is untenable, according to critics. More important, it shields from inquiry decisions made by non-government organizations that are as much public policy as anything that government does. In this manner, such organizations are not responsible or accountable to the people. In a democracy, the people are supposedly the ultimate repository of authority over public policy.

By refusing to see that, for instance, corporations make policy as much as government does, society allows them to continue to operate outside democratic norms, which would call for public policy makers to be responsible and accountable to the people. If businesses make public policy, then—according to the democratic rules of the game—they ought to be directly responsible and accountable to the people as well.

First of all, though, how can it be said that non-governmental actors actually make policy? Some definitions of policy may not specifically mention government, but do emphasize the idea of decisions being made that affect substantial numbers of people.^[8] Once stated in this manner, according to critics, it becomes clear that non-government actors can make policy.

Mark Nadel argued that corporations often make decisions that are intentional, binding, and affecting large numbers of people with respect to some tangible or symbolic allocation of goods or services.^[6,9] United States Steel was once one of the giant American corporations. One of its landmark factories was in Gary, Indiana. Its furnaces belched smoke into the air, affecting the quality of life, health, and property of those living downwind. The company did not address issues of cleaning up its emissions. Thus, it made an intentional decision, binding the people of Gary to that decision and affecting many people with considerable consequences in the process.^[10] That this is public policy,

just as a change in income tax structure or promulgation of some regulation, seems obvious to critics like Nadel. Nadel also noted the decision by the major auto companies in the mid-1950s not to research antipollution technology.^[9] This is as much environmental policy as the Clean Air Act.

Private sector public policy is hidden from normal political discourse. People so commonly define public policy in terms of government actions that what we think of as “private” policy remains hidden. The costs of this are substantial. Factories shut down operations in a community, often devastating a local economy. The decision is considered “private,” even though its consequences are no different from a military base shutting down or a government program ceasing to spend money on a project in a district. People's lives are uprooted; local government's ability to pay for services is seriously compromised because such companies are normally important contributors to the local tax base.

CONCLUSION

This suggests that students of public policy might want to consider more generally the contours of the subject matter. Although the standard conceptualization of public policy as a government decision makes much sense, considering the implications of the expanded definition, along the lines suggested by Dahl and Nadel, might well further enrich understanding of the subject matter.

REFERENCES

1. Anderson, J.E. *Public Policymaking*; Houghton Mifflin: Boston, 2000; 4 pp.
2. Peters, B.G. *American Public Policy*, 3rd Ed.; Chatham House: Chatham, NJ, 1993; 4 pp.
3. Lowi, T.J. American business, public policy, case studies, and political theory. *World Polit.* **1964**, *16*, 677–715.
4. Peterson, P. *City Limits*; University of Chicago Press: Chicago, IL, 1980.
5. Wilson, J.Q. *American Government*; Houghton Mifflin: Boston, 2001.
6. Nadel, M. *Corporations and Political Accountability*; D.C. Heath: Lexington, MA, 1976.
7. Dahl, R.A. *After the Revolution*; Yale University Press: New Haven, 1971.
8. Easton, D. *A Framework for Political Analysis*; Prentice-Hall: Englewood Cliffs, NJ, 1965.
9. Nadel, M. The hidden dimension of public policy. *J. Polit.* **1975**, *37*, 2–34.
10. Crenson, M. *The Unpolitics of Air Pollution*; Johns Hopkins Press: Baltimore, MD, 1971.

Public Policy Formation Under Conditions of Ambiguity

Nikolaos Zahariadis

Department of Government, University of Alabama at Birmingham,
Birmingham, Alabama, U.S.A.

INTRODUCTION

Multiple Streams (MS) is a framework, approach, or lens—I use the terms interchangeably—that can be used to understand (predict and explain) the policy process under conditions of ambiguity. Although it could conceivably be extended to cover the entire process of policy making, it has been developed and applied only as an aid to understanding policy formation (agenda setting and decision making). Drawing inspiration from work on bounded rationality and organization theory as formulated by theorists such as James March, John Kingdon, Herbert Simon, and others, the framework assumes conditions of ambiguity and a temporal sorting order. *The main argument is that policies are the result of problems, solutions, and politics coupled, or joined together, by policy entrepreneurs during open windows of opportunity.*

ASSUMPTIONS AND BASIC LOGIC OF THE FRAMEWORK

MS theorizes at the systemic level, and it incorporates the entire system or a separate decision as the units of analysis. Much like systems theory, it views choice as collective output formulated by the push and pull of several factors. It shares common ground with chaos theories in being attentive to complexity, in assuming a considerable amount of residual randomness, and in viewing systems as constantly evolving and not necessarily settling into equilibria.^[1] Collective choices are not merely the derivatives of individuals but rather the combined result of structural forces and internal processes that are highly context dependent.^[2]

Ambiguity

Ambiguity refers to “a state of having many ways of thinking about the same circumstances or phenomena.”^[3] These ways may not be reconcilable, creating vagueness, confusion, and stress. It is different from uncertainty, a related concept, in that the latter refers to the inability to accurately predict an event.

Ambiguity may be thought of as ambivalence, whereas uncertainty may be referred to as ignorance or imprecision.^[4] While more information may (or not) reduce uncertainty,^[5] more information does not reduce ambiguity. For example, more information can tell us how AIDS is spread, but it still will not tell us whether AIDS is a health, educational, political, or moral issue. MS is a lens that deals with policy making under conditions of ambiguity. It does not explain why politicians sometimes pursue ambiguity but how policies are made when ambiguity is present.

At the heart of the lens lies a garbage can model of choice. In an insightful article, Cohen, March, and Olsen^[6] outline such a model that explains decision making in organizations—or decision situations—they term “organized anarchies” where ambiguity is rampant. These organizations, such as universities or national governments, are characterized by three general properties: fluid participation, problematic preferences, and unclear technology.

First, participation in such organizations is fluid. Turnover is high, and participants drift from one decision to the next. Legislators come and go, and bureaucrats, especially high-level civil servants, often move from public service to private practice.

Second, it is generally observed that people frequently do not have a clear, well-articulated set of preferences. To say that policy makers almost never make their objectives crystal clear is hardly novel, but it is true that quite often, time constraints force politicians to make decisions without having formulated precise preferences. Decisions are made and may even be facilitated by opaqueness. In contrast to most business firms in which the ultimate goal is clear—namely, to make a profit—organized anarchies operate on the basis of multiple and diverse objectives.

Third, technology, that is, an organization’s processes that turn inputs into products, is unclear. Members of an organized anarchy may be well aware of their individual responsibilities, but they exhibit only rudimentary knowledge concerning the way their job fits into the overall mission of the organization. Jurisdictional boundaries are unclear, and turf battles between different ministries are a common occurrence.

Temporal Sorting

Under such extreme conditions, theories based on rational behavior are of limited utility. Because problems and preferences are not well known, selecting the alternative that yields the most net benefits becomes an impossible task. Problems are difficult to define precisely. The problem under conditions of ambiguity is that we do not know what the problem is. Its definition is vague and shifting. In contrast to models that stress consequential action, the logic of multiple streams is based on temporal order. Choice has more to do with the simultaneous evocation of problems and solutions rather than with any inherent correlation between them. Time is a unique, scarce resource. Because the primary concern of decision makers (policy makers, business executives, or top civil servants) is to manage time effectively rather than to manage tasks,^[7,8] it is reasonable to pursue a lens that accords significance to time rather than rationality. This translates into a process in which individuals are viewed as less capable of choosing the issues they would like to solve and more concerned with addressing the multitude of problems that are thrust upon them largely by factors beyond their control.

Assumptions

Three assumptions guide the framework:

1. *Individual attention or processing is serial; systemic attention or processing is parallel.* On the one hand, because of biological and cognitive limitations, individuals can only attend to one issue at a time. On the other hand, division of labor in organizations or governments enables them to attend to many issues simultaneously. This capacity is, of course, not infinite, but it is reasonable to assume that government can put out fires in Colorado, conduct trade negotiations with Greece, and investigate mail fraud at the same time.
2. *Policy makers operate under significant time constraints.* Obviously, these people do not have the luxury of taking their time to make a decision. While this does not imply that all decisions are crises, it suggests that there is a sense of urgency in addressing them. In effect, such time constraints limit the range and number of alternatives to which attention is paid.
3. *The streams flowing through the system are independent.* This assumption is related to the first one in that if systems can do things in parallel, then each element or stream may be conceived as having a life of its own. The

framework postulates three separate streams flowing through a system: problems, solutions (policies), and politics. Problems are concerns that individuals inside and outside the organization have. They involve, but are not limited to, such issues as profitability, work satisfaction, career advancement, and promotion. Solutions are people's products; they are answers to questions that may not be generated only when needed. Thus, each stream seems to obey its own rules and flows largely independently of the others.

THE STRUCTURE OF THE THREE STREAMS

The basic formulation was first put forth by Kingdon,^[9] who used the multiple streams framework to illuminate two predecision processes in the United States: agenda setting and alternative specification. His work has been subsequently revised to examine the entire policy formation process (agenda setting and decision making) in a comparative context and in domestic as well as foreign policy. I proceed by examining the structure of each stream separately, and then I assess the effects of coupling in aiding our understanding of policy formation.

The condensed version of the argument is as follows: *There are three independent streams flowing through every policy-making system. There is a problem stream that includes indicators or crisis events that come to be defined as problems by policy makers. The second stream is the policy, or solution, stream, which comprises ideas generated in narrow policy networks. The third is the political stream that contains the administration, Congressional leaders, etc. The chances that any one idea will be adopted increase dramatically when skillful policy entrepreneurs couple or join together the three streams during open policy windows, defined as changes in the problem or political streams. So, the key to understanding how policies are made under conditions of ambiguity is to elucidate the dynamics of each stream and to specify the circumstances under which coupling is more likely to occur.*

The problem stream consists of various conditions that policy makers and citizens want addressed. Examples are government budget deficits, environmental disasters, inflation, rising medical costs, and so on. The means that policy makers use to find out about these conditions include indicators, focusing events, and feedback. Indicators show the magnitude of a condition as well as the scope of change. An example of a problem is high government borrowing needs.^[10,11] The British Conservatives used the Public Sector Borrowing Requirement to identify those needs. The fact

that it rose to historically high levels by the early 1980s provided impetus to do something about it. The same could be argued about the French. Their borrowing needs were judged to be at historically high levels. But, they were lower than those of Britain. The point that needs to be stressed is that each government looked at the historical record and not to other countries to make this judgement, although cross-national comparisons are often made.^[9] Focusing events also draw attention to problematic conditions. Examples of focusing events include prolonged strikes and train or airplane accidents. Feedback from previous programs is important in that it helps highlight what works and what may not. Successfully implementing a solution in one area may spill over to another, facilitating the adoption of the same solution in a seemingly unrelated area. The case of spillover of privatization from the area of oil to telecommunications in Britain is a good example.^[11]

The politics stream includes a number of political parties or coalitions competing for a chance to govern and thus directly affect the policy process. They are the policy makers who will ultimately make the decisions whether to adopt or not to adopt a specific proposal. Collectively, parties carry an ideological bend that makes them more receptive to certain ideas. Conservatives generally have a proclivity toward more market-oriented policies and socialists/leftists do not, with several exceptions, of course. There are at least two elements in the political stream: national mood and party ideology. Kingdon^[9] advocates keeping them separate, although Zahariadis^[11] proposes they be combined because of significant problems operationalizing the national mood.

The policy stream includes a “soup” of ideas that compete to win acceptance in policy networks. Two criteria affect the availability of proposals in the policy stream: value acceptability and technical feasibility. They are important in securing the chances that any one option will receive attention past the subsystemic level. Ideas that do not conform to the values of many specialists in a policy network and which are more difficult to implement stand a lesser chance of receiving attention by national policy makers.

Not all policy networks are created equal. In the original conception of multiple streams, ideas are recombined and rise to the top only incrementally.^[9] But this need not be the case. An examination of the rise (or not) of privatization in British and German policy networks revealed systematic differences depending on the structure of the networks.^[12] Institutional configurations, or to put it differently, the level of integration, differ across countries, affecting the mode and tempo of ideas, that is, how ideas germinate in the policy stream and how fast they rise to prominence. Integration refers to linkages among participants and is

distinguished by variations in four dimensions: size, mode, capacity, and access. Based on these dimensions, networks can be classified as more or less integrated. Less-integrated networks are larger in size, have a competitive mode, lower administrative capacity, and less restricted access. Conversely, more integrated networks are smaller in size, have a consensual mode, higher capacity, and more restricted access. Less integrated networks are likely to encourage a movement of solutions from conception to prominence through quantum to gradualist evolution, whereas alternatives in more integrated networks are more likely to follow an emergent to convergent evolution.

Policy entrepreneurs are individuals or corporate actors who attempt to couple the three streams. They are more than mere advocates of particular solutions; they are power brokers and manipulators of problematic preferences and unclear technology. Of course, not all entrepreneurs will be successful at all times. The more successful ones are those who have greater access to policy makers. For example, the Adam Smith Institute had greater access to the government during Thatcher's tenure in power in the United Kingdom, because their ideologies matched more closely than those of other groups. Hence, options put forth by individuals associated with the Institute had a greater chance of falling on receptive ears in government.^[13] Second, entrepreneurs with more resources, i.e., the ability to spend more time, money, and energy, to push their proposals, have greater rates of success.

So far, the analysis has focused on entrepreneurial characteristics. Entrepreneurs must also employ strategies to accomplish their goal of coupling the three streams. The following list of strategies is not exhaustive, but it gives a good idea of how entrepreneurs pursue coupling and what factors influence their success: framing, salami tactics, and the strategic use of symbols.^[14] All aim to manipulate the various dimensions of policy images to activate values within with the explicit aim of gaining a receptive audience among policy makers. These strategies are employed largely at the systemic or national level. Framing involves the manipulation of dimensions to represent solutions to specific problems as gains or losses. Entrepreneurs seek to capitalize from people's tendency to hate to lose more than they like to win. If solutions espoused by entrepreneurs are large deviations from the status quo, successful coupling is more likely when problems are represented as losses; if options represent smaller deviations or a preservation of the status quo, coupling success is more likely when problems are framed as gains. A salami tactic is the strategic manipulation of sequential decision making. Entrepreneurs are assumed to have a grand design of the desired outcome. However, because they are reasonably certain

their desired solution will not be adopted because it is too risky, they slice the process into distinct stages. Doing so encourages agreement in steps. By the time policy makers become aware they are being manipulated, they have invested too much to depart from the course of action. Entrepreneurs who are placed at a high level in government, operate under crisis conditions, and pursue salami tactics are more likely to be successful at coupling. Finally, symbols bias choice through emotion and cognitive meaning. They allow entrepreneurs to link the macro to the micro by using the same symbol to appeal to many individuals. Some symbols have predictable effects on individuals in terms of affect and meaning, particularly those commonly termed as higher-order symbols. Higher-order symbols, such as the national anthem or the flag, are relevant to the community at large and possess two characteristics. They are capable of arousing more intense emotions than others and are closer or more relevant to an individual's identity. Entrepreneurs who use higher-order symbols have a greater chance of finding a receptive political audience.

Success in coupling also depends on other factors. Coupling is more likely to be successful during open policy windows. Windows are ephemeral opportunities created by changes in the problem and politics streams. They provide the context within which coupling takes place. The mechanisms that link opportunity to choice are three: urgency, fear, and crisis-mandate. The argument put forth by Kingdon^[9] was that neither problems nor solutions drove the policy process, although he emphasizes the point that solutions frequently search for problems rather than the other way around. But, the argument can be refined by specifying the conditions under which a policy is in search of a rationale.^[13] When windows open in the problem stream, coupling is more likely to be consequential, that is, finding a solution to a given problem. When windows open in the politics stream, coupling is more likely to be doctrinal, that is, finding a problem to a given solution.

SPECIFIC ISSUES AND CONCERNS

The lens can be extended to cover the entire process of policy formation, enable the comparative study of public policy, accommodate a considerable degree of complexity, model a process that is fluid and dynamic, and impose some degree of order and structure while acknowledging the importance of serendipity. The critics, however, point to several drawbacks. I will reflect upon three questions that have been raised: 1) are the streams really independent?; 2) are open policy windows always ephemeral?; and 3) what is the impact of deadlines?

Are the Streams Really Independent?

In a penetrating critique, Gary Mucciaroni^[15] questions the appropriateness of conceptualizing independent streams. The streams can be more fruitfully viewed as interdependent, he maintains, and changes in one stream can trigger or reinforce changes in another, making coupling much less fortuitous and the process more purposive and strategic. For example, the problem of U.S. tax reform was tied to the supply-side tax cuts proposed by conservatives in symbolic and substantive ways long before Reagan's advent to power opened a policy window. Weir^[16] echoed a similar point. Sabatier and Jenkins-Smith's^[17] advocacy coalition framework is partly an attempt to view the politics and policy streams as tightly linked. In other words, coupling, Mucciaroni^[15] concluded, takes place even in the absence of a policy window. Kingdon^[9] opened the possibility that coupling may take place in the absence of an open window, but he was careful to point out that this must be determined empirically.

Problems, policies, and politics are independent of one another in several ways. Politics operate on the systemic level of analysis, policies are developed on a sectoral or subsystemic level, and problems may be developed on a systemic or a sectoral level. Because the framework views choice from the system's point of view, it is possible to conceptualize several things happening simultaneously and independently of one another (first assumption). The policy and political streams are independent, because they involve mostly different people with different preoccupations and orientations.^[9] Specialists in policy communities include experts who share an interest in one narrow area, e.g., telecommunications, health, education, etc., and who concentrate on honing proposals relevant to the area, attend to technical details, conduct cost-benefit analyses, and the like. The political stream involves a wider array of people, most of whom have nothing to do with this narrow area. It includes the entire government, for example, agriculture, finance, or transport ministers, party activists, or religious authorities. They are preoccupied with elections, public opinion, or military operations, which are preoccupations quite distinct from, say, health issues. Solutions are developed in narrow communities and not in broad political discourse, where grandstanding and simplicity rather than technical detail are the key ingredients.

The problem stream differs from the policy stream in that solutions are developed separately from problems. This does not mean that solutions are not developed to solve problems, it simply means that the same solution may be attached to different problems at different times. The example of academic discourse illustrates this distinction well. The garbage can model was conceived as a solution to the problem of

organizational decision making under ambiguity. Kingdon^[9] adapted the same solution to the problem of agenda setting in the U.S. federal government. Another analyst may further adapt the same solution to describe the implementation process of UN decisions. Problems change, while the solution remains the same. Feldman^[3] provided a multitude of examples whereby bureaucratic analysts drafted reports and estimates before any problems arose on the rationale that developing them would take much longer than the time available if they waited to draft the reports after a request was made. Such practice leads to the development of inventories of solutions to address new problems.

Are Open Policy Windows Always Ephemeral?

A policy window is a temporal stimulus for choice. Windows not only signal the temporal limits of the coupling process to policy entrepreneurs, but they also help structure the search for appropriate problems or policies that may be palatable to the policy makers of the day.

Windows may sometimes not be temporary. Sharp^[18] suggests that certain characteristics make an issue continuously activated, i.e., it is always on the agenda, but she does not elaborate on the implications for MS. An issue such as drugs and drug policy has dramatic potential, high visibility, a plethora of organized interests that constantly offer expert diagnoses and prescriptions, and a multitude of consequences—social, personal, political, economic, or physiological—which make it a favorite supplement to a variety of other issues. If certain windows are not fleeting, what are the effects on coupling over time? Do they lead to greater partisan conflict, more symbolic action, or more problem solving?

What is the Impact of Deadlines?

The main way that decisions are made in MS is by entrepreneurial skill. This implies that there are no set deadlines on when decisions need to be made. Weiner^[19] points out that deadlines have a significant impact on the way the lens works. They erect a barrier around the choice that resists the entrance of new problems or the activation of new participants. This causes a dramatic increase in the amount of energy and attention that active participants must exert to make the choice, because they have to withdraw from other choices to stay on top of their obligations to meet the deadline or they must withdraw from making the choice before them. When a deadline is set, policy makers compare the amount of energy it would require to solve or address all problems they have set to solve

and the energy required to make this particular decision. Thus, if they decide they have accepted more problems than they can handle under the new time constraints, they will likely reject some problems and allow them to be addressed in other parts of the system.

Decision cycles are equally important factors to consider, as they close off the process and reduce political stress and instability. The lens paints a fluid picture of an open-ended policy-making process. While a multitude of solutions and problems are dumped into the garbage can, the process has a limit. Cycles of opportunities limit the focusing frames around which groups coalesce, reduce the escalation of demands by winners and losers, and meld differences among competing groups.^[20] Clearly defined cycles, such as the annual passage of the budget, provide a signal as to the duration of the window. To an extent, they regulate conflict by providing multiple opportunities to air out demands. This, in turn, limits options for dissent, because dissent can be spread over several years, and because not every problem can be aired at that point. Past practice suggests which problems and solutions or their relatives may be appropriately debated. The intensity of conflict is also suppressed, because defeat at one point is not necessarily the end but rather the beginning of a new effort to “get your way next time.” Similarly, success is merely temporary and not overwhelming, because another budget and, consequently, another compromise will have to be made next year. Such repetition also leads to a better understanding of what opponents want, which tends to reduce differences among frames held by individuals and to aggregate them into broad coalitions. For example, the annual debate over salaries or other benefits helps distribute information about various grievances. Individual members learn more about how other individuals feel about the same issue and are better capable of coalescing around common themes. From an individual member’s point of view, cycles eventually promote the creation of coalitions, because there is greater likelihood for success in numbers. From the system’s point of view, cycles reduce diversity and conflict to a more manageable level. In short, cycles are hypothesized as promoting order and coordination in otherwise disorderly organized anarchies.

MULTIPLE STREAMS AND THE STAGES HEURISTIC

By virtue of focusing attention on the entire policy formation process (agenda setting and decision making), multiple streams enable research to move beyond the stages heuristic, which is the dominant framework for understanding the policy process. It provides a better

theory of the policy process by facilitating cumulation of knowledge and integration of models across various stages that have been previously examined separately and in isolation. The end result is the ability to explain more with less.

Early in the development of policy studies as a field of inquiry, Lasswell^[21] proposed a conceptual map that would enable analysts to maintain a general image of the process of collective choice, while cutting the process down to analytically manageable proportions. The list of stages he proposed, which was subsequently refined by others, is important not simply for practical reasons, but primarily for theoretical ones. Each stage, such as initiation, implementation, evaluation, and so on, has a distinctive characteristic that gives it “a life and presence of its own.”^[22] It is argued that the actors involved in each stage are different as well as the purpose and dynamics of activities, making the distinction between stages an analytically useful enterprise.^[23]

While the stages heuristic has served analysts well by stimulating considerable research, it has been criticized recently as conceptually inadequate and stifling rather than facilitating theory development. Nakamura^[24] views the issue as one of boundary specification. Are the boundaries between the stages as precisely defined as their proponents suggest, and if not, is this not evidence that the distinctiveness is artificially created? Sabatier^[25,26] and Sabatier and Jenkins-Smith^[17] go further. Apart from questioning the assumption that there is one rather than many interacting cycles, they argue that the stages heuristic is not a causal theory but a taxonomy, which can be misleading, because research in each stage proceeds without reference to work in other stages. Moreover, they maintain that the description of stages is often inaccurate. Each stage is affected by others. For example, agenda setting is affected by evaluation of existing programs; implementation difficulties influence the likelihood of adopting specific legislation. Unfortunately, it appears “the map” initially created to bring a broad overview of the process without sacrificing analytical manageability has strayed off its original intention. Models and theories are proposed and tested for each stage in isolation of other stages, fostering disjointed theory development. As a result, the broad general picture of the process is lost in a sea of stage models which, while acknowledging their interconnectedness, nevertheless discourage any effort to link them in a theoretically meaningful way.

Multiple streams offer a way to overcome the limitations of the stage heuristic. While it does not cover the entire policy process, although it conceivably could, it provides a way to understand the policy-making process of two stages with one model of reference. First, it does not draw boundaries between stages. The benefit from such “non-distinction” is the ability to develop

concepts and processes that are internally consistent and meaningful across stages. It promotes rather than impedes theoretical links between stages. Second, it provides a better, clearer view of the larger process. Third, it elaborates on causal mechanisms that give rise to theoretical coherence and hypothesis testing. Fourth, multiple streams is a better alternative to the stage heuristic in that it helps explain a greater portion of the policy process by reference to the same factors. If one accepts the argument put forth by Lakatos^[27] that cumulation of knowledge is the basic aim of science and that parsimony is of great value, then the ability to use the same set of factors to explain phenomena that used to demand two or more models gives multiple streams a qualitative edge. Although many conceptual refinements have yet to be made, relative to the stage heuristic, multiple streams is a better theory.

CONCLUSION

Multiple streams is a lens that examines the process of making policies under conditions of ambiguity. According to multiple streams (MS), policies are the result of actions by policy entrepreneurs who couple three streams—problems, policies, and politics—during open windows of opportunity. In this concluding section, I chart a course for future research.

Does policy affect politics? Are there issue properties that alter the policy process in predictable ways?^[15,28] Rommetveit^[29] suggests that prime candidates for garbage cans are those decision processes that take place under changes in normative structures (basic value-priorities in a polity) and when no active participant dominates the policy process. When a society is in the process of reordering its values, established norms that underlie state–society relations are challenged. As a result, the role of experts and the wisdom of bureaucrats is questioned, and dissenting groups might use this opportunity to pursue their diverging claims. The possible activation of new groups and the wide disagreement as to the relevant values upon which to base the decision will, in turn, increase ambiguity and permit the evocation or appearance of new problems and solutions. Such desegmentation of previously established links between windows, problems, and politics complicates the process as new and perhaps unrelated elements are dumped into the can. In this light, the issues of privatization or government reform are good candidates for applying MS.^[30,31]

MS has been investigated so far as if decisions are discrete and unrelated to one another. Policy making, however, involves sequences of decisions that are linked together through the important process of feedback. Kingdon^[9] noted that feedback plays a pivotal

role in helping policy makers define or redefine problems. Feedback, however, may operate in all three streams as well as in coupling. Experts in policy communities formulate solutions partly on the basis of previous experience with similar problems. Entrepreneurs keep inventories of coupling techniques and experience. Precedent makes it easier to transplant issues or success from one area to the next. There is a lot of learning going on in public policy, and feedback plays an important role in fostering it. Exploring further into the mechanisms that facilitate this process might yield insights into entrepreneurial strategy and sense of timing. What effect does sequential coupling have on the lens? How do entrepreneurs come to recognize opportunities, and what influences their timing strategies? When are old decisions likely to be reexamined?

How should policy makers cope with an ambiguous world? What lessons does MS have to offer to democratic governance? Although reasonable and obvious, these questions have not really been explored. In order for democracy to work properly, there needs to be a strong sense of accountability and efficacy. Policy makers are supposed to make decisions that conform to the wishes of at least the majority of the citizenry, but they must also solve as many problems confronting the polity as possible. In an earlier study of leadership in universities, a situation slightly different from democracies in that university presidents are hired and not elected, Cohen and March^[32] suggested *inter alia* persistence and overloading the system by making many proposals on the rationale that some will pass. Looking specifically at bureaucratic hierarchies, Padgett^[33] concluded that managers had better follow a “hands off” approach, provided they tend to personnel policies and structural design. For example, presidents should not fight subunit conflict but rather reorganize high- and low-saliency issues into relevant divisions and then hire activist and conservative secretaries (or vice presidents) to deal with the respective divisions. High saliency issues tend to flow upward in the hierarchy, and they are precisely those issues in which upper management attempts to be involved. It is best to hire someone who seeks that involvement. Low saliency issues tend to be addressed at a lower level and require a conservative manager who is not obtrusive. Crecine^[34] added to this strategy the manipulation of windows, which in effect alters the context. The lens assumes windows are opened by changes in the political or problem streams.^[9] Some windows are subject to manipulation, particularly those associated with budgets or public procurement efforts. The same can be said of elections. They can be called in parliamentary democracies at the prime minister’s discretion. For example, Margaret Thatcher’s 1983 decision to hold legislative elections a year ahead of schedule caught the opposition divided and

busy with the fallout from the Falkland War. Her subsequent victory enabled her to pass the politically costly and complex bill authorizing the sale of British Telecom with relative ease.^[10]

Democracies differ in that policy makers flow in and out of the system as one is elected (or reelected) to the post and the others are not. What are the implications for governance? March and Olsen^[35] sketched some ideas, but they did not model them. Zahariadis^[14] suggested that the policy process is replete with manipulation of facts and people, but he did not draw implications for democratic governance. While the multiple streams framework has not yet provided answers to these important questions, it is a good starting point for investigating policy making under conditions of ambiguity.

REFERENCES

1. Kingdon, J.W. Agendas, ideas, and policy change. In *New Perspectives on American Politics*; Dodd, L.C., Jilison, C., Eds.; CQ Press: Washington, DC, 1994; 215–229.
2. Ashford, D., Ed.; *History and Context in Comparative Public Policy*; University of Pittsburgh Press: Pittsburgh, 1992.
3. Feldman, M.S. *Order Without Design: Information Production and Policy Making*; Stanford University Press: Stanford, 1989.
4. March, J.G. *A Primer on Decision Making*; Free Press: New York, 1994.
5. Wilson, J.Q. *Bureaucracy*; Basic: New York, 1989.
6. Cohen, M.D.; March, J.G.; Olsen, J.P. A garbage can model of organizational choice. *Adm. Sci. Q.* **1972**, *17* (1), 1–25.
7. Drucker, P. *The Effective Executive*; Random House: New York, 1967.
8. Mackenzie, R.A. *The Time Trap*; AMACOM: New York, 1972.
9. Kingdon, J.W. *Agendas, Alternatives and Public Policies*, 2nd Ed.; Harper Collins: New York, 1995.
10. Zahariadis, N. To sell or not to sell? Telecommunications policy in Britain and France. *J. Public Policy* **1992**, *12*, 355–376.
11. Zahariadis, N. *Markets, States, and Public Policies: Privatization in Britain and France*; University of Michigan Press: Ann Arbor, 1995.
12. Zahariadis, N.; Allen, C.S. Ideas, networks, and policy streams: privatization in Britain and Germany. *Policy Stud. Rev.* **1995**, *14* (1/2), 71–98.
13. Zahariadis, N. Selling British rail: an idea whose time has come? *Comp. Polit. Stud.* **1996**, *29*, 400–422.
14. Zahariadis, N. *Ambiguity and Choice in Public Policy*; Georgetown University Press: Washington, DC, 2003.
15. Mucciaroni, G. The garbage can model and the study of policy making: a critique. *Polity* **1992**, *24*, 459–482.
16. Weir, M. Ideas and the politics of bounded innovation. In *Structuring Politics: Historical Institutionalism in*

- Comparative Analysis*; Steinmo, S., Thelen, K., Longstreth, F., Eds.; Cambridge University Press: Cambridge, 1992; 188–216.
17. Sabatier, P.A., Jenkins-Smith, H.C., Eds.; *Policy Change and Learning: An Advocacy Coalition Approach*; Westview: Boulder, CO, 1993.
 18. Sharp, E.B. Paradoxes of national antidrug policymaking. In *The Politics of Problem Definition: Shaping the Policy Agenda*; Rochefort, D.A., Cobb, R.W., Eds.; University Press of Kansas: Lawrence, KS, 1994; 98–116.
 19. Weiner, S.S. Participation, deadlines, and choice. In *Ambiguity and Choice in Organizations*; March, J.G., Olsen, J.P., Eds.; Universitetsforlaget: Bergen, Norway, 1976; 225–250.
 20. Huff, A.S. Politics and argument as a means for coping with ambiguity and change. In *Managing Ambiguity and Change*; Pondy, L.R., Boland, R.J., Jr., Thomas, H., Eds.; John Wiley & Sons: New York, 1988; 79–90.
 21. Lasswell, H.D. *The Decision Process*; University of Maryland Press: College Park, MD, 1956.
 22. DeLeon, P. The stages approach to the policy process: What has been done? where is it going? In *Theories of the Policy Process*; Sabatier, P.A., Ed.; Westview: Boulder, CO, 1999; 19–32.
 23. Brewer, G.D.; DeLeon, P. *The Foundations of Policy Analysis*; Brooks/Cole: Monterey, CA, 1983.
 24. Nakamura, R. The textbook policy process and implementation research. *Policy Stud. Rev.* **1987**, 7 (2), 142–154.
 25. Sabatier, P.A. Toward better theories of the policy process. *PS Polit. Sci. Polit.* **1991**, 24, 147–156.
 26. Sabatier, P.A. The need for better theories. In *Theories of the Policy Process*; Sabatier, P.A., Ed.; Westview: Boulder, CO, 1999; 3–17.
 27. Lakatos, I. Falsification and the methodology of scientific research programmes. In *Criticism and the Growth of Knowledge*; Lakatos, I., Musgrave, A., Eds.; Cambridge University Press: Cambridge, 1970; 91–195.
 28. Lowi, T.E. Four systems of policy, politics, and choice. *Public Adm. Rev.* **1972**, 32, 298–310.
 29. Rommetveit, K. Decision making under changing norms. In *Ambiguity and Choice in Organizations*; March, J.G., Olsen, J.P., Eds.; Universitetsforlaget: Bergen, Norway, 1976; 140–155.
 30. Brunsson, N.; Olsen, J.P. *The Reforming Organization*; Routledge: London, 1993.
 31. March, J.G.; Olsen, J.P. Organizing political life: what administrative reorganization tells us about government. *Am. Polit. Sci. Rev.* **1983**, 77, 281–297.
 32. Cohen, M.D.; March, J.G. *Leadership and Ambiguity*; McGraw-Hill: New York, 1974.
 33. Padgett, J.F. Managing garbage can hierarchies. *Adm. Sci. Q.* **1980**, 25, 583–604.
 34. Crecine, J.P. Defense resource allocation: garbage can analysis of C3 procurement. In *Ambiguity and Command*; March, J.G., Weissinger-Baylon, R., Eds.; Pitman: Marshfield, MA, 1986; 72–119.
 35. March, J.G.; Olsen, J.P. *Democratic Governance*; Free Press: New York, 1995.

Public Procurement

Robert E. Lloyd

U.S. Department of State, Ft. Lauderdale, Florida, U.S.A.

INTRODUCTION

When private individuals or corporations buy things, they may use the terms “purchasing” or “buying” to describe their actions. In government organizations, many different terms and definitions are used when referring to the buying function, from “public procurement” to “contracting” or “acquisition” and many other variations. The wide variety of terms and definitional ambiguities permeate the practice and profession. Despite the fact that public procurement has a centuries-old history, clarity in what practitioners and scholars mean when they refer to “public procurement” is sometimes lacking. Nevertheless, there are common themes, roles, and responsibilities that span the wide range of governmental organizations involved in procurement.

LEGAL DEFINITIONS

Definition of the Federal Government

In 1984, the US Government converted to a new system of regulations governing how federal agencies buy things. The former Federal Procurement Regulations published by the General Services Administration (GSA) and the Defense Acquisition Regulation (known as the Armed Services Procurement Regulation until 1976) issued by the Department of Defense, along with the National Aeronautics and Space Administration Procurement Regulation were combined into a new set of rules called the Federal Acquisition Regulation (FAR; Title 48, Code of Federal Regulations, Chap. 1). There was a deliberate attempt to replace the term “procurement” with “acquisition,” in part to distinguish the new rules from the old, as noted in the following excerpt from the FAR (Section 2.101, Definitions):

“Acquisition” means the acquiring by contract with appropriated funds of supplies or services (including construction) by and for the use of the Federal Government through purchase or lease, whether the supplies or services are already in existence or must be created, developed, demonstrated, and evaluated. Acquisition begins at the point when agency needs

are established and includes the description of requirements to satisfy agency needs, solicitation and selection of sources, award of contracts, contract financing, contract performance, contract administration, and those technical and management functions directly related to the process of fulfilling agency needs by contract.

“Procurement (see “acquisition”).”

Although the terms “procurement” and “contracting” have generally fallen into disfavor at the federal level, they still persist in several forms. For example, the federal office responsible for Government-wide policy in the Office of Management and Budget is known as the Office of Federal Procurement Policy, and the job series for buyers is referred to as “contracting.” At the federal level, Congress passed a new legislation (the “Services Acquisition Reform Act” or Defense Authorization Act for 2004, Public Law 108-136, title XIV, Section 1411) that defines in detail what “acquisition” means:

“The term “acquisition”—

- A. means the process of acquiring, with appropriated funds, by contract for purchase or lease, property or services (including construction) that support the missions and goals of an executive agency, from the point at which the requirements of the executive agency are established in consultation with the chief acquisition officer of the executive agency; and
- B. includes—
 - i. the process of acquiring property or services that are already in existence, or that must be created, developed, demonstrated, and evaluated;
 - ii. the description of requirements to satisfy agency needs;
 - iii. solicitation and selection of sources;
 - iv. award of contracts;
 - v. contract performance;
 - vi. contract financing;
 - vii. management and measurement of contract performance through final delivery and payment; and

- viii. technical and management functions directly related to the process of fulfilling agency requirements by contract.”

Although the new definition is similar to that appearing in the FAR, by promulgating this new, more expansive definition, Congress signaled its intent to view the buying process in a broader perspective.

In addition, federal legislation also distinguishes a “procurement contract” from a grant or cooperative agreement. A contract is to be used when “the principal purpose of the instrument is to acquire (by purchase, lease, or barter) property or services for the direct benefit or use of the United States Government” (31 U.S.C. 6303). Grants and cooperative agreements, on the other hand, involve a transfer of funds to a source outside the federal government for a public purpose, such as research.

Definitions of the State and Local Government

The American Bar Association’s Model Procurement Code^[1] defines “procurement” as “buying, purchasing, renting, leasing, or otherwise acquiring any supply, service, or construction, including description of requirements, selection and solicitation of sources, preparation and award of contract, and all phases of contract administration.” Many jurisdictions have adopted all or part of this model for their public procurement rules. For example, the State of Arizona’s Revised Statutes (Article 1, Section 41-2503, paragraph 15), made effective in 1985, adopt this definition for public procurement by the state.

GOALS AND PROCEDURES

Public procurement is often most visible owing to the strict procedures it follows. At all levels of government, the emphasis is on transparency, enabling the public to learn of contracting opportunities, and facilitating competitive private sector responses. The three primary goals of public procurement are fairness, efficiency, and promotion of socioeconomic objectives. Public procurement is held to a higher standard than private sector buying in that governments are expected or required to publish advance notices of upcoming contracts. Fairness demands that the public be allowed to participate, and Congress passed the Competition in Contracting Act of 1984 to mandate “full and open competition” or allowing all responsible sources to compete for contracts awarded by federal agencies.

Efficiency is likewise a goal of public procurement, as revenue from income taxes is being spent and must be safeguarded from waste. Many efficiencies have been introduced in public procurement in recent years, including the acquisition of commercial items rather than items meeting government-unique specifications, expanded use of simplified acquisition techniques and electronic commerce tools such as posting solicitations for upcoming contracts on internet web sites, and the implementation of purchase cards to simplify transactions of small value. For example, the federal purchase card program allows qualified cardholders to buy using a bank-issued card up to the “micropurchase” limit of \$2500 per transaction.

As a result of the large amount spent in public procurement (over \$200 billion per year at the US federal level, and State and local government procurement exceeds even that amount in the aggregate), many see public procurement’s role as a means of implementing socioeconomic policy.^[2] Public procurement is often used as a means of leveraging this significant purchasing power. In addition to the goals of fairness and efficiency, the added aims of assisting small businesses (as described in considerable detail in FAR Part 19), promoting the buying of environmentally friendly products, implementing wage increases and fringe benefit packages, as in the “living wage” ordinances found in local government procurement,^[3] and similar programs comprise an important aspect of public procurement.

The operating procedures followed in public procurement generally fall into three phases:

- i. *Planning, publicizing, and soliciting:* During this stage, the public procurement office works with its clients to develop a complete description of the government’s needs. The procurement office then publishes some form of public notice regarding the upcoming contract (on an internet web site such as www.fedbizopps.gov, by newspaper advertisement, etc.). A written solicitation document is then published describing the government’s needs and asking for responses from the private sector. The most common types of competitive solicitations are the Request for Proposals (RFP) and the Invitation for Bids (IFBs). RFPs (negotiated procurement, which allows negotiation with those submitting proposals) are typically used for more complex efforts, and IFBs (which involve a public bid opening and no negotiation) are normally limited to contracts for routine supplies or construction. Federal agencies generally use RFPs and rarely use IFBs, while the situation is sometimes the reverse in local government procurement. The most important functions of the solicitation are to define the government’s

needs, explain how to submit an offer, outline how the government will evaluate the responses and pick the winning firm, and include the clauses, terms, and conditions that will become part of the contract.

- ii. *Evaluation and award:* Once the proposals or bids have been received, they must be fairly evaluated by the procurement office and its customers. A contract may be awarded based on the initial response from the offering firms, or in the case of a negotiated procurement, the government may choose to conduct negotiations with one or more of the highest rated firms. This phase of public procurement is also known as source selection, and its intent is to choose the most advantageous contractor after following the procedures stated in the solicitation. For negotiated procurement, the procurement office typically adheres to an evaluation plan that describes how the proposals will be evaluated to ensure impartiality in the process. Ethical issues are addressed during this phase as well, such as avoiding conflicts of interest by requiring proposal evaluators to certify that they do not own stock in any of the competing companies. At the conclusion of the evaluation, someone with delegated contracting authority will sign the contract with the successful firm. During this process, disappointed firms may have an opportunity to file a protest if they believe they were unfairly treated.
- iii. *Contract administration:* After the contract is signed, the procurement office must administer it. That means monitoring the work performed by the contractor, including the important functions of receipt, inspection, and acceptance prior to payment. Warranties must be administered, such as when computer equipment is purchased, or for construction contracts that often include a one-year warranty for the work and extended warranties for equipment. Public procurement may involve resolving contract disputes arising during contract performance. Disputes may end up in court or in a designated board of contract appeals. Some contracts extend for many years, with either annual renewal options, or they are awarded on a multiyear basis. At the conclusion of the contract, it must be closed out so that all outstanding funding and performance issues are resolved.

Separate procedures exist for different types of public procurement, such as when buying architect-engineering services, which uses a qualifications-based selection procedure (known in the federal government as “Brooks Act procedures”) rather than price

competition. Also, most jurisdictions allow the use of simplified procedures for small dollar value purchases. For federal agencies, a transaction of \$100,000 or less may be conducted by requesting quotations over the telephone or by electronic mail, resulting in an abbreviated purchase order rather than a lengthy contract. (This amount increases to \$5 million if the item being purchased is a commercial item sold to the general public.) Federal agencies may place delivery orders under Federal Supply Schedule contracts issued by GSA on a streamlined basis for supplies and services. State and local governments that are interested in cooperative purchasing for information technology items may also use the GSA schedules, as a result of recent legislation. Although the emphasis in public procurement has long been toward greater competition in contracting, noncompetitive contracts may also be awarded, and these are typically subject to higher level approvals than other contracts.

POSITIVE AND NORMATIVE DEFINITIONS

One approach to defining public procurement is to examine what its practitioners do on a daily basis. The body of knowledge defined by GSA’s Federal Acquisition Institute^[4] (FAI) and the various professional associations in public procurement can be considered a positive approach to what public procurement is. Johnson, Leenders, and McCue^[5] describe the results of a survey of governmental purchasing groups in terms of organizational structure, responsibilities, reporting relationships, team-based procurement, and involvement in major organizational activities. The authors found significant differences in the responsibilities of public procurement buyers in the supply chain, namely, there was a lower level of involvement in the planning process as well as risk management by public sector procurement officials than their private sector counterparts.

Discussion of public procurement often involves opinion as to what public procurement “should” be, as noted by various commentators:

Public Procurement as Business

One common thread in all public procurement is that it is carried out by obtaining needed items from businesses. Recent commentators such as Rindner^[6] have advocated the view that “acquisition is business” by highlighting the need for public procurement professionals to understand the commercial marketplace, become customer focused, maintain relationships with customers, and expand their ability to draft requirements or manage programs. The notion of the public

procurement practitioner as “business advisor” or “business manager” has grown considerably in recent years. To a certain extent, public procurement is now facing up to the challenge posed to private sector procurement by Kraljic^[7] much earlier when he argued for a greater role of procurement in supply chain management. As a related development, federal legislation has promoted both the acquisition of commercial items and the adoption of procurement practices more similar to those used by private sector businesses (FAR Part 12). The US federal government may now use simplified procedures for buying commercial supplies and services up to \$5 million (increased from \$100,000 in 1994), and buyers are encouraged to look to the private sector when developing terms and conditions for federal contracts, rather than using government-unique specifications.

Transactions, Systems, and Process-Based Approaches

Public procurement has been defined in terms of its means of handling specific purchase transactions or “making the buy.” The goal in recent years has been to move beyond viewing this transaction processing as a clerical function, but instead to view it in terms of the efficiencies that public procurement can offer. Especially in the area of procurement of supplies, public procurement has the potential to find quantity discounts and to provide data and analysis of spending patterns to identify opportunities for improvement.

In other respects, public procurement has been regarded as a process-based approach to government. The amount of law and regulation governing how public procurement is conducted is staggering. For federal agencies alone, the FAR prescribes in great detail the procedural rules for how supplies and services must be bought. In fact, the overabundance of rules has caused some to draw attention to the need to focus on results rather than simply the process of contracting.

Another view of public procurement is to consider it as a “nested structure of systems within systems,”^[8] such as the rulemaking system for public procurement at the federal level as typified by the voluminous FAR rules, the procurement cycle for how buys are made, and the organizational structure involved (requiring office, procurement office, receiving/inspection office, etc.).

PROFESSIONALISM

Professional Associations

The dominant professional associations for buyers who work for government organizations are the

National Institute of Governmental Purchasing (NIGP), the National Contract Management Association (NCMA), and the National Association of State Procurement Officials (NASPO). All members of these organizations arguably do the same kind of work, despite the differences in terminology or definition. Periodicals issued by various organizations involved in the field likewise reflect the diversity of terms for the buying done by governmental organizations. The American Bar Association issues *Public Contract Law Journal*, NCMA prints *Contract Management*, the Department of Defense has published *Acquisition Review Quarterly*, a popular web site used by practitioners is called “Where in Federal Contracting?” (www.wifcon.com), and there is a *Journal of Public Procurement* published by Florida Atlantic University.

Lately, professional organizations for public procurement practitioners have sought to expand the reach and influence of the discipline. For example, NCMA^[9] offers the following definition:

“Contract management is a strategic management discipline employed by both buyers and sellers whose objectives are to manage customer and supplier expectations and relationships, control risk and cost, and contribute to organizational profitability/success.”

Similarly, NASPO^[10] has broadened its view of the role played by public procurement, noting: “The emphasis is shifting rapidly from just a “buy transaction” to supply chain management, which includes the requirement definition and disciplines such as supplier development and global sourcing.” This is part of a transition “from process-based to knowledge/accountability-based procurement organizations.”

Public Procurement as a Profession

The NIGP and NCMA refer to a body of knowledge that those working in public procurement should possess. Federal Acquisition Institute has developed “Contract Specialist Training Blueprints” to define, in terms of daily work, what public procurement professionals do. There are 71 different critical tasks and duties performed in public procurement, grouped into three phases: acquisition planning, contract formation, and contract administration. Examples of some of the 71 common tasks or duties are market research, negotiation strategy, and contract modification. Similarly, NCMA has developed a “Contract Management Body of Knowledge” used for its certification program for members who pass a rigorous examination leading to professional designation. National Institute of Governmental Purchasing also has two different certification programs intended to

“indicate to the public that, having mastered a body of knowledge, one can make sound decisions that reflect maximum value for the taxpayer’s dollar.”^[11]

CONCLUSIONS

Overall, public procurement is one of the most significant areas of governmental activity in terms of money. It is also one of the most highly regulated fields of government endeavor. Although there is no single definition of public procurement, and commentators are fond of saying what public procurement should be, the themes of working with customers to define requirements, publicizing upcoming procurements, soliciting and awarding contracts, and administering those instruments are common to all forms of public procurement. In the future, an expansion of the role of public procurement is likely as its strategic role is increasingly recognized.

REFERENCES

1. American Bar Association. In *The Model Procurement Code for State and Local Governments*; Section 1–301; Washington, DC, 2000.
2. Thai, K.V. Public procurement re-examined. *J. Public Procurement* **2001**, 1 (1), 9–50.
3. Miami-Dade County. In *Living Wage Ordinance for County Service Contracts and County Employees*; Section 2–8.9 of the Code of Miami-Dade County; Florida, 2000.
4. Federal Acquisition Institute, U.S. General Services Administration. In *Contract Specialist Training Blueprints*; 2003, <http://www.fai.gov/prodev/contract.htm>.
5. Johnson, P.F.; Michiel, R.L.; Clifford McCue. A comparison of purchasing’s organizational roles and responsibilities in the public and private sector. *J. Public Procurement* **2003**, 3 (1), 57–74.
6. Rindner, C.M. Acquisition as business. *Contract Management* **2001**, 41 (5), 22–26.
7. Kraljic, P. Purchasing must become supply management. *Harvard Business Rev.* **1983**, 109–117.
8. Thai Purchasing must become supply management. *Harvard Business Rev.* **1983**, September–October, 40.
9. National Contract Management Association. In *About NCMA*; 2004, <http://www.ncmahq.org/membership/about.asp>.
10. National Association of State Procurement Officials. In *State Procurement: Strategic Positioning for the 21st Century*; White Paper, January 1999; <http://www.naspo.org/whitepapers/whitepaper-21century.cfm>.
11. National Institute of Governmental Purchasing. In *Why Certify?* 2004; <http://www.nigp.org/certification/WhyCert.pdf>.

Public Procurement Ethics

David Seth Jones

Department of Public Policy, University of Brunei Darussalam, Gadong, Brunei

INTRODUCTION

One of the hallmarks of good governance in the modern state is adherence to strict ethical standards by government departments and agencies in procuring goods and services for public programs (these organizations are commonly referred to as government procurement entities or GPEs). Such standards are necessary for several reasons. The procurement of goods and services by public authorities often comprises a significant percentage of both the operating and capital budgets of a government and involves spending a substantial part of the revenue provided by the taxpayer. In addition, how public procurement is undertaken has a crucial impact upon the efficiency and effectiveness of public services and may also affect growth and employment within the wider domestic economy. For ethical principles to be upheld, there should exist a clear framework of rules and procedures based upon them, governing the various stages of the procurement process: devising specifications, advertising the intended procurement, evaluation of tender and quotation submissions, recommendation of a submission, final selection, drafting the contract, and post-contract adjustment.^[1] Where they exist, such rules and procedures are usually enshrined in legislation and/or stipulated in administrative directives, often contained in public service operational manuals, with which procurement officers and other public officials as well as suppliers must comply.^[2,3] Explained below are the key principles upon which such a framework of procurement should be based.

FAIRNESS

Principles of Fairness

An important part of the ethical responsibility of government and GPEs when purchasing goods and services is to follow fair practices. Fairness can be construed in various senses: rewarding desert, providing for need, avoiding harm to others, and honoring a promise. In the first sense, it entails providing rewards according to merit, achievement, or endeavor. In the second sense, it involves responding to the needs

arising from circumstances or inborn disadvantages, regardless of merit, achievement, or endeavor.^[4,5]

These two interpretations of fairness have given rise to two different approaches to public procurement. The third interpretation of fairness prohibits actions which harm the legitimate interests of others, while the fourth sense of fairness, based on the duty to keep a promise, requires that the terms of a contract, freely entered into, be honored by the parties concerned.^[6]

Desert Principle: Equal Access

Fairness related to desert requires that all potential suppliers be given equal access to the public procurement market through open competition, and that contracts be awarded to those who offer the lowest price or best meet the stipulated specifications.^[2,7] To ensure equal access and a level-playing field, governments and GPEs must disseminate information about an intended procurement as widely as possible through the relevant advertising channels: the government gazette, the local and overseas newspapers, the trade journals, and, increasingly, the internet.^[8,9] Equal access also necessitates, for the benefit of overseas companies and businesses, that a reasonable time be allowed for both the receipt of tender and quotation forms and other detailed documentation and also the submission of tender and quotation proposals. Once tenders and quotations have been submitted, the GPE is obliged to objectively evaluate the merits of each in terms of price and specification fulfillment to determine which proposal is the most advantageous.^[7-9]

After the contract has been awarded, fairness based on desert prescribes that unsuccessful tenderers be informed of the reasons why they have failed. This would be helpful in enabling them to submit better proposals for future tenders and quotations. In addition, in the interest of fairness, any unsuccessful tenderer who has grounds to believe that his/her submission was not evaluated on the basis of merit or that procurement officers did not follow official procurement procedures should be given the right of appeal to a procurement adjudication body.^[3,8]

Furthermore, postcontract adjustment should be carefully limited.^[3] For contracts that extend over a lengthy period, some adjustments may be necessary

as a result of changing rates of inflation, currency fluctuations (for overseas suppliers), newly improved product designs, and changing needs of the GPE itself. The contract alterations may be negotiated with the contracted supplier and should only be allowed under conditions stipulated in the contract or with the agreement of both parties. However, where the postcontract adjustments are either unrelated to the above and/or are substantial, then it is only fair to other potential suppliers that a new tender or quotation be called for.^[3]

Open competition based on equal access has been enshrined as a central principle in the WTO's Government Procurement Agreement (GPA), which was drafted in 1994. More than 30 countries are now signatories to the agreement. The key objective is to liberalize the public procurement market in as many countries as possible and so enable foreign suppliers or foreign-linked suppliers to compete for government contracts on equal terms with local suppliers. Governments which acceded to the GPA cannot, under its provisions, discriminate against foreign suppliers, even those that do not include any domestic content in the their goods and services, nor can they discriminate against local suppliers with foreign affiliation or ownership and/or trading in goods and services originating in another country.^[8] Goods and services of a security or strategic interest may be exempted (e.g., construction of a military building). In this respect, the GPA has followed the system of competitive procurement adopted in the European Union based on EU Directives following the creation of the European Single Market in 1987.^[7,10,11]

Need Principle: Unequal Access

Alternatively, public procurement practices may be considered as fair if suppliers are chosen on the basis of their needs (or the needs of the population or section of the population to which they belong). Such suppliers are then given preferential treatment with contracts drafted which take into account those needs. Included in this category may be local suppliers within the country, small-scale suppliers, and those drawn from disadvantaged minority communities, most of whom could not compete effectively for government procurement contracts if equal access through open competition was permitted.^[7,12] Affirmative access, it is argued, is necessary to enable such businesses to survive and expand and to generate income for owners and employees alike. For suppliers from a minority community, affirmative access may be important in raising living standards and creating business opportunities as part of a wider policy to improve its circumstances. To facilitate affirmative access, selective

tendering may be adopted which restricts tender submissions to suppliers that fall within the categories stated above. To the same end, open tendering may still be adopted, but with special consideration given to submissions from the favored categories during the evaluation stage.^[2,12] Such special consideration may entail allowing a preferential percentage margin on the prices offered by the suppliers concerned.

Avoidance of Harm Principle

Fairness also imposes a responsibility upon GPEs and suppliers to ensure that goods and services are produced or supplied by means that do not lead to harmful consequences for others (negative externalities), especially in terms of health and safety. Examples of such externalities are the production of goods in which little care is taken to protect the safety of employees or which results in environmental waste or pollution. It is then justified for GPEs to restrict access to the public procurement market for businesses engaged in such practices. For this reason, certain governments, as a matter of principle, discriminate against businesses with poor industrial safety records or which despoil the environment or do not supply environmentally friendly products.^[13,14]

Principle of Contractual Obligation

It is only fair, in accordance with the moral duty to keep a promise, that once a GPE and a supplier have signed a procurement contract, both then honor its terms. During the period of the contract, the GPE might want to alter the terms of the procurement, as a result of its changing needs, as mentioned above. This might be allowed under certain conditions and within certain limits, which should be stipulated in the contract. However, fairness dictates that, having entered the contract, the GPE should secure the agreement of the supplier to an alteration of the terms beyond what is allowable and should not terminate the contract if this is not forthcoming. Likewise, the supplier is duty-bound to honor his/her obligations under the terms of the contract, and any adjustment it wishes to make must have the consent of the GPE.^[9] The key obligations for suppliers that relate to honoring a contract include not committing a novation of contract, refraining from unauthorized subcontracting, meeting the specifications in the type and quality of goods and services supplied, adhering to the price agreed, and abiding by the stipulated time frame for delivery.^[9,15] Where the supplier is in serious default of the above, the GPE has the right to impose penalties, e.g., fines, termination of contract, and debarment from future contracts.

VALUE-FOR-MONEY

As major spenders of tax revenue, GPEs are under an obligation to provide value-for-money for the public as both taxpayers and recipients of public services. This reflects the “social contract” at the heart of a democratic state, by which voters/taxpayers are entitled to receive from an elected government efficient and effective public services in return for the income and wealth they have foregone through taxation.^[7,16]

Value-for-money is achieved through the procurement of goods and services for a public program, appropriate to its specific requirements, which are of a sufficiently high standard and purchased at a reasonably low price. If such criteria are not adhered to, the taxpayers’s money will be wastefully spent and public managers will be denied the means of providing a high standard of public services. Ensuring that a procurement meets the requirements of a public program may necessitate decentralizing the task of procurement to the implementing department or agency. However, where materials are commonly used in a standard form by all departments and agencies, then value-for-money can best be guaranteed by bulk purchasing by a centralized body.^[1]

To ensure value-for-money, certain procedures must be established and observed. These include justifying the need for the procurement and devising procurement specifications in light of program requirements. In addition, tender proposals should be thoroughly vetted to determine which is the most advantageous, and contracts should be properly and precisely drafted. It is also important that procurement decisions are taken by committees and bodies that include not only procurement officers, but also technical specialists with the expertise to understand technical specifications and the end users who can determine if the products and services to be procured are appropriate to their programs.^[7]

In most cases, value-for-money is most likely to be achieved through open competitive tendering. However, sometimes, even selective tendering and limited sourcing can result in equally positive outcomes. This depends upon negotiation and hard bargaining by procurement officers on matters of price, design, scope, and quality of the product or service to be purchased.^[7] In such cases, procurement officers will be aided by having access to full information of the prospective supplier’s cost breakdown, expected profit margin, and supply capacity.^[17–19] Limited sourcing may involve the use of just one supplier for repeat procurements of the same goods or services. This is justified if the supplier is reliable and, on the basis of a guarantee of repeat orders, is prepared to offer discounts and improved products and services, resulting in better value-for-money.^[11,19]

PROBITY

A further ethical obligation of governments and GPEs is to maintain strict standards of probity and avoid corruption and dishonesty in the procurement of goods and services. All too often, public procurement is undermined by corrupt practices in which political leaders, senior administrators, procurement officers, and other officials take advantage of the opportunities provided to make illicit personal gains or secure benefits for their family and friends. All the main forms of corruption can occur. One is misappropriation in which suppliers are overinvoiced and provide more goods than needed or specified, with the surplus then falling into the hands of public officials and used for their benefit or sold for personal gain. Another prevalent form of corruption is bribery. Money and favors (kickbacks) are offered to, or demanded by, officials in return for the award of a contract.^[2,9,15] Equally common is nepotism and cronyism by which preferential treatment is accorded to businesses linked to the families and friends of government officials who influence or decide the choice of tenderer.^[20] For their part, suppliers may act unethically not only by offering bribes, but also by committing fraud, such as by deliberately disclosing false information about their past financial record, resources and capital value, and their capacity to deliver the goods and services in question.^[9]

To ensure high standards of probity, measures may be enforced to both prevent and deter corruption and dishonesty. Preventive measures may include a requirement that officers who are involved in any significant stage in the procurement process declare any possible conflict of interest. This especially applies to a personal or family connection with a business (or any affiliated business) submitting a quotation or tender, through ownership, partnership, shareholding, and loans.^[9,20] As further measures to prevent corruption, the different functions and stages of the procurement process should involve separate officers (including separate officers for making procurement policy and undertaking procurement operations), and if manpower availability permits, they should be rotated on a regular basis.^[2] In addition, more than one officer should be involved in the key stages of the procurement process.

Preventive measures should be reinforced by deterrence measures to punish those discovered to be engaged in corruption and dishonesty. Penalties would include stiff fines; in more serious cases, prison terms; in addition, for suppliers, debarment from future contracts; and for procurement officers, demotion or dismissal from post. In some countries such as Hong Kong and Singapore, legislation has been enacted under which an anticorruption agency has been created with wide-ranging powers to pursue inquiries and unearth any evidence of corrupt practices, with courts

given powers to pass stiff sentences in such cases.^[21] In many countries, the government audit agency is given the investigative powers and the responsibility to identify corruption in the procurement process.

TRANSPARENCY

Maintaining fairness, value-for-money, and probity in part depends upon making the public procurement process as transparent as possible. This requires the disclosure of all relevant information at critical points in the process.^[3] At the outset, information relating to an intended procurement, including tender or quotation specifications, should be made available to all prospective suppliers, as mentioned above. The various tender or quotation proposals submitted should then be publicized, followed by the disclosure of the name of the successful bidder, together with the content of the contract and the cost of the purchase.^[3] Transparency also requires that reasons be given to the unsuccessful tenderers why they had not been selected, as stated above. In addition, to ensure proper accountability in the procurement process, government auditors, internal regulatory authorities, and relevant watchdog committees within the legislature should be given details of how a procurement was undertaken, together with all the necessary documentation, if so requested. Where necessary, senior administrators and procurement officers of a GPE should be subject to cross-examination by relevant committees of the legislature.

Transparency underpins the three other ethical principles in public procurement in two senses. First, it is an essential condition in enabling open competition for tenders to prevail which is both the basis of fairness related to desert and arguably a guarantee of value-for-money. Second, it subjects GPEs to greater scrutiny and accountability in respect to various aspects of fairness, value-for-money, and probity. Any deviation from such standards can be readily identified and exposed, allowing action to be taken to rectify the shortcoming.

CONCLUSION

The above principles governing procurement in many respects supplement each other. For example, measures that promote fairness based on open competition necessitate transparency, arguably ensure increased value-for-money, and even safeguard standards of probity. Value-for-money is more likely to be achieved with greater transparency that allows procurements to be open to scrutiny leading to greater political and public accountability. Value-for-money can be further enhanced if procurement practices are governed by

probity, in view of the fact that corruption can lead to a lower standard of procurement and greater waste and inefficiency.

However, in one respect, one of the principles of procurement may be at variance with the others. While much may be said in favor of fairness based on the need to allow unequal access to the public procurement market, this may lessen the possibility of achieving value-for-money because the suppliers chosen for a contract through special preference may not offer the most advantageous terms. Equally, it may be at variance with fairness based on open competition and merit because suppliers offering better terms and more capable of meeting the specification requirements may be rejected. This requires the government or GPE to determine where the balance of ethical responsibility lies between need and value-for-money. If consideration of need-based fairness in the procurement process does not result in the acquisition of substandard goods and services at an excessive price to the taxpayer, then preferential access may be allowable. However, if such access does indeed lead to procurements that fall well short of giving value-for-money, so impairing public programs and excessively draining tax revenue, then the alternative of selecting the best suppliers through open competition within a level-playing field may be the better option.

REFERENCES

1. Coe, C.K. *Public Financial Management*; Prentice Hall: Englewood Cliffs, NJ, 1989.
2. Coe, C.K. Government purchasing: the state of the practice. In *Handbook of Comparative Public Budgeting and Financial Management*; Lynch, T., Martin, L., Eds.; Marcel Dekker: New York, 1993; 207–224.
3. Sherman, S. *Government Procurement Management*; Wordcrafters Publications: Germantown, MD, 1999.
4. Benn, S.; Peters, R. *Social Principles of the Democratic State*; George Allen & Unwin: London, 1961.
5. Barry, B. *Political Argument*; Routledge & Kegan Paul: London, 1965.
6. Thiroux, J. *Ethics: Theory and Practice*, 7th Ed.; Prentice Hall: Upper Saddle River, NJ, 2001.
7. Erridge, A. Competitive tendering and partnership in the public sector. In *Working in Partnership*; Burnes, B., Dale, B., Eds.; Gower: Aldershot, UK, 1998; 21–38.
8. WTO. *Plurilateral Agreement on Government Procurement*; Author: Geneva, 1997.
9. Farrington, B.; Waters, D. *Managing Purchasing: Organizing, Planning and Control*; Chapman & Hall: London, 1994.
10. Arrowsmith, S.; Linarelli, J.; Wallace, D. *Regulating Public Procurement*; Kluwer Law International: The Hague, 2000.
11. McDonald, F.; Winkelmann, T. Public procurement rules in the european union and the prospects for

- efficient supply: the cases of UK and germany. In *Innovations in Procurement Management*; Cox, A., Ed.; Earlsgate Press: Boston, UK, 1996; 283–304.
12. Leenders, M.R.; Fearon, H.E. *Purchasing and Materials Management*, 11th Ed.; Irwin: Homewood, IL, 1996.
 13. Russel, T. Introduction. In *Greener Purchasing: Opportunities and Innovations*; Russel, T., Ed.; Greenleaf Publishing: Sheffield, UK, 1998; 9–21.
 14. Pento, T. Implementation of public green procurement programmes. In *Greener Purchasing: Opportunities and Innovations*; Russel, T., Ed.; Greenleaf Publishing: Sheffield, UK, 1998; 22–30.
 15. Finkler, S. *Financial Management for Public, Health, and Not-for-Profit Organizations*; Prentice Hall: Upper Saddle River, NJ, 2001.
 16. Central Unit On Procurement. *Supplement to CUP Newsletter*; HMSO: London, 1989.
 17. Fishner, S. *A Report on Government Procurement Practices*; Camelot Publishers: Merrifield, VA, 1989.
 18. Baily, P. Procurement policy. In *Handbook of Management*, 3rd Ed.; Lock, D., Ed.; Gower: Brookfield, VT, 1992; 339–406.
 19. Deverill, N. Change and innovation in government procurement. In *Innovations in Procurement Management*; Cox, A., Ed.; Earlsgate Press: Boston, 1996; 239–271.
 20. Zenz, G. *Purchasing and the Management of Materials*, 7th Ed.; John Wiley: New York, 1994.
 21. Quah, J. Controlling corruption in city states: a comparative study of Hong Kong and Singapore. *Crime, Law Soc. Chang.* **1995**, 22, 391–414.

Public Reporting

Mordecai Lee

Department of Governmental Affairs, University of Wisconsin-Milwaukee,
Milwaukee, Wisconsin, U.S.A.

INTRODUCTION

Public reporting refers to the actions of a government agency to account to the citizenry for its activities in the recent past, including its record of performance and its stewardship of public funds. The term refers both to the normative obligation of public entities in a democracy to perform such activities and to the internal management activities necessary to accomplish this external communications goal.

Public reporting was developed by public administration theorists in the first half of the 20th century in an effort to harmonize the emergence of the administrative state with democracy. Inasmuch as permanent and professionalized bureaucracies were not subject to elections, there was a need to define the role of public administration in democracy. By assigning to senior civil servants the duty to report directly to the citizenry on agency activities, theorists sought to make government departments accountable to the public-at-large, not just to elected officials. Public reporting would contribute to an informed citizenry, the ultimate source of power and decision making in democracy.

ORIGINS

With the emergence of the administrative state in advanced western societies in the late 19th and early 20th centuries, a natural concern arose regarding the need to harmonize democracy with the inherently undemocratic nature of bureaucracy. Political, legal, and constitutional responses were the understandable initial solutions. Administrative agencies would be formally accountable to the elected and appointed oversight institutions of democracy, such as through budget and policy controls possessed by the chief elected executive officer, oversight activities by legislative bodies, and the jurisdiction of the courts to rule on agency policies and procedures. These elected and appointed constitutional branches of government would be able to assure that democracy would reach into bureaucracy through external controls and interventions. Yet these approaches all assumed that

bureaucracies were only *indirectly* accountable to the citizenry—the source of democratic sovereignty and power.

However, another line of thinking suggested that bureaucracies also needed to be *directly* accountable to the voters. It was not enough, this theorizing went, to be accountable to the citizenry through other democratic institutions. Instead, agencies should also be directly responsible to the electorate. Thus was born public reporting.

As it gradually developed in the 1920s and 1930s in the United States, public reporting referred to the duty of the public administrator to report regularly to the public-at-large on the activities of the agency. Inasmuch as an informed public was the sine qua non of democracy,^[1] then government agencies should be one of the sources of information for the citizenry about the operations of the public sector. For public opinion to be the central engine of decision making in a democracy, the voters needed to have basic information upon which to construct their opinions. In the administrative state, information should come not only from traditional sources, such as newspapers, civic groups, and elected officials, but also from the newest component of the public sector, government departments. That meant government managers had the obligation of preparing and releasing reports on a regular schedule, helping to assure that the public would become familiar with agency activities.^[2–4]

Public reporting first emerged as part of the efforts by reformers to “clean up” American city governments in the early 20th century. A knowledgeable citizenry would become a bulwark against patronage, waste, and corruption in city halls. One of the first references to public reporting was in 1912, when Bruère^[5] argued that the concept of “efficient citizenship” required regular reporting by city government to the public. In the 1930s, the International City Management Association sponsored contests to promote improved public reporting^[6] and through the 1950s continued actively encouraging municipalities to conduct robust reporting programs.^[7]

The public reporting imperative of the government manager is one of the activities that operationalizes the difference between business and public administration.

Given the public-sector environment it operates in, government management is qualitatively a different activity from business administration. The duty of public reporting is one of the components of public administration that reflects this inherent intersectorial difference.^[8]

DEFINITION

Public reporting is only one element of the array of interactions that government agencies have with the public-at-large. In that respect, it needs to be delineated in a way that clearly separates it from other related, but distinctly different, activities.

In the context of an agency's external communication programs, public reporting is an activity that is not related to the efficient implementation of the agency's core mission. Many public relations programs are aimed at furthering the *raison d'être* of an agency, such as informing the public of new programs and services, using publicity as a substitute for regulation and encouraging the public to serve as the "eyes and ears" of the agency.^[9] In that respect, public reporting is a communication activity that does not "do" anything in context of the pragmatic get-the-job-done pressures of daily management and communications. Rather, reporting simply provides information for information sake.

Similarly, public reporting focuses on efforts to contribute to an informed public. In that sense, reporting contributes to the functioning of democracy. However, public reporting focuses only on the transmission of information from the administrative agency to the public and not on the subsequent developments that it triggers. Thus public reporting is different from citizen participation in agency decision making or agency "listening" to customers and clients in an effort to improve the quality of consumer interactions.

Given that the modern definition of communication is a two-way flow of information, then public reporting focuses solely on the first half and front end of the entire communications loop and only on communication intended to inform for democratic purposes. Therefore public reporting is defined as:

the management activity intended to convey systematically and regularly information about government operations, in order to promote an informed citizenry in a democracy and accountability to public opinion. It consists of direct and indirect reporting of the government's record of accomplishments and stewardship of the taxpayers' money. Public reporting is presented in many different communication formats, but always uses vocabulary that is understandable and meaningful to lay citizens.^[10]

Public reporting occurs when an agency conveys information about its performance to the citizenry rather than to political and elected oversight institutions. Certainly, public opinion gains information about the operations of the government from elected officials, whether those from the executive or the legislative branches. However, the central concept of reporting focuses on government agencies communicating with the public *besides* the formal, legal, and constitutional channels they are subject to.

Generally, a bureaucracy can report to the public in two ways: directly and indirectly. Indirect reporting refers to communication efforts *through* the intermediary institution of the news media. Direct reporting focuses on non-mediated communications from the agency to the public-at-large for the purpose of conveying summary and performance information.

INDIRECT REPORTING

One way for an agency to inform the public is by facilitating media coverage of its operations. As an instrument of democracy, the news media can provide the public with ongoing information about the activities and performance of the government. Given this constitutional role of the press in contributing to an informed citizenry, agencies have a counterpart obligation to assist the press in its efforts to report about agency operations. This perspective is the underlying rationale for public-sector agencies to maintain media liaison offices, staffed with specialists trained in dealing with reporters.

Some news coverage of an agency is headline-driven. In those situations, the agency is passively reacting to inquiries from the press. This form of interaction rarely leads to providing the public with a more systematic and comprehensive understanding of agency performance. Instead, it provides highly selective—and often negative—snapshots driven by the media's need for controversy. Therefore a better opportunity for indirect reporting occurs when an agency is able to initiate and generate coverage. Events such as observing a milestone in a program's operations, the ribbon cutting at a new site, and the kickoff of a new service all provide an opportunity to report indirectly to the public about the general and overall activities of an agency. While, by necessity, an incomplete snapshot of the entire agency's operations, this category of news coverage is nonetheless a channel for reporting to the citizenry on the more routine and daily aspects of government, in contradistinction to news coverage that is crisis- and controversy-driven.

One of the advantages of indirect reporting is that it is free. The agency incurs no additional cost when

the media reports to the public about it. Yet while agencies should make efforts to “tell their story” to the citizenry via the news media, indirect reporting—by itself—is rarely adequate to accomplish the broad democratic goals intended for public reporting.

DIRECT REPORTING

The more prominent and effective approach to public reporting is through direct contact with the citizenry. In the predigital era when public reporting first began, direct reporting often focused on printed materials. Items such as annual and quarterly reports were viewed as one of the basic ways that agencies could inform the public about their accomplishments. Two key elements have been identified as necessary for effective printed reports: contents and distribution.

First, printed reports should be prepared with the audience of the lay public in mind. Reports need to be in plain English and must focus on the key items that represent the agency’s accomplishments. Long and detailed descriptions, use of technical jargon, and impenetrable statistical presentations are examples of what a good report should not contain. Besides use of plain language, effective public reports extensively rely on graphics, photos, and other visual techniques to make the documents user-friendly. Similarly, linking agency information to comparisons that are understandable to average citizens (“for the cost of an average breakfast, our agency delivers to each of our clients . . .”) is another way to contribute to an effective public report. This focus on content and presentation also helps highlight the difference between public reporting and financial reporting. Annual reports prepared by accountants and auditors to satisfy various legal and constitutional requirements rarely qualify as public reporting. The density of numerical information in financial reports generally precludes creating a product that is meaningful to a lay reader.

Second, effective reporting also requires significant efforts to disseminate reports as broadly as possible. Some agencies indeed prepare annual reports, but their dissemination efforts are half-hearted and incomplete. Certainly, copies should be sent to elected officials, key civic leaders, the media, public libraries, and major civic groups. However, the concept of public reporting calls for efforts to put the report directly in the hands of the citizenry. In the middle of the 20th century, municipalities experimented with a wide variety of dissemination channels. Some examples included inserting the annual report in the annual property tax bill or in quarterly bills from municipal utilities, inserting the report in the daily newspaper, having sanitation workers distribute the reports while on their rounds and recruiting boy and girl scout troops to

adopt, as their civic project, the distribution of the report door-to-door.

However, annual reports and other regularly issued publications (such as quarterly ones) are not the only way that government agencies can engage in direct public reporting. Referring again to examples from the mid-20th century, other efforts included open houses, speakers bureaus, billboards, signs on subways and buses, annual exhibits keyed to the budget adoption process, movie shorts, public service advertising, regular radio programs, and film versions of annual reports. Like other marketing and advertising efforts, multiple products and channels need to be utilized to maximize the potential of reaching as much of the citizenry as possible.

ROOSEVELT’S OFFICE OF GOVERNMENT REPORTS

In the federal government, the high water mark of traditional public reporting was the Office of Government Reports (OGR), which President Roosevelt established in 1939. In the Executive Order creating it, one of its central missions was to “distribute information concerning the purposes and activities of executive departments and agencies.”^[11] Roosevelt viewed OGR as the executive branch’s agency that would report to the public on the activities of the federal government. In a national radio address, Roosevelt explained that once informed factually of the record of the Administration, public opinion could then express its will to Congress regarding the continuation or cessation of those federal programs.^[11] As a reflection of the importance he assigned to OGR, Roosevelt made it one of the five original agencies comprising the Executive Office of the President (EOP), which he also established in 1939. However, OGR was a very controversial agency, especially with the conservative coalition in Congress. Overcoming strong resistance on Capitol Hill to the concept of such a federal reporting agency, in 1941, Roosevelt succeeded in shepherding through Congress a law authorizing permanent appropriations for OGR.^[12] However, continued opposition to OGR by legislators forced Roosevelt in 1942 to merge it for the duration of World War II into the temporary Office of War Information (OWI). When President Truman tried to reestablish OGR after the war, as Roosevelt had intended, the Republican 80th Congress refused to appropriate any additional funding for it. OGR went out of existence in 1948.^[12]

The political blowout relating to OGR presented a high-profile case of how easily public reporting can become politically controversial. Whether an accurate assessment or not, the lesson learned was that for public reporting to occur, it needed to be blander and less

engaging. In fact, for the second half of the 20th century, public reporting gradually ossified into boring annual reports at most levels of government. Those half-hearted reports were given limited circulation and triggered even less interest.

21ST CENTURY TRANSFORMATION: E-REPORTING

While the rationale for public reporting retained its validity in the post-OGR era of the second half of the 20th century, as a concrete activity, it gradually faded from prominence in both public administration theory and practice. However, the emergence of the communications age and the digital era has provided an opportunity to revive and reestablish the value of public reporting in the 21st century. New communications technologies now permit government agencies to communicate quickly, efficiently, and inexpensively directly to the citizenry. Annual and special reports posted on agency websites and distributed by e-mail provide an opportunity to perform public reporting in ways that could not have been envisioned a century earlier.

Similarly, the increased attention in public administration to performance measurement has provided a new format by which agencies can report on their operations to the public-at-large. While performance measurement was originally developed largely for management control purposes—whether for the senior management of the agency or for oversight by the executive and by the legislative branch—it has much potential and value to the lay citizenry as well.

Therefore the combination of digital technology with the focus on performance measurement has provided an opportunity to reformulate and reinvigorate the original conception of public reporting. Now called e-reporting, it consists of:

the administrative activity that uses electronic government technology for digital delivery of public reports that are largely based on performance information. E-reporting is a tool of e-democracy that conveys systematically and regularly information about government operations that is valuable to the public-at-large, in order to promote an informed citizenry in a democracy and accountability to public opinion. E-reports are planned to be citizen-friendly, by being understandable and meaningful to the lay public.^[10]

CONCLUSION

Public reporting has been making a comeback in public administration. Herzlinger^[13] has suggested that

one of the ways to restore public trust in government is through improved reporting. That would include an increase in the amount of important information disclosed to the citizenry as well as better and broader dissemination of such regular reports. The Organisation for Economic Cooperation and Development (OECD), an association of the most developed western countries, issued two reports in 2001 calling for improved public reporting and providing guidelines for such efforts.^[14,15]

Whether widely or well practiced, public reporting is a permanent and inherent element of public administration in a democracy. Civil servants are expected to engage in systematic efforts to report to the citizenry on their agencies' activities, performance, and stewardship of taxpayer funds. The profession of public management entails more than effectively running an agency. It also involves contributing to the working of democracy. Public reporting is the vehicle for such efforts. For more information, see Refs.^[16–20]

REFERENCES

1. Brown, R.D. *The Strength of a People: The Idea of an Informed Citizenry in America, 1650–1870*; University of North Carolina Press: Chapel Hill, NC, 1996.
2. Beyle, H.C. *Governmental Reporting in Chicago*; University of Chicago Press: Chicago, 1928.
3. National committee on municipal reporting. *Public Reporting, With Special Reference to Annual, Departmental, and Current Reports of Municipalities*; Municipal Administration Service: New York, 1931.
4. Ridley, C.E.; Simon, H.A. *Specifications for the Annual Municipal Report*; International City Managers' Association: Chicago, 1948.
5. Bruère, H. Efficiency in city government. *Ann. Am. Acad. Polit. Soc. Sci.* **1912**, *41*, 3–22.
6. Simon, H.A. Inter-City contests. In *Municipal Year Book, 1937*; Ridley, C.E., Nolting, O.F., Eds.; International City Managers' Association: Chicago, 1937; 137–144.
7. Snyder, R.W. Municipal reporting in 1953. In *Municipal Year Book, 1954*; Ridley, C.E., Nolting, O.F., Eds.; International City Managers' Association: Chicago, 1954; 269–276.
8. Lee, M. Intersectoral differences in public affairs: the duty of public reporting in public administration. *J. Public Aff.* **2002**, *2* (2), 33–43.
9. Lee, M. Public information in government organizations: a review and curriculum outline of external relations in public administration. *Public Admin. Manage.* **2000**, *5* (4), 183–214. http://www.pamij.com/00_5_4.html (accessed February 2004).
10. Lee, M. *E-reporting: Strengthening Democratic Accountability*; IBM Center for the Business of Government: Washington, DC, 2004, pp. 5, 9, also available online: http://www.businessofgovernment.org/pdfs/Lee_Report.pdf (accessed February 2004). Reprinted as Chapter 4 in *Managing for Results 2005*; Kamensky, J.M.,

- Morales, A., Eds.; Rowman & Littlefield: Lanham, MD, 2004; 141–195.
11. Roosevelt, F.D. *The Public Papers and Addresses of Franklin D. Roosevelt*; Macmillan: New York, 1941; Vol. 1939, pp. 305–310, 494.
 12. Lee, M. *The First Presidential Communications Agency: FDR's Office of Government Reports*; State University of New York Press: Albany, NY, 2005.
 13. Herzlinger, R.E. Can public trust in nonprofits and government be restored? *Harvard Bus. Rev.* **1996**, *74* (2), 97–107.
 14. Caddy, J.; Vergez, C. *Citizens as Partners: Information, Consultation and Public Participation in Policy-Making*; Organisation for Economic Co-operation and Development: Paris, France, 2001.
 15. Gramberger, M.R. *Citizens as Partners: OECD Handbook on Information, Consultation and Public Participation in Policy-Making*; Organisation for Economic Co-operation and Development: Paris, France, 2001.
 16. Campbell, O.W. San Diego's 1951 annual report. *Public Admin. Rev.* **1953**, *13* (1), 30–32.
 17. Graves, W.B. Public reporting in the American states. *Public Opin. Q.* **1938**, *2* (2), 211–228.
 18. Lee, M. Is there anything new under the sun? Herbert Simon's contributions in the 1930s to performance measurement and public reporting of performance results. *Public Voices* **2003**, *6* (2–3), 73–82.
 19. Scott, J.D. Local Government Publications. In *Effective Communication: A Local Government Guide*; Wheeler, K.M., Ed.; International City Management Association: Washington, DC, 1994; 190–216.
 20. Wall, N.L. *Municipal Reporting to the Public*; International City Managers' Association: Chicago, 1963.

Public Sector Productivity: Barriers

David N. Ammons

*Institute of Government, University of North Carolina at Chapel Hill,
Chapel Hill, North Carolina, U.S.A.*

INTRODUCTION

Successful managers of public sector organizations find ways to cope with a host of obstacles that jeopardize the efficiency and effectiveness of their operations. Ideally, they devise strategies to circumvent or defeat these barriers to public sector productivity.

Some impediments to organizational performance are common across sectors, confronting public and private sector managers alike, but other barriers faced by public sector managers are distinctive. Common barriers to productivity improvement in the public sector are sorted into the three clusters noted below and described briefly:^a

- Environmental barriers—including factors that most clearly distinguish the public sector environment from that of the private sector and confound productivity initiatives.
- Organizational barriers—including common characteristics found at varying degrees in public sector organizations.
- Personal barriers—including individual traits, attitudes, and behavior that may influence an administrator's inclination and ability to tackle productivity improvement opportunities.

ENVIRONMENTAL BARRIERS

Although many candidates for public office pledge to run government like a business, once elected most officeholders discover that the public sector environment is, in many ways, less conducive to productivity improvement than is the private sector environment. Among the key environmental barriers are the following:

Absence of Market Pressures: Except in cases of “managed competition” or other relatively rare circumstances that introduce the pressure of market

competition, public sector agencies tend to behave as “unregulated monopolies.”^[1–6] If a public sector agency is to be aggressive, responsive, innovative, and cost conscious, its inspiration typically must come from something other than a competitive market environment.

Political Factors that Influence Decision Making: Unlike the private sector where the financial “bottom line” guides corporate decisions, the public sector is influenced by multiple objectives that often conflict with one another. Political biases, personal whims, and the legislative process often can trump a seemingly more rational and productive strategy.

Public's Lack of Patience with Operational Changes and Lack of Enthusiasm for Gradual Gains: The gains from productivity initiatives usually take time to materialize, but the disruptions to the operational status quo imposed by change often are immediate. Unless quick and substantial improvements are made in the quality or quantity of a given service, affected clientele are apt to rail against productivity-inspired program change, whereas even the cost-conscious public at large, unimpressed by cost-saving benefits of perhaps only pennies a day for individual taxpayers, are unlikely to rally in defense of productivity improvement.

Dominant Preference for the Status Quo: If groups having the power to influence change are generally satisfied with the services provided to their members and see little reason to push changes for the benefit of powerless groups, little encouragement is likely to be generated for experimentation and innovation in service delivery.^[7]

Lack of Political Appeal: As a political cause, improving the quality and efficiency of government services lacks the appeal of promises to resolve social issues, the allure of program expansion or capital improvements, the grand implications of long-range planning and economic development strategy, or the widespread personal implications of tax policy. Policy issues capture the morning headlines and the evening news much more often than do stories of operational successes and failures in routine governmental functions.

Short Time Horizon: Many productivity improvement initiatives are long-term strategies that carry large front-end price tags and the promise of only

^aFor a more extensive discussion of productivity barriers, see Ammons, D.N. Productivity barriers in the public sector. In *Public Productivity Handbook*, 2nd Ed.; Holzer, M., Lee, S.-H., Eds.; Marcel Dekker: New York, 2004; 139–163.

gradual returns on investment. Public officials frequently have a short time horizon for such matters—one that extends only as far as the next election. After all, why incur major expenses and perhaps short-term tax increases that are almost sure to earn the wrath of voters, while producing long-term benefits that may only be enjoyed by one's successor?

Subordination of Efficiency to Secondary Status: In public sector deliberations that pit the values of efficiency, effectiveness, responsiveness, and equity against one another, efficiency frequently loses. Professional associations, employee organizations, client groups, neighborhood associations, and service regulators all constitute formidable interest groups that tend to place efficiency on a level secondary to other goals, if they consider efficiency at all. Not surprisingly, the values that compete with efficiency are often given political and administrative priority.

Other Barriers Common to the Public Sector Environment: Other barriers that distinguish the public sector from the private sector environment include the following:

- Limited options for achieving economies of scale (i.e., options for market expansion typically are more constrained than in the private sector).
- More cumbersome procurement and personnel procedures.
- Legal restrictions and court rulings that can dramatically affect the nature of public services, how these services are delivered, and how they are funded.
- Anti-innovation and antiproduktivity effects of some grants and other intergovernmental programs, especially those that impose rigid prescriptions, limit local discretion, or reward increased local expenditures.
- Intergovernmental mandating of state and local expenditures.

ORGANIZATIONAL BARRIERS

Although less pervasive than the environmental barriers that tend to permeate the public sector, a second category of productivity barriers includes impediments that are sufficiently common in public sector organizations to warrant mention here. These organizational barriers differ widely in their incidence and severity. Each may be a major impediment in some agencies or jurisdictions and relatively insignificant in others.

Bureaucratic Socialization Process: When veteran employees introduce newcomers to the ways of the organization, instructing them on what is expected and "showing them the ropes," the result can be

positive if the culture encourages service and innovation, or it can be extremely damaging if the organization's veterans have lost their zeal for excellence. An organization facing the latter circumstance may find itself staffed with a perpetual cadre of bureaucrats resistant to innovation.

Lack of Accountability: If it is true, as critics claim, that few government managers or employees are held accountable for their actions or for their failures to act, then government operations are being deprived of a fundamental mechanism of direction and control. Without the pressure that comes with accounting for resources and results, many managers are likely to be satisfied with the status quo. Why strive for productivity improvement?

Focus on Outputs Rather than Outcomes: Focusing on their department's activities or workload—its outputs—is far easier for most government managers than trying to track progress on the bigger picture of whether their efforts are achieving positive results. However, keeping an eye on results is what leads top-performing organizations to excellence. Their lower-performing counterparts are unlikely to discover opportunities for productivity gains by focusing simply on maintaining or expanding the activities of their departments.

Perverse Reward System: Too often, governments fail to recognize, promote, or adequately reward managers who find ways to do more with less. Instead, they typically reward managers who have expanded their budgets and increased the number of persons they supervise. Such a reward system discourages the pursuit of greater efficiency.

"Spend it or Lose it" Budget Rules: Because few governments allow their departments or agencies to carry forward to subsequent years all or even a portion of any savings remaining in the current year's budget, many operating officials believe that it is more rational to spend the money than to lose it to the central treasury. Besides, any evidence of residual resources might lead budget officials to cut next year's appropriation. The frequent result of these rules is a rush to make end-of-the-year expenditures for purposes that would not have been considered high priorities earlier in the year.

Other Organizational Barriers: Other organizational barriers that challenge public sector managers include the following:

- Inadequate management commitment to productivity—The support of some managers is more rhetorical than real.
- Barriers to monetary incentive plans—Plans for monetary rewards sometimes confront citizen or legislative objections over cost and appropriateness, and even the opposition of supervisors and some

employees worried that the plan will fall prey to favoritism.

- Union resistance—Unions often push for restrictive work rules and higher wages and oppose initiatives that would reduce the workforce, although exceptions in the form of management–labor cooperation in managed competition and other efforts are also noteworthy.
- Perceived threat to job security—When the term productivity is assumed by the workforce to be a pseudonym for “speed up” or “automation” with an accompanying loss of jobs, employee cooperation can be difficult to secure.
- Excessive layers of middle management.
- Supervisory resistance—Supervisors are ideally situated in the organizational ranks to thwart any productivity improvement initiatives they oppose.
- Ambiguous objectives—This is very much like having no objectives at all.
- Reluctance to abandon—Once in place, even unsuccessful programs die hard.
- Insufficient analytical skills or analytical staffing.
- Inadequate cost-accounting systems.
- Inadequate performance data for one’s own operation and inadequate comparison data from others.
- Inadequate performance evaluation/appraisal.
- Inadequate dissemination of information about successful programs, accompanied by a reluctance to use what *is* known.
- Fragmentation of government—The belief shared by many, but rejected by others, that having a host of small, overlapping governments obscures lines of accountability, promotes unnecessary duplication, and forfeits opportunities for economies of scale.
- Bureaucratic rigidities, including an extreme focus on rules and procedures.
- Fragmented authority and uncertainty about where decisions can be made.
- Turf protecting tactics by entrenched public officials.
- Inadequate research, development, and experimentation with technological improvements and other advances for government operations.
- Requirement of large initial investment for productivity improvement efforts—Major gains often require substantial front-end investments.
- Performance myths confound rational management—Myths, for instance, regarding the inadequacy of one-officer patrol cars and the unsuitability of competitive systems for public services have limited managerial thinking in the past.
- Overselling productivity—Overly ambitious promises made to “sell” a productivity improvement initiative can come back to haunt program officials if program results, although positive, fail to match the promises.

PERSONAL BARRIERS

Some barriers are largely personal in nature. Although often influenced by characteristics of the public sector environment or a given organization, these are obstacles that are more internal to the manager or the manager’s job. Among the key personal barriers are the following:

Inadequate Control of Time/Workday: Most government managers, like their counterparts in the private sector, work long hours. Time constraints and the pressures of competing demands for managerial attention constitute an obstacle to productivity improvement, because they limit the manager’s ability to explore new ideas for operational improvement. For some, the enormous demands on managerial time constitute a major barrier; for others, more adept at time management, including prudent priority setting and the art of effective delegation of duties, the obstacle is somewhat less formidable.

Conceptual Confusion: Confusion over the meaning and ramifications of productivity in the public sector complicates communication, undermines analytical and planning efforts, and limits achievement.^[8] Those who view productivity improvement narrowly, simplistically, or apprehensively restrict opportunities and reduce the likelihood of success. The notions of “more bang for the buck” and “working smarter, not harder” only begin to capture the essence of productivity improvement; equating it to forced speedups or cut-back management captures even less.

Proclivity to “Paralysis by Analysis”: A conscientious administrator or researcher would prefer to base important decisions on logic and thorough analysis of complete data. Unfortunately, the information can never be complete; the analysis can never be perfect. At some point, the productive manager has to move forward with a decision based on careful analysis of good, but rarely perfect or complete, data.^[1] Managers who are unable to move forward following a *conscientious and reasonably complete* analysis may inadvertently lock in place a status quo operation, despite opportunities for improvement. They are paralyzed by the thought that a little more analysis would clarify the decision and reveal an obvious course of action. “Paralysis by analysis” is a self-inflicted malady. In some cases, it is produced by perfectionism; in others, it is merely an avoidance mechanism for an administrator who would prefer not to have to make a difficult recommendation or decision.

Risk Avoidance: Successful risk taking in the public sector offers few of the personal rewards enjoyed by successful entrepreneurs in the private sector. Entrepreneurial public managers who take well-reasoned risks that benefit their government and its citizens are inspired instead by commitment, professionalism, and

their perspective on public service. Because the penalty for risk-taking failure can be substantial, many public executives choose to curb their entrepreneurial efforts. Some observers, in fact, even perceive a “relatively low risk threshold of most public officials.”^[9] Others detect an even more serious barrier than the failure to encourage risk taking, suggesting that “innovative people do not choose careers in the public sector.”^[10] Evidence of entrepreneurial behavior in the public sector shows clearly, however, that some entrepreneurs do choose public service careers.

Managerial Alibis: Public sector managers sometimes blame bureaucratic regulations, civil service restrictions, political forces, and the power of long-standing traditions for their inability to adopt innovations or to make other changes that would correct operational inefficiencies. In many cases, these are merely alibis for the managers’ unwillingness to mount the effort necessary to bring change. Innovative, determined managers find ways to work within legal restrictions, civil service regulations, and other administrative rules, or they find ways to work around or amend them. For these officials, bucking tradition may be difficult but not impossible.

CONCLUSION

All public sector managers confront barriers to productivity improvement. Some are particularly adept at recognizing potential obstacles, assessing their

magnitude, and devising appropriate strategies for neutralizing or circumventing them. These are the managers who are most likely to lead their organizations or departments to greater achievements and higher levels of productivity.

REFERENCES

1. Osborne, D.; Plastrik, P. *The Reinventor's Fieldbook: Tools for Transforming Your Government*; Jossey-Bass: San Francisco, 2000; 42, 157.
2. Savas, E.S. *Privatization and Public-Private Partnerships*; Chatham House: New York, 2000; 31 pp.
3. Savas, E.S. *Privatization: The Key to Better Government*; Chatham House: Chatham, NJ, 1987; 97 pp.
4. Niskanen, W.A. *Bureaucracy and Representative Government*; Aldine, Atherton: Chicago, 1971.
5. Downs, A. *Inside Bureaucracy*; Little, Brown, and Company: Boston, 1967.
6. Tullock, G. *The Politics of Bureaucracy*; Public Affairs Press: Washington, DC, 1965.
7. Marris, P.; Rein, M. *Dilemmas of Social Reform*; Aldine: Chicago, 1973; 45 pp.
8. Balk, W.L. Productivity in government: a legislative focus. *Public Prod. Rev.* **1984**, 8 (2), 148–161.
9. Steiss, A.W.; Daneke, G.A. *Performance Administration: Improved Responsiveness and Effectiveness in Public Service*; Heath, D.C., Lexington Books: Lexington, MA, 1980; 170 pp.
10. Borins, S. Loose cannons and rule breakers, or enterprising leaders? Some evidence about innovative public managers. *Public Admin. Rev.* **2000**, 60 (6), 500.

Public Service

Richard A. Loverd

Acton, Massachusetts, U.S.A.

INTRODUCTION

Public service, literally serving the public, can be viewed from several perspectives. On the one hand, it can be considered as a place or “locus of employment . . . embracing all those who work in the public sector.”^[1] Then again, as Elmer Staats has noted, it can also be perceived as “a concept, an attitude, a sense of duty—yes, even a sense of public morality,”^[2] suggesting a public service ethic. In this article, both points of view will be explored.

PUBLIC SERVICE AS A PLACE: THE UNIQUENESS OF PUBLIC SECTOR WORK

Public service is very different from private enterprise, so much so that Wallace Sayre has suggested that “public and private management are alike in all unimportant respects.”^[3] While such a view may be a bit extreme, there are major constraints that uniquely shape public-sector work. In particular, those constraints are reflected in a public environment that is diffuse, political, bureaucratic, and very public.

Diffusion

At base, our system of government was not designed to be manageable; it is purposefully, indeed constitutionally, difficult and diffuse. Such ideas are reflected in the words of Graham Allison when he notes that “In business, the functions of the general management are centralized in a single individual: the Chief Executive Officer (CEO). In contrast, in the U.S. government, the functions of general management are constitutionally spread among competing institutions: the executive, two houses of Congress, and the courts. The constitutional goal was “not to promote efficiency but to preclude the exercise of arbitrary power,” as Justice Brandeis observed.” Moreover, Allison further stresses that, “as *The Federalist Papers* make starkly clear, the aim was to create incentives to compete so as to provide security against a gradual concentration of the several powers in the same branch.” In so doing, “the general management functions concentrated in the CEO of private business are,

by constitutional design, spread in the public sector among a number of competing institutions and thus shared by a number of individuals whose ambitions are set against one another”^[3]. Indeed, with power divided and ambition pitted against ambition, it is sometimes a wonder that the public service works as well as it does!

Politics

The very diffusion of power, pitting ambition against ambition, encourages political activity as public servants try to marshal forces that will help them accomplish tasks, protect prerogatives, and expand their power bases. In addition to the power vested in their formal positions, they might use rules, funds, jobs, expertise, tradition, charisma, information, timing, group pressure, and any number of other power sources to influence those within and around their work settings.

As an especially instructive example, jobs have served as a source of power, particularly political party power, throughout the history of American public service. During the 1800s, from around the time of Presidents Jackson to Garfield, government jobs were viewed as “spoils” or prizes to be given to those who worked for the party and helped victorious candidates win public office. Consequently, partisan service, rather than individual competence, was used as the basis for awarding public jobs; and the jobs themselves were deemed of little importance except as they contributed to party power.

Unfortunately, the spoils system soon became a two-edged sword, sharper on one side than the other. For while it proved helpful to the party leadership, it became quite harmful to the country. Because their jobs tended to last only as long as their party candidates were in office, appointees were tempted to plunder their public trusts for personal gain. Thus, the use of jobs as a political means for enhancing party power came to be associated with a lack of competence and an excess of corruption in the government. And when a disappointed office seeker named Charles Guiteau assassinated President James Garfield in 1881, that use was associated with murder as well. Obviously, something had to be done.

Bureaucratic Behavior

To curb the excesses of party interests and restore some measure of responsibility and competence to public servants, bureaucratic reforms were initiated through legislation, first at the federal level, with the passage of the Civil Service Act of 1883, and soon after by states and localities.

These reforms, by being bureaucratic, stressed the use of a rational-legal rather than a partisan-political approach to public management. In so doing, to use Max Weber's terms, impersonal "calculable rules," rather than personal partisan preferences, were used to provide the basis for uniform, even-handed judgments about public employees; and, with no pejorative overtones intended, public managers were to "go by the book" and see that others did the same within their specified spheres of authority, responsibility, and competence. Indeed, competence itself would be judged on the basis of technical proficiency rather than political affiliation or social status; and, as much as possible, employees would be kept away from the corrupting, debilitating influences of partisan politics.

The bulk of these bureaucratic ideas is included in what has come to be known as the "merit system concept" of civil service reform, a concept that provides the cornerstone for civil service policies and processes. This concept includes three fundamental tenets set forth for civil servants (i.e., public bureaucrats): "First, among the historic principles is recruitment via competitive examinations, or on the basis of job ability or individual competence. A second major principle is absence of arbitrary removals, or relative security of tenure. The third fundamental principle is that of political neutrality in the civil service."^[4] Thus, competent people are selected through open, competitive examinations instead of partisan politics and subsequently protected from politically motivated removal by job tenure. And in exchange for that tenure, they are expected to remain politically neutral by making their competence available to the political party leaders of the day.

It should be noted that while the notions of competence and tenure seem straightforward enough, political neutrality has caused problems from the outset. Might not merit employees be tempted to resist the wishes of the political party machinery, prove unresponsive, and attend to their own bureaucratic machinery instead? Indeed, according to a report by the Committee for Economic Development just such bureaucratic merit system problems have come to pass, including: 1) bureaucratic delays, which prevent the prompt filling of job vacancies; many times leading to the loss of top applicants to other job opportunities; 2) recruitment by personnel offices, which is often slow, unimaginative, and unaggressive; 3) rigid

classification systems, which impede the efficient assignment of work; 4) excessive reliance on written examinations, seniority, and other criteria that have little or no relation to job performance; and 5) a lack of authority available to managers to reward superior performers and discipline or fire non-performers.^[5] All these bureaucratic constraints reflect what Wallace Sayre once tagged as a "triumph of techniques over purpose,"^[6] techniques that, in turn, can heighten inefficiency and unresponsiveness and diminish the executive's capacity to manage.

Publicity

James Forrestal, the nation's first secretary of defense after World War II, once stressed that "the difficulty of government work is that it not only has to be well done, but the public has to be convinced that it is being well done. In other words, there is a necessity both for competence and exposition, and I hold it is extremely difficult to combine the two in the same person."^[7] This need for exposition, to have the public's business openly conducted and judged for all to see, can prove quite challenging to managers in government, particularly those used to the privacy of the business world. In the private sector, far more of its managerial activities are, by definition, private: more internal, less visible, and less subject to public and media scrutiny. By contrast, in the public sector, with instant global communications and an intrusive press, managers can expect to be placed in a goldfish bowl, scrutinized about anything and everything that can or may come their way.

PUBLIC SERVICE AS AN ATTITUDE: THE APPEAL OF PUBLIC SERVICE

Given all of the discomfort inherent in the constraints described above, what stirs people to want to seek public service? Obviously it is much more difficult to manage, more political, less efficient, less financially rewarding and more intrusive than private-sector work. So where lies the appeal?

James Perry and Lois Wise suggest that there are different kinds of motives unique to the public service that draw people and keep them, ones that lean less on comfort and economic prosperity as the main incentives. Instead, such attractions as the opportunity to express loyalty to one's country, to serve the public interest and some greater good, to participate in the making of public policy, to redress inequities, to advocate a special interest, or to commit to a public program are more readily provided through government service.^[1] And it is these types of motivational factors

that need to be emphasized when governments are seeking to attract new employees to their ranks.

Of course, communicating the appeal of public service does not come without its challenges, not the least of which is the degree of confidence the public has in its government, a level of support that has not always been high. For example, from the mid-1960s to the 1990s, an ongoing decline in public confidence in American institutions occurred to the point where only one in four Americans believed that the government would “do what is right.”^[1] This decline in public trust has made it all the more difficult to recruit and retain a high quality public workforce, leading many to choose, and trust, private over public-sector employment.

To remedy such low confidence, expectations need to change outside, as well as within the confines of public employment, a condition that will likely require environmental shifts regarding the way in which the public views its public servants as much as any improvements within the public service itself. To wit: The more the public perceives a need for, and appreciation of, its public servants, the more likely its appeal will rise. Therefore, when events requiring a need for governmental intervention present themselves, if public servants should demonstrate their ability to fill that public need in the most productive and effective way possible, confidence and trust will grow.

Perhaps no better recent example of public service at its best can be cited than its response to the attacks by Muslim terrorists on the Pentagon and the World Trade Center on September 11, 2001. The response by police, fire, and emergency medical teams at the local level, and a host of state and federal agencies at higher levels, was impressive, leading Americans to rethink their relationship with their government and perhaps turn back to, and trust it, once again. Indeed, by October 2001, one *Washington Post*/ABC News poll noted that public confidence in Washington to “do what was right” rose to 64%—“within the range of public confidence in 1964, before the Vietnam protest era.”^[8] A good start for not only renewing confidence but also motivating people to reconsider the appeal of public service.

CONCLUSION

Public service can be analyzed as a place of employment or an attitude. As a place, its uniqueness has been emphasized through examination of the constraints that shape its activities. As an attitude, its appeal has been viewed through consideration of the distinct motives that draw people to want to work in such a place. While the challenges are great, so too can be the rewards. Hopefully, the examples of service demonstrated by public employees during times of need will encourage more of the public to appreciate what they do and lead more potential employees to want to participate in its noble tasks.

REFERENCES

1. Perry, J.; Wise, L. The motivational bases of public service. In *Contemporary Public Administration*; Rosenbloom, D., Goldman, D., Ingraham, P., Eds.; McGraw-Hill: New York, 1994; 251–255.
2. Staats, E. Public service and the public interest. *Public Adm. Rev.* March/April, **1988**, 48, 601–605.
3. Allison, G. Public and private management: are they fundamentally alike in all unimportant respects? In *Public Administration: Concepts and Cases*, 6th Ed.; Stillman, R.J., II, Ed.; Houghton Mifflin: Boston, 1996; 288, 291 pp.
4. McMurty, V. Merit principles and the federal service: 1789 to the present. In *History of Civil Service Merit Systems of the United States and Selected Foreign Countries*; U.S. House of Representatives, Committee on Post Office and Civil Service: Washington, DC, 1976; 7 pp.
5. Committee for Economic Development. *Improving Management of the Public Work Force*; Committee for Economic Development: New York, 1978; 44–46.
6. Sayre, W. The triumph of techniques over purpose. *Public Adm. Rev.* Spring, **1948**, 134–137.
7. Corson, J. Distinguishing characteristics of public administration. *Public Adm. Rev.* Spring, **1952**, 124.
8. Shribman, D. A renewed interest, trust in government. *Boston Globe* October 16, **2001**, A13.

Public–Private Partnerships for Economic Development

Steven G. Koven
Stuart C. Strother

School of Urban and Public Affairs, University of Louisville, Louisville, Kentucky, U.S.A.

INTRODUCTION

A public–private partnership is a formal complementary relationship between public and private entities to achieve a common objective. The formal partnership arrangement delineates each partner's roles and responsibilities, states the level of investment and risk of each partner, and describes how financial and non-financial benefits will be distributed between the partners. Although partnerships represent government “power sharing,” they also entail “risk shifting” from government to the private sector. The basic purpose of partnering is “to take advantage of the potential for all parties to gain greater benefit than they could on their own.” When public–private partnerships are created, the overall objective of the partnership is usually to increase jobs, to enhance the numbers of employers in a region, or to revitalize the physical assets of an urban area. This encyclopedia entry delineates various public strategies to achieve growth, privatization models, and future directions of public–private partnerships.

PUBLIC–PRIVATE STRATEGIES TO ACHIEVE GROWTH

State and local public actors enter into partnerships with private organizations to accomplish the objective of economic growth.^[1] Some of the most common objectives of an economic development public–private partnership are to improve the general image of an area and to improve an area's business climate.^[2]

Some public–private partnerships are specifically designed to market their area to outside firms. Surveys indicate that about 40% of U.S. cities engage in marketing their cities, and about 21% of cities partner with private economic development foundations. Private foundations typically engage in activities such as developing promotional materials, establishing Web sites, advertising in the media, hosting special events, and sponsoring trade missions abroad.^[3]

Various strategies are utilized to improve the business climate of an area. In general, these strategies can be placed into categories of: 1) providing direct financial incentives; 2) improving the tax environment;

and 3) providing non-financial incentives.^[4] Each of these strategies involves some type of partnership and agreement between governmental entities and representatives of the private sector. These strategies are reviewed below.

Financial Incentives

Common financial incentives provided by government entities to the private sector include grants and loans. Grants have been a part of American history since its founding. Land grants to railroads and homesteading helped open up the nation's west. The Morrill Act of 1862 subsidized schools that offered instruction in agricultural and mechanical arts. Contemporary examples of federal grant programs include funding for urban renewal, small businesses, community development, and urban development. A federal government agency involved in providing grants to private entities is the Economic Development Administration (EDA). Initially, EDA programs were targeted to depressed rural communities. After the mid-1970s, programs were increasingly directed to urban areas. Public works grants, business loans, technical assistance grants, and grants to assist communities affected by economic dislocations are programs administered by EDA. Communities adept at marketing their needs were more likely than others to gain access to these federal programs.

Public entities provide loans for the private sector. The federal EDA provides loans under a Business Development Loan program. States provide loans to enterprises that promise to enhance economic development. These loans are often targeted for specific purposes. For example, the state of New Jersey targeted loans to firms that promise to locate in low-income/high-unemployment areas. Pennsylvania targeted a proportion of its loans to firms that deal with advanced technology. Alaska and Hawaii offer loans to owners of fishing vessels who are unable to secure conventional financing. Ohio gave priority for loans to businesses already operating in the state.

Public loan programs carry the risk of default and financial loss to taxpayers. This risk is assumed by the public sector in efforts to spur economic growth. Usually, state or local departments of economic

development or industrial development authorities make direct loans. Revolving loan funds recycle payments of outstanding loans to other businesses. In theory, once a revolving loan fund is established, it exists in perpetuity and is self-financing as long as default rates as well as administrative costs are relatively low. Loans are often made by community development loan funds at below-market interest rates to help entrepreneurs start new businesses.

Tax Environment

Tax incentives are often offered to businesses in hopes of attracting major employers or retaining existing firms. Typical incentives offered to businesses include tax abatements, tax exemptions, tax credits, and tax increment financing.

Tax abatements are legal agreements between a government entity and a property owner to forgo taxing some share of assessed real estate value for a given period of time. It is expected that when the abatement period expires, rates will be reapplied, tax revenues will increase, and tax base will be strengthened. Tax abatements are the most popular type of tax incentive. They are often used to encourage economic development in depressed or blighted areas. Tax exemptions refer to reductions in the base upon which taxes are calculated. Businesses may be granted exemptions on corporate income taxes, commodities such as fuel, inventories, land, capital improvements, and equipment. Tax credits directly reduce tax obligations. They are more desirable than tax abatements and tax exemptions because they produce a dollar for dollar reduction in the amount of taxes one pays. Tax credits are often granted for activities (such as investment in new plant and equipment) that will produce long-term benefits for the jurisdiction.

Tax increment financing allows jurisdictions to borrow funds and earmarks increased tax revenues (attributable to improvements made with the borrowed funds) to repay debt. When bonds are fully paid, a tax increment finance district can be dissolved, and the full share of the tax increase is assigned to a jurisdiction's general revenue fund. Proponents of tax increment financing cite the facts that it does not incur out-of-pocket expenses, it does not obligate jurisdictions to raise taxes, and it does not place added obligations on property owners who reside outside the tax increment district. Critics of tax increment financing claim that it bypasses voter approval, it increases the need for essential services that must be paid by citizens living outside the increment district, it is more expensive than general obligation debt, and it subsidizes development that would have been undertaken in the absence of government assistance.

Non-Financial Incentives

Site development, enterprise zones, and human resource development represent non-financial public activities that assist in attracting private development. These activities are not cost-free and are justified in the name of economic development.

State and local governments acquire and improve sites for industrial and commercial use. The object of site development is to attract private firms by offering land at a reduced cost. Capital improvements are often made to land in terms of water and sewer lines, landscaping, lighting, gutters, and curbs. The land is then sold to developers below cost. Jurisdictions provide catalogs of available industrial sites or build infrastructure to suit the needs of the private firm.

Jurisdictions often engage in land banking or the practice of acquiring and improving contiguous parcels of land suitable for development. To construct a land bank, public officials search for underused, underdeveloped, and/or misused properties. Officials catalog properties by size and location, keeping the information up to date for quick reference. Land banking usually requires a substantial amount of funds for purchases. Potential sources of funds include surtaxes on local real estate and the sale of bonds.

Enterprise zones provide tax incentives and regulatory relief to businesses that are willing to locate in a designated zone. Proponents of enterprise zones maintain that lowering taxes and red tape will induce companies to make investments in areas that typically are neglected. Such investments, in theory, can provide employment in declining neighborhoods. Critics of enterprise zones claim that they do not create new jobs but only shuffle jobs from one location to another. Furthermore, enterprise zones tend to benefit large, capital-intensive corporations with high tax liabilities. The public sector loses the tax revenue that would be generated from those large corporations when enterprise zones are created.

Many companies will not invest in areas where the workforce is viewed as inadequate. Inadequacies can be traced to insufficient numbers of the types of workers demanded by the enterprise. To alleviate workforce concerns, jurisdictions often provide customized training. This training is typically provided at the employer's work site or a local community college. Many state and local governments tailor training to the specific needs of the individual firm. In a specific example of an innovative public-private partnership, the commonwealth of Kentucky, colleges in the Louisville area, and the large package-handling firm United Parcel Service joined together in a collaborative arrangement that helped to expand the pool of nightshift part-time package-handling employees.^[5] This partnership assured the retention of the largest employer

in the commonwealth of Kentucky, increased tuition revenue for the participating colleges, and expanded access to education.

General economic development strategies that are discussed above have produced numerous success stories. These success stories generally occur when public leaders alter investment decisions of the private sector through the provision of inducements. An alternative perspective on public-private partnerships is viewed from the lens of service delivery. Government services can be delivered through a myriad of arrangements ranging from mostly public delivery to mostly private. Efforts on the part of public sector leaders to shift the delivery of services to private sector providers represent a major development in American governance. Some of the models that represent shifts from public to private provision of goods and services are described below.

PUBLIC-PRIVATE PARTNERSHIP MODELS FOR SERVICE DELIVERY

The level of collaboration between public and private actors in a partnership varies from case to case. Many public services have traditionally been provided directly by the government with no collaboration with the private sector. Intergovernmental agreements (such as a county and city sharing jail facilities, road maintenance responsibilities, or an airport) represent an example of collaboration between multiple public organizations. But the public sector is increasingly looking to the private sector for partners who can produce public services. Savas^[6] lists 10 service delivery arrangements and ranks them by their degree of privatization or their degree of reliance on market forces. According to Savas,^[6] free market, voluntary service, and self-service arrangements are the ultimate in privatization because they contain the least government involvement. Four other types of service delivery arrangements utilize private-sector production. Ranked in decreasing degrees of privatization, these arrangements include franchises, vouchers, grants, and contracts. Service delivery arrangements that utilize a government producer include government vending, intergovernmental agreements, and government service. Government service is considered to have the lowest level of privatization.

The *free market* provides most ordinary goods and services. Consumers select a producer, which is a private firm and pays the private firm. Government is not involved in the transaction in any significant way. *Voluntary service* is provided through organizations that provide services that ordinarily might be supplied by the public sector. Examples of such services include recreational programs run by sports enthusiasts,

protective patrols run by neighborhood associations, and fire protection provided by volunteer fire departments. *Self-service* is the most basic of all delivery systems. It refers to people providing their own services such as installing burglar alarms in their homes to prevent crime.

Franchises are awarded by governments to private firms who agree to provide a unique public service such as airport operation, utilities, and toll roads. Government gives the firm permission to operate in a specific geographic area, and citizens pay the firm directly for services. *Vouchers* are subsidies given directly to consumers to purchase goods such as food (food stamps) or education (school vouchers). Unlike grants where the government decides which producers get the resources, consumers decide where to spend their vouchers. Job-training vouchers and the G. I. Bill are voucher programs designed to increase human capital and are therefore relevant from an economic development perspective. *Grants* are resources given by government to private firms who can then provide more affordable good or service to citizens. Grants are often subsidies, in the form of money, tax exemptions, tax abatements, or low-cost loans. For economic development, many state and local governments give away land at no cost to new firms. *Contracting* involves government payment to a private entity to provide a specific service. Virtually all governments procure some goods or services by contracting with private firms. Services related to economic development that are often provided by contractors include economic development attraction activity, road and building construction, convention center management, industrial development, and urban planning.

Government vending involves the purchase of goods or services from a governmental agency. An example of government vending is the sale of rights for water, mineral, timber, and grazing on livestock on government-owned land. *Intergovernmental agreements* involve one government hiring another to supply a service. For example, a state may contract with a city or county to provide social services, or a small community may purchase library services from a specialized government unit that was organized to sell services to other governments in the region. *Government service* denotes the delivery of a service by a government unit using its own employees. An example of government services is sanitation service provided by a municipal sanitation department or education provided by the local public school system.^[6]

Infrastructure can also be supplied and delivered through various public-private partnerships. Some partnerships that utilize private sector arrangements to supply infrastructure include lease-build-operate, build-transfer-operate, build-own-operate transfer, and build-own-operate.^[6]

Under *lease-build-operate* partnerships, a private firm is given a long-term lease to develop (with its own funds) and operate a facility. The firm pays a rental fee and reaps a reasonable return on its investment over the term of the lease. The facility remains publicly owned. Under the *build-transfer-operate* partnership, a private firm builds a facility and transfers ownership to a government agency. The government agency then leases the facility back to the private firm. The private firm operates the facility and has the opportunity to recover its investment. The firm earns a reasonable return from user charges and commercial activities.

In the *build-operate-transfer* arrangement, a private developer is awarded a franchise to finance, build, own, and operate a facility and to collect user fees for a specific period of time. At the end of the time period, ownership is transferred to the public sector. In contrast to a sale, government retains control over the project. Under the *build-own-operate* partnership, the private developer finances, builds, owns, and operates a facility in perpetuity under a franchise agreement. Long-term property rights provide a significant financial incentive for capital investment. Examples of build-own-operate transfer arrangements include the construction of a bridge in Shanghai, China, the construction of the “chunnel” under the English Channel, the development of private toll roads in California, and the creation of power plant facilities in Indonesia.^[6,7] The build-own-operate arrangement resembles outright privatization.^[8]

TRENDS AND DIRECTIONS OF PUBLIC-PRIVATE PARTNERSHIPS

The use of public-private partnerships is believed to be on the rise at the start of the 21st century because of partnership's ability to leverage private capital.^[8] Public sector leaders primarily enter into public-private partnerships because “they are seeking additional capital for economic expansion.”^[9] In many instances, partnerships facilitate the construction of worthy goods, allow private sector contractor to make a profit, and permit eventual public sector ownership of facilities without government expenditure.

Another factor contributing to the popularity of public-private partnerships appears to be the decentralization of decision making from the national to local governments. Local governments often lack the fiscal resources necessary for public projects. Therefore creative entrepreneurial approaches, such as partnering with the private sector, are useful to provide needed public goods and services. Partnerships between public and private sector officials offer the promise of increasing efficiencies in the delivery of public services as well as increasing the leveraging of

private sector funds. These private sector funds are utilized for public projects.

A third reason for the rise in partnerships between government and industry relates to structural changes in the economy. In the new marketplace, firms increasingly compete across national boundaries, and public officials have a vested interest in maintaining international competitive advantages. In the United States, with its so-called “New Economy,” economic growth relates to technological advances and corporate strategy.^[10] In efforts to maintain and expand comparative advantages, technology partnerships between public and private actors have evolved.^[11]

A leading example of a high-profile public-private partnership is SEMATECH, a partnership formed in 1987 between U.S. computer-chip makers and the U.S. federal government. Under this partnership, the federal government provided funding for a collaborative effort to increase the computer-chip market share of American companies. At the time of the creation of the SEMATECH, companies based in the United States lagged behind Japanese-based companies in the sale of semiconductors.^[10] After almost a decade, the United States became the world leader in the semiconductor market, and the federal government withdrew public funding for SEMATECH.

A number of factors coalesce to expand the trend toward public-private partnerships. Decentralization of political power, an entrepreneurial public sector management style, fiscal stress, and the need for infrastructure all contribute toward this trend. The use of partnerships is viewed as especially critical in regard to infrastructure creation.^[12]

CONCLUSION

Savas noted that privatization of government functions is in the ascendancy, and that a majority of Americans no longer believe that an expanding state is the answer. He stated that “throughout the world we are experiencing a reorientation of government away from a top-down approach, an abandonment of the reigning assumption that a powerful, active, and interventionist government, manned by a caring intellectual elite and driven by good intentions, is the basis for good society.”^[6] Public-private partnerships are likely to further evolve as acceptance of privatization as a basic strategy for societal governance grows in acceptance.

Some authors, however, caution against the blind acceptance of increased privatization. Rosenau^[13] noted that private and public partnerships may suffer because of conflicting interests. Conflict is attributed to the fact that the private sector is oriented toward profit, risk, competition, and corporate goals, whereas

the public sector is oriented toward legislation, public opinion, democratic decision making, minimizing risk, and realizing social goals. Because of this inherent conflict, policymakers must guard against providing services to vulnerable populations (the children, the elderly, the disabled, and the cognitively impaired) in such a manner as to maximize profit at the expense of quality, access, or equity. In addition, public-private partnerships should avoid cronyism or improper awarding of contracts to private firms who enrich themselves with public resources.^[14] Public-private initiatives that erode participation, equity, access, and democracy concerns must be minimized. In general, however, public-private partnerships have proven to be an effective tool for economic development projects, and it is expected that they will continue to be utilized in the future.

REFERENCES

1. Linder, S.H. Coming to terms with the public-private partnership: a grammar of multiple meanings. *Am. Behav. Sci.* **1999**, *43* (1), 35–51.
2. Mullin, S.P. *Public-private partnerships and state and local economic development: leveraging private investment*; Economic Development Administration: Washington, DC, 2002.
3. *Economic Development 1999 [Data file]*; International City/County Management Association: Washington DC, 2003.
4. Koven, S.; Lyon, T. *Economic Development: Strategies for State and Local Practice*; International City/County Management Association: Washington, DC, 2003; 27–53.
5. Koven, S.; Strother, S. Saving jobs in Louisville, Kentucky. *Econ. Dev. J.* **2002**, *1* (1), 19–22.
6. Savas, E. *Privatization and Public-Private Partnerships*; Seven Bridges Press, LLC: New York, 2000; pp. 67–86, 104, 243–247, 327.
7. Wu, A.; Wu, R. A trip to the construction site of dreams. In *Zhejiang Online*; 2003. Retrieved September 20, 2003, from <http://www.zjol.com.cn/gb/node2/node138665/node139012/node149968/node162419/node162422/userobject15a1729735.html>.
8. Williams, T.P. *Moving to Public-Private Partnerships: Learning from Experience Around the World*; IBM Endowment for the Business of Government: Arlington, VA, 2003.
9. Walzer, N.; Jacobs, B.D., Eds.; *Public-Private Partnerships for Local Economic Development*; Praeger: Westport, CT, 1998; 16 pp.
10. Carayannis, E.G.; Alexander, J. Revisiting SEMATECH: profiling public- and private-sector cooperation. *Eng. Manag. J.* **2000**, *12* (4), 33–43.
11. Stiglitz, J.E.; Wallsten, S.J. *Public-private technology partnerships: promises and pitfalls*. *Am. Behav. Sci.* **1999**, *43* (1), 52–73.
12. Klitgaard, R.; Treverton, G.F. *Assessing Partnerships: New Forms of Collaboration*; IBM Endowment for the Business of Government: Arlington, VA, 2003.
13. Rosenau, P.V. The strengths and weaknesses of public-private policy partnerships. In *Public-Private Policy Partnerships*; Rosenau, P.V., Ed.; The MIT Press: Cambridge, MA, 2000; 217–241.
14. Rosenau, P.V. Introduction: the strengths and weaknesses of public-private policy partnerships. *Am. Behav. Sci.* **1999**, *43* (1), 10–35.

Public–Private Partnerships in Developing Countries

Steven G. Koven
Stuart C. Strother

School of Urban and Public Affairs, University of Louisville, Louisville, Kentucky, U.S.A.

INTRODUCTION

In recent years, developing countries have been embracing innovative methods of service delivery including privatization, quasi-private arrangements, and public–private partnerships. Over time, officials in these nations have come to realize that government initiatives alone cannot properly address the enormous economic and social challenges that they confront. The needs of developing countries are truly vast and are beginning to be addressed through various combinations of partnerships between philanthropic organizations, non-government organizations, corporations, community-based organizations, and government entities.

Partnerships exist in developing countries for an array of activities including health care, transportation, energy policy, technology development, protecting the environment, making agriculture more productive, generating revenue from tourism, and many more activities. These partnerships offer the promise of generating profits for private companies, increasing government revenue, and improving the lives of people who often do not have access to safe water, electricity, or all-season roads.

Leaders in developing countries have come to recognize that government alone has failed to meet developmental goals. In turn, these leaders have begun to look at innovative partnership types in an effort to improve the living conditions of their constituents. This encyclopedia entry will describe why delivery systems such as privatization have grown in popularity, the decline of statist thinking, constraints on effective public–private partnering, and examples of successful partnerships.

WHY PRIVATIZE?

The movement away from monopolistic, government provision of goods and services has steadily advanced since the latter half of the 20th century. Numerous economic and social reasons have been presented to explain events such as the collapse of Soviet as well as Eastern European communism after 1989, the rise of leaders such as Margaret Thatcher in Britain and Ronald Reagan in the United States, and the renewed

acceptance of neoclassical economics in developing countries.

Savas^[1] identified five factors that have propelled the rise of privatization throughout the world. These factors include the impact of 1) pragmatic forces, 2) economic forces, 3) ideological forces, 4) commercial forces, and 5) populist forces. According to Savas, pragmatic forces promote privatization because there exists a need to continuously provide services to constituents in environments of rising costs and resistance to higher taxes. It is believed that privatization can assuage these pressures by enhancing productivity, reducing waste, and promoting efficiency. Productivity may be enhanced in a variety of manners. In instances where government-owned enterprises are dominant, they can be forced to compete with private companies. Competition between public and private entities can transpire for government contracts or for satisfying consumer desires. Economic forces that promote privatization relate to a growing affluence in some parts of the world and affluence's corollary of less dependence on government. For example, if more people own their own automobiles, there will be less reliance on public transportation.

Ideologically, some political theorists see an expanding government sector as a threat to freedom and liberty. This threat to individuals has been well documented historically. Nations such as Germany (during the Nazi regime of Adolf Hitler) and Russia (after the 1917 Bolshevik revolution) witnessed the growth of massive state entities that exercised almost total control over the lives of individuals. It is therefore not surprising that individuals living in pluralistic societies are wary about ceding power to impersonal government institutions. People in developing countries have often found themselves in situations where their individual rights were subsumed to a collective ideal that was defined by the state. Seizure of private property often preceded tight, central control over the economy, an economic paradigm that recently has been discredited.

Countries with large nationalized interests may also be pressured by business leaders to provide more business opportunities to private citizens (commercial forces). Savas^[2] noted that much of the work performed by government consists of routine commercial activities such as maintenance of buildings, grounds,

vehicles, and ships, data processing, collecting trash, and repairing streets. Business groups advocate the privatization of more of these activities to taxpaying, private companies. They claim that these services can be provided more efficiently and effectively if competition was permitted in contracting for services. Another segment of the private sector identifies business opportunities in financing, building, or operating large government projects, such as roads, bridges, airports, and waste-to-energy plants. In countries where nationalized industries exist, commercial pressure comes from business leaders who believe that the nationalized industries are mismanaged, slothful, and overtly underutilize existing assets. In such cases, denationalization (a particular form of privatization) is advocated. Denationalization presents the potential for innovation, whereas continued stagnation seems likely if nationalized industries remain protected from market forces. A lack of competition and protection of nationalized industries is viewed as counterproductive to innovation. State-owned enterprises or nationalized industries include a variety of enterprises such as manufacturing, mining, oil production, and transportation.

Finally, populist forces are believed to promote privatization based on widespread popular antipathy to the ideal of "big government." According to this viewpoint, government systems have become too institutionalized, too bureaucratized, too professionalized, and too protective of their own interest. Populist forces (common views of average citizens) advocate greater choice in public services, less reliance on cumbersome bureaucracies, and greater use of neighborhood, civic, church, and voluntary associations. Populist sentiment is believed to exist among citizens who are fearful of both big government and big business. These individuals find allies among advocates of family values and communitarians.

Other reasons to privatize are cited in the literature. Brinkerhoff^[3] contends that enhancing efficiency and effectiveness represents just one of a number of reasons to privatize. In addition to improving efficiency/effectiveness, public-private partnerships can also provide more integrated solutions to problems, improve outcomes for stakeholders, and promote a broader conception of the public good. More integrated solutions and improved outcomes are likely to arise when diverse actors cooperate with each other, with diverse actors free to present their own perspective. Expertise and relationships are also enhanced with greater cooperation. For example, multinational corporations and national government agencies often partner with local health agencies to learn more about decentralized/local health-care needs. In partnerships, local agencies often gain access to more capital, which, in turn, is used to address their problems. Creativity may emerge as diverse actors come together to address specific issues.

Partnerships can also create a broader understanding of the public good. Through such an understanding, sustainable benefits can be created. Social capital can be created in public partnerships, and this capital can be applied to other issues. For example, the U.S. Agency for International Development found that social capital could be developed and could facilitate continued relationships between donor and recipient organizations. This would foster support for long-term cooperation on issues of common concern.

DECLINE OF STATIST THINKING

It is no coincidence that the ascendancy of privatization thinking has coincided with a decline in "statist" ideology.^[4] Critics of countries characterized by "extreme statist experiments" (such as Cuba, Ethiopia, Mozambique, Vietnam, Laos, Cambodia, Burma, North Korea, and China) cite numerous reasons for the collapse of centrally planned societies. First, the assumption of self-centered behavior (looking out to maximize one's well being) is now recognized as applicable to both the public and private spheres. In statist societies, self-interest manifests itself in widespread corruption. Government officials use the pursuit of the common good and their ability to interpret the common good as a means to satisfy materialistic/individualistic desires. In the extreme, the selfless pursuit of common goals becomes a facade that masks individual self-interest. In contrast to classical economic thought, individual self-interest does not produce market competition/efficiency but allows public officials to extract personal rewards on the basis of their privileged government positions. Second, it is believed that "statism" fails to deal with the issue of incentives. Market incentives that imposed discipline on private sector managers and encouraged efficient production were simply not provided by the state. Third, building large-scale production units in urban areas of planned economies was detrimental to rural dwellers. Developing such units harmed those in rural areas and enriched only a small minority. Fourth, central planning neglected the question of quality. For example, Soviet oil fields were notorious for the amount of oil lost or wasted in production because Soviet managers were concerned with increasing output and not with making extraction less costly. Typically, managers in the Soviet Union were reluctant to develop new products for fear that production would temporarily lag. The private sector consumer-based market is more effective in promoting quality because profits are linked to a company's ability to attract customers. These customers are attracted on the basis of improved quality or reduced cost.^[5]

Inefficient state-run farms were commonly associated with statist countries. Analysts claim that state

entities consumed resources that would have better uses in other areas. For example, research has found that in developing countries where labor is abundant, small household farms are more cost-effective than large mechanized farms.^[6] In some nations such as Cuba, the cost of agricultural workers was found to be greater than the value of the sugar they produced. Empirical data indicate that the consequences of collectivization were not positive. In many developing nations, real per capita incomes were lower at the end of the 20th century than they were when the nations achieved independence from their colonial rulers.^[7]

Caustic criticism of the performance of statist regimes began to appear in the academic literature around the mid-1960s. By 1970, the Organization for Economic Cooperation and Development (OECD) published a study critical of state-centered import substitution industrialization (ISI) strategies that were prevalent in developing nations such as Argentina, Brazil, Mexico, India, Pakistan, the Philippines, and Taiwan. Soon after, the World Bank, in conjunction with the Inter-American Development Bank, published a study supporting the OECD's findings.

In 1983, the U.S. National Bureau of Economic Research (NBER) issued a report advocating export-oriented industrialization in contrast to import substitution characteristic of ISI-based regimes. According to the NBER, import substitution regimes were clearly less successful than regimes that focused on export-oriented industrialization. After adopting a market-based export model, some nations (particularly the Four Little Tigers or Dragons of East Asia: Hong Kong, Taiwan, Singapore, and South Korea) became models of efficiency, innovation, and prosperity. In per capita terms, these nations were the world's fastest growing in the latter decades of the 20th century.^[7] Few now dispute the NBER study's assertion that nations adopting export-based policies have performed better than nations adopting import substitution. The NBER study was so widely accepted that some began to call it the core of the neoclassical critique of statism.^[8]

As neoclassical economics regained acceptance, developing nations sought to forge partnerships with private sector companies in the developed world. In direct contrast to the prescriptions advanced in the ISI model (such as limiting imports of finished goods from developed countries and trade between poor countries), the neoclassical paradigm advocated trading between rich and poor nations.^[8] This model was embraced by economists who proclaimed that there was no need to articulate an economics for development because the market offered the best mechanism for poor countries to develop. Leaders of developing countries, however, still confronted numerous impediments to economic advancement.

CONSTRAINTS TO PARTNERING IN DEVELOPING COUNTRIES

An array of factors clearly influences the feasibility of creating partnerships in developing countries. Two factors that impinge on the ability to create partnerships relate to governance and institutional capacity.^[9] Governance in the form of laws, decrees, policies, actions, and regulations can encourage or discourage potential arrangements between public and private sector actors. Laws or decrees that inhibit profit-making, prohibit currency transfers, or limit market-based actions are likely to produce hostile environments for public-private partnerships. Governance factors considered hostile to partnerships include an absence of the rule of law principle, widespread corruption, and minimal access to policymakers. Legal impediments include overly burdensome administrative and financial requirements on potential partners. In the extreme, existing laws may outlaw the possibility of specific types of partnerships. Access to government decision makers and inefficient bureaucracies may be a source of concern. Finally, low levels of institutional capacity (witnessed in low levels of legitimacy and inexperience of major actors) can inhibit partnership creation.

Political risk also discourages investment. Five components of political risk have been identified in the literature. These risks include the danger of 1) violence or war, 2) currency restrictions, 3) contract interference, 4) expropriation, and 5) unfair regulatory environments.^[10] Violence from civil war, insurrection, and coups clearly has a deleterious impact on partnership investments. Violence often coincides with property destruction, theft, arson, and an inability to attract dependable workers. Developing countries such as Sri Lanka, Pakistan, India, Algeria, Liberia, Rwanda, and others have been plagued by large-scale violence.

Government policies restricting the ability of foreign investors to convert local currency into their home currency also constrain partnership development. Flows of funds from host states can be entirely blocked by local/host country restrictions. Such restrictive policies will severely reduce the likelihood of foreign investment. Investors find no compelling reason to risk their capital unless they can convert future profits to the currency of their home country. In some instances, foreign-exchange brokers are licensed by the host government to handle foreign exchange.

Major international events may lead to repudiation of contracts between parties. For example, the U.S.-imposed embargo against Iraq caused the cancellation of some contracts between the previous government of Iraq and some European companies. Fears of expropriation (the government confiscation of property) represent another risk. Notable examples of expropriation of assets include the takeover of foreign property

by the Castro regime following the Cuban revolution, the nationalization of foreign copper companies by the Chilean government, and the seizing of French assets by Tunisia following their independence in 1954. Finally, unfair regulatory treatment may inhibit investment. Unfair treatment includes government policies such as discriminatory tax treatment (taxes favoring the host's domestic firms), limits on access to materials needed for production, and limitations of access to local distribution systems. Private sector investors have an interest in clarifying these issues prior to making monetary commitments to enterprises in foreign nations. Whereas numerous constraints exist, there have been notable examples of successful public-private partnerships in developing nations.

EXAMPLES OF PUBLIC-PRIVATE PARTNERSHIPS IN DEVELOPING COUNTRIES

Innovative public-private partnerships have been created in both Africa and Asia. For example, in Africa, there is evidence that government officials have leveraged their niche as a center for tourism, particularly in regard to their notoriety in "big game" hunting.^[11] In various African public-private ventures, contractual arrangements are formed between a community or local institution and a private investor. These parties work together in establishing and/or operating tourism or hunting enterprises. Both the community/local institution and the private investor have rights and responsibilities under the contractual agreement. Two common forms of partnership include 1) establishment of tourism lodges and 2) leasing of hunting quotas. In the case of tourism lodges, communities either have an equity stake or receive payments from lodge operators. In the case of the sale of hunting quotas, bids are taken, and the income from the sales is either distributed to residents in the area where the hunting takes place or is used for community projects.

Both lodge construction and quota agreements have grown in popularity as rights over wildlife have devolved from the central government to the local community level. Numerous cash-starved communities in Africa view tourism as an opportunity for local economic development. Private investors have identified African tourism for its profit potential. This confluence of interests led to the creation of joint ventures between public sector officials and private investors. Specific examples of tourism joint ventures include the sale of hunting quotas in Zimbabwe, the lease of large blocks of land around game reserve in Botswana, and the tripartite agreements in South Africa between public sector conservation departments, private investors, and tribal authorities. The sale of rights to engage in trophy hunting (especially for the big five of lions,

leopards, elephants, rhinos, and buffaloes) is viewed as a potential source of revenue. Attention to sustainability and mutuality of interest is thought to be essential to the development of successful joint ventures. Written agreements and contract are important in developing the shared vision.

Another case of public-private partnership is the development of a partnership between government officials and private sector shippers to upgrade efficiency in the South Korean port city of Pusan.^[12] In the early 1990s, this was identified as a high priority of the South Korean government. The need for improved efficiency was evident to South Korean officials who were influenced by the fact that ports in general handled approximately 99.7% of Korea's trade, and the Pusan port handled about 90% of port trade. In 1998, the Pusan port was ranked as the fifth largest container port in the world.

South Korean officials focused on improving the flow of information, reducing the burden of government regulation, and shortening delays in unloading cargo. Prior to the 1990s, an estimated 50–150 documents were needed for delivering and unloading cargo. As a result of this and other factors, Pusan suffered from a perception of poor service. To improve the performance of the Pusan port, government officials concluded that better information was needed in regard to equipment availability (especially cranes and tractors for unloading), cargo losses/damages, and the availability of cargo pickup times. Each of these factors was dependent on accurate and timely flows of vital information.

To address the problem of poor service, the Korean government acquired state-of-the-art computer hardware and software between 1989 and 1990. This equipment was used in the creation of a new electronic information system termed electronic development interchange (EDI). Training of port workers for the system was instituted, and studies of the new system were conducted. In 1996, the new system was fully implemented by administrative fiat. Use of the new electronic system became mandatory among all private sector port users. Once it became clear that the Korean government had mandated participation, the private sector cooperated in the establishment of a public-private entity, the Korea Logistics Network Corporation (KL-Net). This company attracted investment from companies such as Hyundai Merchant Marine, Global Corporation, and others. By virtue of their investment, the private corporations became fully involved in KL-Net operations.

In this case, the government sector was a major initiator of the public-private partnership. The public sector invested in the creation of an EDI system, they used the coercive power of the state to guarantee use, and they received private sector contributions to run the system. From the perspective of improving port

performance, the partnership is viewed as a great success and a model for others in developing countries.^[12]

CONCLUSION

The repudiation of the centrally controlled economic system suggests that public-private partnerships will play an enhanced role in the economies of developing nations. These partnerships can be created as a consequence of state intervention (such as the case with the Pusan port) or more localized decision making such as the case of tourism in Africa. Whereas centralized economies such as those found in the former Soviet Union and Eastern Europe have been renounced, it is also doubtful that the 19th century style capitalism can be reintroduced. Given these parameters, it is likely that public and private sector officials will seek out cooperative arrangements as the mutuality of benefits becomes evident.

REFERENCES

1. Savas, E.S. *Privatization and Public-Private Partnerships*; Seven Bridges Press, LLC: New York, 2000; 5-14.
2. Savas, E.S. *Privatization and Public-Private Partnerships*; Seven Bridges Press, LLC: New York, 2000; 12-13.
3. Brinkerhoff, J.M. *Partnership for International Development: Rhetoric or Results?* Lynne Rienner: Boulder, CO, 2002; 6 pp.
4. Rapley, J. *Understanding Development: Theory and Practice in the Third World*; Lynne Rienner: Boulder, CO, 2002.
5. Rapley, J. *Understanding Development: Theory and Practice in the Third World*; Lynne Rienner: Boulder, CO, 2002; 42 pp.
6. Young, C. *Ideology and Development in Africa*; Yale University Press: New Haven, 1982; 155 pp.
7. Rapley, J. *Understanding Development: Theory and Practice in the Third World*; Lynne Rienner: Boulder, CO, 2002; 46 pp.
8. Rapley, J. *Understanding Development: Theory and Practice in the Third World*; Lynne Rienner: Boulder, CO, 2002; 58 pp.
9. Brinkerhoff, J.M. *Partnership for International Development: Rhetoric or Results?*; Lynne Rienner: Boulder, CO, 2002; 26 pp.
10. Baker, J.C. *Foreign Direct Investment in Less Developed Countries: The Role of ICSID and MIGA*; Quorum Books: Westport, CT, 1999; 13-14.
11. Ashley, C.; Jones, B. Joint ventures between communities and tourism investors: experience from southern Africa. *Int. J. Tour. Res.* **2001**, 3, 407-423.
12. Bagchi, P.I.; Paik, S.K. The role of public-private-partnership in port information systems development. *Int. J. Public Sect. Manag.* **2001**, 14 (6/7), 482-500.

Punctuated Equilibrium

Scott E. Robinson

*Department of Political Science and Public Affairs, University of Texas-Dallas,
Richardson, Texas, U.S.A.*

INTRODUCTION

Punctuated equilibrium theory has become a prominent tool for explaining policy change processes. At the most basic level, punctuated equilibrium theory is an account of policy change that predicts long-term policy equilibria that are infrequently, but dramatically, interrupted by periods of large change. The origins of punctuated equilibrium theory in the disputes between incremental and non-incremental theories of budgetary change are reviewed in this article. Contemporary punctuated equilibrium theory and some of the recent trends in the research into this policy phenomenon are also discussed in this article.

Recent research on public policy has shifted its attention from studies of cross-sectional variation in policy adoption to temporal variation. Instead of studying, say, why some states have adopted high-stakes tests while other states have not, temporal policy change researchers have asked why states have adopted policies at some times rather than others. Interest in the temporal dynamics of policy adoption in fact go decades back, but the recent resurgence of interest in the temporal dimensions of policy changes are, in part, because of the emergence of the punctuated equilibrium theory of policy change in the 1990s.

In this article, I will trace the historical roots of the punctuated equilibrium model of policy change, lay out the basic theoretical propositions of the punctuated equilibrium model, and consider the recent directions in research on punctuated equilibrium models of policy change.

HISTORICAL ROOTS OF PUNCTUATED EQUILIBRIUM

Incrementalism

The scholar most responsible for bringing attention to questions of policy change was Charles Lindblum.^[1] In his famous article “The Science of Muddling Through,” he analyzed the tendencies for slow change in administrative and policymaking groups. He argued that the slow change was a product of three pressures on policymakers. First, the policymakers had to seek consensus on any policy change—particularly in the

American government, the focus of Lindblum’s article. Second, the policymakers had very little information about potential large changes. The larger the change, in fact, the less reliable the information policymakers likely have to consider for the implications of the change. Third, policy change tends to involve very high stakes. If a large mistake is made, it may be difficult to reverse the decision.

Lindblum’s conclusion was that incrementalism dominates the policymaking process. Slow, considered, and politically non-controversial policy changes are much more likely, Lindblum argued, than large, abrupt, and politically controversial policy changes. This approach was applied most prominently to the study of federal budgeting.^[2] In the adaptation of incrementalism to federal budgeting, Wildavsky emphasized the need for consensus as a source of budgetary instrumentalism. To avoid opening up old political struggles, Wildavsky argued that politicians would use previous budgets as a baseline and only modify budgets slowly year-to-year. The result was a political consensus on an incrementally changing budgetary process, Wildavsky contended.

Non-Incrementalism

While the incrementalist school of policy change was popular from the time of Lindblum’s seminal article, many policy adoption scholars disagree that incrementalism could fully capture the dynamism inherent in policy change. Notably, Schulman^[3] published a famous critique of incrementalism that attacked Lindblum’s theory using counterexample. Convincingly, Schulman argued that an incremental theory could not account for many important policy areas where one could not incrementally build up a program. While one may be able to incrementally start a social welfare program (say by having pilot programs with steadily increasing scope), one could not, Schulman pointed out, build a space program incrementally. The space program was one where there had to be an initially large, non-incremental investment for the program to ever start. Schulman argued that policies that are essentially indivisible (one cannot, for example, have 12% of the space program—you pretty much have one or do not have one) will be characterized by non-incremental policymaking.

The Schulman critique of incrementalism was characteristic of the dispute between supporters of the incrementalist and non-incrementalist perspectives. The supporters of incrementalism would contend that they identified a tendency in the political process. The detractors would point to counterexamples to illustrate the non-universality of the theory. Along with conceptual ambiguities in what constituted “incremental” change, the debate became bogged down and researchers lost interest in the question.^[4]

THE EMERGENCE OF THE PUNCTUATED EQUILIBRIUM THEORY

The basic conflict between incrementalist and non-incrementalist theories was easy to identify. Incrementalists were right in that the predominant mode of policy change was slow and deliberate. However, non-incrementalists were correct in that many (most even) important policy changes were non-incremental. This left incremental theory incapable of explaining many important policies. What was needed was a theory that accounted for the common experiences with incrementalism while still allowing for occasional non-incremental policy change.

The solution to the problem came in the form of a book, written by Baumgartner and Jones, *Agendas and Instability in American Politics*.^[5] In their study of agenda setting in American politics, Baumgartner and Jones had seen a common pattern emerge. Attention to policy issues seemed to change slowly most of the time. Occasionally, however, there would be dramatic episodes of large change.

The best example of this was the attention to issues surrounding nuclear power. Attention to nuclear power issues changed slowly over time through most of the 20th century. One notable exception to this pattern of incremental evolution of attention was the Three Mile Island incident. Immediately following the incident, attention to nuclear power issues shot up. This was clearly an episode of non-incremental change in attention. Baumgartner and Jones identified similar non-incremental shifts in attention in such issues as pesticide controls.

To explain the mixed pattern they observed, Baumgartner and Jones borrowed a popular model of evolutionary change proposed by Eldredge and Gould.^[6] In punctuated equilibrium models of evolution, one does not expect to see gradual change over time in the fossil record. Instead, punctuated equilibria suggested that one should find long periods of stability (especially in the phenotypic characteristics of species) rarely interrupted by periods of abrupt change. These periods of abrupt change were called “punctuations,” while the periods of stability were seen as “equilibria.”

Baumgartner and Jones adapted this approach for studying change processes. They argued that agenda setting processes produced mechanisms of negative and positive feedback. When the mechanisms of negative feedback are dominant, as is the case in most situations, institutions militate against large changes in attention and policy. In these periods, policymaking proceeds through the traditional policy subsystems rather than broad public participation. The denial of tools to expand the conflict makes it difficult for actors to substantially change the status quo.

However, events can create mechanisms of positive rather than negative feedback. The example of Three Mile Island was again instructive. The public nature of the event expanded the conflict to include almost everyone in the country. The policy subsystem was broken wide open by the broad attention generated by the Three Mile Island event. This set in motion a series of positive feedback loops. The more people were paying attention to the issue, the more media and political attention was brought to the policy area. The more media and political attention, the more stable the public attention. The broad attention fed itself. More and more people were aware of nuclear power issues. Soon it became one of the leading issues of the day, not long removed from its previous state of agenda obscurity.

The combination of positive and negative feedback mechanisms in the policymaking system, Baumgartner and Jones argued, created the characteristic pattern of punctuated equilibrium theory—long periods of stasis punctuated with short periods of abrupt, large change. This was the pattern that had caused so many disputes between the incrementalists and the non-incrementalists. The incrementalists could point to the long periods of stasis as evidence of their approach. The non-incrementalists could point to the episodes of punctuation as evidence of the limitations of the incremental approach. Punctuated equilibrium promised the possibility of integrating the incrementalist theories (to explain negative feedback processes) and the non-incremental theories (to explain the positive feedback processes) in one encompassing theory.

ADVANCES IN PUNCTUATED EQUILIBRIUM THEORY

The first step in proving the importance of punctuated equilibrium theory was to demonstrate the pervasiveness and importance of punctuated equilibrium processes. While the original Baumgartner and Jones text^[5] had identified some examples of punctuated equilibria, it had not demonstrated how common these processes were. The disputes between incrementalists and non-incrementalists had suggested that the pattern may be common—even universal—but a general

demonstration was needed. Given the centrality of budgeting to the dispute between incrementalists and non-incrementalists, it was only natural to start by looking at the distribution of budgetary change.

True, Jones, and Baumgartner^[7] provided a demonstration that punctuated equilibrium described the actual distribution of U.S. federal budgetary changes better than did traditional incremental or rational choice explanations. What they found when they looked at the distribution of federal budget appropriations was a characteristically leptokurtic distribution—i.e., a distribution characterized by a large number of observations at its peak (in this case, around a slow growth value), a large number of observations of large change (in both the positive and negative direction), and a smaller than expected number of moderate changes. The distribution of federal budget appropriations indicated that small and large changes were easier to make than would have been expected, while moderate change was harder than expected. This is exactly the descriptive prediction of the punctuated equilibrium theory. Long periods of stasis would produce a large number of observations of small change. The rare, but dramatic, punctuations would show up as an unexpectedly large number of large changes. These unexpected observations (unexpectedly common small and large changes) would come at the cost of reducing the frequency of moderate changes. The analysis of U.S. federal appropriations budget convincingly demonstrated the existence of punctuated equilibrium.

Recent work in the punctuated equilibrium tradition has moved from the demonstration of the existence of punctuated equilibria to testing of the causes of punctuated equilibria. Jones, Sulkin, and Larsen^[8] showed that punctuated equilibrium theory could provide the basis for predictive hypothesis testing. They found that the stages of the policymaking process grew increasingly “punctuated” from agenda setting to policy budgeting. The degree of punctuation seemed to increase as one moved from settings like popular media and popular elections to highly institutionalized setting like legislative budgeting. The authors contended that this pattern indicated an increase in institutional friction as one moved through the policy process. Other authors found that factors such as the nature of the policy activity^[9] or bureaucratization of the policymaking institutions also affected the degree of punctuation.^[10] This research reassured scholars that punctuated equilibrium was a theory with predictive power and a subject worthy of continued investigation.

THE FUTURE OF PUNCTUATED EQUILIBRIUM RESEARCH

The future looks bright for research into punctuated equilibrium models of policy change. The recent

attention to hypothesis testing has opened up new avenues of research and theory development.

The most obvious recent development is the broadening of the scope of punctuated equilibrium research. While previous research had been largely limited to U.S. federal data^[7,8] (with some forays into local budgets,^[9,10]) recent research increasingly deals with policymaking institutions in other areas. Punctuated equilibrium patterns at other levels of policymaking—including international comparisons^[11] and comparisons between states in the U.S. system—have been investigated in a recent work.^[12] Expanding the research into new institutional contexts has allowed for the investigation of institutional contributions to punctuated equilibrium in ways that the federal budgeting studies have not been able. International studies have been able to investigate the impact of parliamentary systems of punctuated equilibrium,^[11] while the studies in U.S. states have allowed investigators to consider the role of federal specialization on punctuated policymaking.^[12]

CONCLUSIONS

More work remains to be done to better understand punctuated equilibrium theory. First, work remains to uncover the causes of punctuated equilibrium. A great amount of work is still to be done to help us understand why some policy processes seem more prone to punctuated equilibria than other processes. Suggestions like “institutional friction”^[8] help us pin down the exact causes—though this is only a beginning. There are also methodological challenges ahead. Most of the studies have heretofore compared samples of policy outputs rather than individual policy outputs. This limits research substantially because the comparisons tend, then, to be discrete and univariate. Recent work has sought to provide a system to study punctuated policy processes at the individual input unit of analysis, but these are only the first steps at multivariate analyses.^[13]

Punctuated equilibrium theory has proven to be a useful way to resolve the disputes between incremental and non-incremental theory. While it is still a new theory, the existing work leads to the suggestion that it is a promising tool in understanding policy change process.

ARTICLES OF FURTHER INTEREST

Decision Making, Incrementalism and Transformational Change, p. 514.

Decision Making, Open System and Nonequilibrium, p. 519.

Incrementalism, Budgetary, p. 992.

Models of the Policy Process, p. 1260.
Policy Change, p. 1473.
Public Budgeting, Politics in, p. 1605.

REFERENCES

1. Lindblum, C.E. The science of muddling through. *Public Admin. Rev.* **1959**, 19, 79–88.
2. Wildavsky, A.B. *The Politics of the Budgetary Process*; Little Brown: Boston, MA, 1964.
3. Schulman, P. Non-incremental policymaking: notes toward an alternative paradigm. *Am. Polit. Sci. Rev.* **1975**, 69 (4), 1354–1370.
4. Berry, W.D. The confusing case of budgetary incrementalism: too many definitions for a single concept. *J. Polit.* **1990**, 52, 167–196.
5. Baumgartner, F.R.; Bryan, D.J. *Agendas and Instability in American Politics*; University of Chicago Press: Chicago, IL, 1993.
6. Eldredge, N.; Gould, S.J. Punctuated equilibria: an alternative to phylogenetic gradualism. In *Models in Paleobiology*; Schopf, T.J.M., Ed.; Freeman, Cooper: San Francisco, CA, 1972; 82–115.
7. True, J.L.; Jones, B.D.; Baumgartner, F.R. Punctuated equilibrium theory: explaining stability and change in American policymaking. In *Theories of the Policy Process*; Sabatier, P.A., Ed.; Westview Press: Boulder, CO, 1999.
8. Jones, B.D.; Sulkin, T.; Larsen, H.A. Policy punctuations and American political institutions. *Am. Polit. Sci. Rev.* **2003**, 97 (1), 151–169.
9. Jordan, M.M. Punctuations and agendas: a new look at local government expenditures. *J. Policy Anal. Manage.* **2003**, 22 (3), 345–360.
10. Robinson, S.E. Punctuated equilibrium, bureaucratization, and school budgets. *Policy Stud. J.* **2004**, 32 (1), 25–39.
11. Mortensen, P. American theory and Danish practice: does punctuated equilibrium theory apply to budgeting in Denmark, National Meeting of the Midwest Political Science Association, Chicago, IL, April 7–10, 2005.
12. Koske, C.; Breunig, C. Sine waves or strokes: incrementalism, punctuations, and their consequences, National Meeting of the Midwest Political Science Association, Chicago, IL, April 7–10, 2005.
13. Robinson, S.E.; Meier, K.J.; O'Toole, L.; Caver, F. Explaining policy punctuations: a multivariate model of the punctuated equilibrium theory of public agency budgets, National Meeting of the Midwest Political Science Association, Chicago, IL, April 15–18, 2004.

Q Methodology

Dan Durning

Carl Vinson Institute of Government, University of Georgia, Athens, Georgia, U.S.A.

INTRODUCTION

Q methodology is used for the systematic study of subjectivity; that is, it provides a way to investigate empirically how an individual, separately or as part of a group, thinks about a topic or issue of interest. As Brown expressed it, Q methodology “examines the world from the internal standpoint of the individual being studied” (on page 1 of Ref.^[1]).

The name Q methodology is intended to differentiate it from the “R” methods of conventional statistics that typically measure variables such as the characteristics, attitudes, or behaviors of populations (or their samples) and seek to determine the relationships among those variables. Although Q methodology makes use of factor analysis, one of the more sophisticated R methods, it has more in common with qualitative methodologies and their quest to understand humans and their behavior in their full complexity.

This article introduces the origins of Q methodology, describes its developments, and illustrates how it has been used in public administration and public policy research. After that, the article explains both Q-sort technique and its underlying framework. It concludes by suggesting that Q methodology is attractive to some researchers for the types of insights it provides and to other researchers for philosophical reasons.

HISTORY AND DEVELOPMENT OF Q METHODOLOGY

The origins of Q methodology can be traced to a letter written in 1935 by Stephenson to the editor of *Nature*, a British journal, which suggested the value of “inverting” the matrix of observations and variables in conventional factor analysis to correlate persons rather than tests (on pages 9–11 of Ref.^[1]). He proposed that researchers begin with a population of tests (“or essays, pictures, traits or other measurable material”) that would be “measured or scaled” by a set of individuals. The intercorrelations of the measures would then be subject to conventional factor analysis. In this correlation matrix, the observations would be the tests (or other measures), and the variables would be the people who completed them. Thus, factor analysis of

this inverted matrix would identify variables (that is, persons) with similar responses to the tests or other measures.

Stephenson’s 1935 letter and subsequent articles made it clear that Q methodology had two main elements. The first, highlighted in the letter, was the technique of “inverted factor analysis.” The second element was present, but less developed, in the 1935 letter: the nature of the observations—the tests or measures submitted to the individuals for them to complete. According to Stephenson, these tests or measures differ from those used in R methodology (on page 15 of Ref.^[2]). In fact, he emphasized that data collected for R method analysis cannot be inverted for use in Q methodology. Stephenson’s later work on Q methodology, which lasted until his death in 1989, linked the measures to be used in Q methodology to communication and other theories.

At the time Stephenson proposed the Q methodology innovation, he had completed two doctorates, the first in physics at Durham University (1926) and the second in psychology at the University of London (1929). As a psychology doctoral student, he was the last graduate assistant of Charles Spearman, the inventor of factor analysis. In 1936, he was appointed Assistant Director of the Institute of Experimental Psychology at the University of Oxford, where he remained until 1948. That year he moved to take a position in the University of Chicago’s Psychology Department. In 1958, he became a faculty member at the University of Missouri’s School of Journalism, where he taught until he retired.

Stephenson’s definitive book on Q methodology, *The Study of Behavior: Q-Technique and Its Methodology*, was published in 1953. It set out systematically the theoretical foundations of Q methodology.^[2] Although Q methodology, as formulated by Stephenson, was ignored, even shunned, for decades in psychology, many of his students at Chicago, Missouri, and other universities found it to be a valuable research method, and some became the second generation of Q methodologists.^[3]

A leader among the second generation has been Steven R. Brown, professor of political science at Kent State University, who has both promulgated Q methodology tirelessly and contributed important new

insights into it as a tool for the study of subjectivity in the social sciences, especially political science. Brown's 1980 book, *Political Subjectivity: Application of Q Methodology in Political Science*, joined Stephenson's *Study of Behavior*, as an essential book on Q methodology for social scientists.

In recent years, several scholars have employed Q methodology in their public administration and public policy research and practice, and several how-to guides are available for them and others interested in carrying out Q methodology research [e.g., Refs.^[4–6]]. The utility of Q methodology can be seen in the following four examples of ways that Q methodology has been employed by researchers:

- *Investigating a major issue of interest in public administration:* Brewer et al. used Q methodology to understand better why people choose to work in the public sector.^[7]
- *Identifying the subjectivities of stakeholders in policy issues:* Steelman and Maguire are among several researchers who have employed Q methodology in recent years to clarify competing stakeholder perspectives on particular policy issues or broader issues areas.^[8] They focused on the values of stakeholders in forest policy.
- *Managing conflict over policy issues:* Some scholars have turned to Q methodology to help resolve controversial policy issues. For example, Focht used Q methodology to assist in formulating a strategy to manage conflict over the use of the Illinois River Watershed in Oklahoma.^[9]
- *Identifying policy alternatives for intractable problems:* Van Eeten showed how Q methodology was used to “recast intractable problems” so as to make them more amenable to solution. He described the efforts to determine the future policy for expanding Amsterdam's Schipol Airport.^[10]

For a summary of public administration and public policy research carried out with Q methodology, see Ref.^[5].

It should be noted that Stephenson's approach to Q methodology is not the only one. In fact, the dominant definition of Q methodology in psychology has, from its beginnings, portrayed it simply as “inverted factor analysis.” This view was promoted by Cyril Burt, a major presence in British psychology for decades (see Refs.^[1,2]). He maintained that Q technique could be carried out using the same data (from “objective” measures) collected for R studies. With that data, R factor analysis would identify correlated traits and, transposing the matrix, Q factor analysis would identify persons with related characteristics. Burt's definition of Q methodology is still found in many methodology books for psychologists and social scientists.

Q-SORT TECHNIQUE

The process of conducting a Q method study consists of three major elements: 1) constructing the measurement instrument (the Q sample) and selecting the persons to be asked to complete the measurement instrument (the P set); 2) administering the measurement instrument (Q sorting) and analyzing the completed instruments (the sorts) using factor analysis; and 3) understanding or interpreting the results of the factor analysis to identify the different subjectivities (structures of thought) about the issue or topic of interest among the individuals who completed the Q sorting.

The Q Sample and P Set

The starting point for creating the measure instrument for a Q methodological study is to identify the “subjective communicability” about the topic of interest.^[11] For example, to research individual subjectivity on a topic such as illegal immigration, the researcher identifies what is being said or written about that topic. The body of written and spoken communication on illegal immigration makes up its “concourse.”^[12] This communication appears in newspaper articles, journals, websites, blogs, and television talk shows—wherever people have an opportunity to express their opinions. It can also be found by interviewing people who have different perspectives on the issue and by listening to conversations that occur every day in life.

When collecting statements of individual opinions, beliefs, and feelings about the topic, the researcher does not edit the statements for clarity nor assume that each statement has one specific meaning that can be determined. The researcher's goal is to collect enough communications in their natural form so that all perspectives are heard and included.

From the tens, hundreds, or thousands of statements on a topic, the researcher selects a Q sample, a set of statements that are a good representation of the diversity of the concourse. The selection of representative statements is not done randomly but is a purposeful activity by the researcher to ensure that the full range of subjectivity is included in the Q sample. Often, the researcher will use an experimental design to assist in selecting the statements (see pages 186–191 of Refs.^[1,10]).

While creating the Q sort, the researcher can also select the individuals who will be asked to participate in the study. These individuals (the P set) may be selected for any number of reasons. For example, the researcher may be interested in how members of a city council view an important policy issue and will ask each member to take part in the research. Or the researcher may want to know the views of social

service professionals on the issue and will select as research participants several of them who will likely have a diversity of views.

Because Q methodology is an intensive methodology, the P set will be small. In fact, Q methodology studies can be carried out with one individual, who sorts Q sample statements from different viewpoints (or “conditions of instruction”). Alternatively, the P set might reach a hundred or more; however, most will include much fewer than a hundred persons. Whatever be the number, the purpose of Q methodology is to identify different perspectives on a topic; it does not provide information on the distribution of those perspectives among P populations.

Administering and Analyzing Q Sorts

The Q sample is the measurement instrument administered to the P set through Q sorting. The mechanics of Q sorting include a Q-sort deck, instructions for sorting statements from a particular perspective and along a specified scale, and interaction with the individual completing the sort to understand better why statements were placed as they were on the continuum.

The Q-sort deck consists of the statements from the Q sample typed one to a card and numbered randomly. The individual is provided with the deck and asked to place a specified number of statements at each point on the scale (see Fig. 1) in a way that reflects the individual's views on the topic being researched. Typically, the statements are sorted in a quasinormal distribution. The forced sort makes sure that the person will place each statement in relation to all of the other statements. During the sorting process and after, the researcher can ask the participant to explain the rationale for placing certain statements at the end points on the scale (e.g., +4, -4). This conversation assists with the researcher's efforts to understand and interpret factors.

After the Q sorting is completed, the sorts are inter-correlated and the correlation matrix is factor analyzed. The factor analysis includes rotation, either through standard procedures or by hand rotation using the judgment of the researcher. The result is the identification of factors on which people with similar views have significant loadings. These factors represent different perspectives (subjectivities) on the topic being researched.

Understanding/Interpreting the Factors

The perspective of each factor is made clear by calculating its weighted average Q sort. The average is weighted so that the views of the persons with the highest factor scores have the greatest influence on the average sort. With these weighted average sorts for each factor, it is possible to determine how each factor ranked (or scored) each statement.

Examining the average sorts for each of the factors, the researcher seeks to understand the viewpoint of each factor. The researcher pays particular attention to the statements put at the end points of the continuum (e.g., +4 and -4). Also, the researcher identifies statements for which the scores of different factors diverge significantly and those for which they have similar positions. Free software (PQ Method) is available to assist with all aspects of the analysis of completed Q sorts and the interpretation of the identified factors.

The researcher will not know in advance how many factors (perspectives) will be identified or what the nature of the factors will be. The subjectivities that emerge from the sorts have not been constrained nor determined a priori by the researcher.

METHODOLOGICAL FRAMEWORK

In creating Q methodology, Stephenson's main innovation was not using factor analysis to “measure persons” instead of tests, but rather his stance that inner experience (individual subjectivity) could be measured and analyzed by the same scientific methods used to measure externally observed behavior.^[2] He insisted that the statistical procedures that correlate traits across populations could be employed to gain insight into the thoughts, opinions, and feelings of one individual. Stephenson maintained that subjectivity (thoughts, attitudes, and feelings) is behavior as real as any physical act, and he devised Q technique to make subjectivity manifest and to analyze it.

The evolution of Q methodology from its focus on single-case psychological studies to multi-individual research in other disciplines has accelerated in the past 25 yr. Part of the acceleration came as a result of further development of the theory behind the construction of measures to be used in Q-sort technique. In the 1970s, Stephenson linked these measures to his

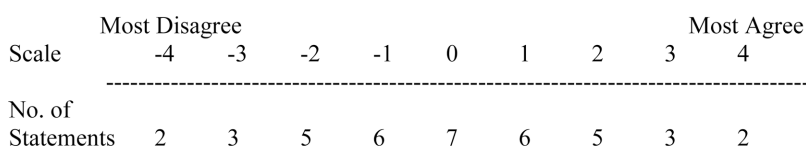


Fig. 1 Example of Q sorting scale and distribution.

concourse theory, which suggests that the universe of subjectivities on a particular issue or topic can be found in the flow of communication about it.^[12]

Just as R methodology rests on a foundation of assumptions about the nature of the knowledge, Q methodology has its own foundation. R methodology is employed to provide an objective analysis of an issue or topic. It is founded on logical positivism, in which the researcher is an outside observer with a privileged (maybe superior) vantage point from which the researcher can observe without affecting the subjects. The researcher seeks to obtain—or approach as closely as possible—objective truth. In contrast, Q methodology is closely related to the postpositivist perspective that rejects the possibility of observer objectivity and denies that the observer occupies a privileged position.^[13] It posits the existence of multiple realities and seeks to measure them. Thus, Q methodology is consistent with phenomenological, hermeneutic, and quantum theories.^[14,15]

Reflecting the Q methodology framework, the work of the researcher in a Q methodology study is visible and explicit: The researcher purposefully selects the Q sample and P set, may hand rotate the factor analysis using insights gathered during the research process, and interprets the factors to understand the different subjectivities. However, the purpose of the involvement is simply to enable individuals to express their subjectivities freely and to assist the researcher to understand those subjectivities.

CONCLUSION

In the past 15 yr, Q methodology has been utilized by researchers in growing number of fields and disciplines, including public administration and public policy.^[5] Part of Q methodology's attraction is that it provides information not accessible through R methods. Some researchers find it attractive because of its philosophical stance, including its postpositivist foundation, and have suggested that Q methodology should be taught as a research method in policy schools.^[16]

At present, Q methodology does not have a place in most public administration and public policy methods courses. However, its increasing use by public administration and policy researchers shows its potential contributions to both disciplines.

REFERENCES

1. Brown, S.R. *Political Subjectivity: Applications of Q Methodology in Political Science*; Yale University Press: New Haven, CT, 1980.
2. Stephenson, W. *The Study of Behavior: Q-technique and its Methodology*; University of Chicago Press: Chicago, IL, 1953.
3. Brown, S.R., Brenner, D.J., Eds.; *Science, Psychology, and Communication: Essays Honoring William Stephenson*; Teachers College Press: New York, 1972.
4. McKeown, B.F.; Thomas, D.B. *Q Methodology*; Sage: Newbury Park, CA, 1988.
5. Brown, S.R.; Durning, D.W.; Selden, S.C. Q methodology. In *Handbook of Research Methods in Public Administration, Public Administration and Public Policy Series*; Miller, G.J., Whicker, M.L., Eds.; Marcel Dekker: New York, 1999; Vol. 71, 599–637.
6. Durning, D.W.; Brown, S.R. Q methodology and decision making. In *Handbook of Decision Making*; Morçöl, G., Ed.; CRC Taylor and Francis: Boca Raton, FL, 2007; 537–563.
7. Brewer, G.A.; Selden, S.C.; Facer, R.L. Individual conceptions of public service motivation. *Public Admin. Rev.* **2000**, *60* (3), 254–264.
8. Steelman, T.A.; Maguire, L.A. Understanding participant perspectives: Q-methodology in national forest management. *J. Policy Anal. Manage.* **1999**, *18* (3), 361–388.
9. Focht, W. Assessment and management of policy conflict in the Illinois River Watershed in Oklahoma: an application of Q methodology. *Int. J. Public Admin.* **2002**, *25*, 1311–1349.
10. van Eeten, M.J.G. Recasting intractable policy issues: the wider implications of the Netherlands civil aviation controversy. *J. Policy Anal. Manage.* **2001**, *20* (3), 391–414.
11. Stephenson, W. Consciring: a general theory for subjective communicability. In *Communication Yearbook*; Nimmo, D., Ed.; Transaction: New Brunswick, NJ, 1980; Vol.4, 7–36.
12. Stephenson, W. Concourse theory of communication. *Communication* **1978**, *3*, 21–40.
13. Dryzek, J.S. The measurement of political man—and woman. In *Discursive Democracy*; Cambridge University Press: Cambridge, UK, 1990; 173–189.
14. McKeown, B. Circles: Q methodology and hermeneutical science. *Operant Subjectivity* **1998**, *21*, 112–138.
15. Stephenson, W. Quantum theory and Q-methodology: fictionalistic and probabilistic theories conjoined. *Psych. Rec.* **1983**, *33*, 213–230.
16. Durning, D. The transition from traditional to post-positivist policy analysis: a role for Q-methodology. *J. Policy Anal. Manage.* **1999**, *18* (3), 389–410.

Quantitative Decision Making: Benefit–Cost Analysis

Jay P. Hamilton

*Department of Public Management, John Jay College of Criminal Justice,
New York, New York, U.S.A.*

INTRODUCTION

Benefit–cost analysis (or cost–benefit analysis) is a quantitative decision-making tool, used by policy makers and public administrators to assess the efficiency of public policy. Benefit–cost analysis uses a single metric to determine the net benefits of a policy or project for a well-defined group of people. The policy’s consequences are measured in currency units. For the United States, the dollar value of all benefits accruing to all people in the well-defined group is compared to the dollar value of all costs accruing to all people in the well-defined group. If the benefits exceed the costs (positive net benefits), then the policy maker or public administrator can use the analysis to justify proceeding with the project. If the policy maker rejects implementation of a project with positive net benefits or proceeds with a project that fails to produce positive net benefits, then the decision maker must justify their decision in light of the empirical evidence. Benefit–cost analysis helps decision makers find an optimal solution when deciding between different projects in which there are several proposed options. In this situation, benefit–cost analysis informs the decision maker about which projects return the greatest net benefits or the greatest rate of return.

OVERVIEW

Benefit–cost analysis has been applied to many areas of public decision making. Major public works such as dams, levees, and roadways are frequently subject to benefit–cost analysis prior to construction.^[1] Benefit–cost analysis has been applied to various health-care policy discussions, such as drug abuse prevention programs.^[2] Social regulations designed to protect the environment and worker safety are commonly, but not always, subject to benefit–cost analysis.^[3] In the field of criminal justice, benefit–cost analysis has been applied to evaluate the efficiency of early childhood intervention programs, such as preschool for the disadvantaged, in preventing future crime.^[4] Economic regulations such as minimum wages and tariffs are also subjects of benefit–cost analysis.^[5] In principle, any policy or public decision can be subject to benefit–cost analysis.

Controversy surrounds benefit–cost analysis in its theoretical underpinnings and in its application. Although benefit–cost analysis appears to be objective, its theoretical foundations are based on at least two value judgments.^[6] First, society’s welfare is derived from individual preferences. Thus, benefit–cost analysis is the sum of the net benefits received by all individuals affected by the policy or project. Second, the importance of individuals’ preferences is not equal, because money is used as the metric of comparison. Individuals with greater wealth have a greater influence on the valuation of benefits and costs. In practice, benefit–cost analysis can be reweighed to compensate for the wealth bias, but the reweighing depends on a value judgment made by the analyst or the decision maker. Those who support benefit–cost analysis argue that all public decision making will involve some implicit or explicit value judgments, but a thoroughly documented benefit–cost analysis will highlight the sensitivity of any value judgments and will produce a quantified report with details that can be scrutinized.

The controversies of benefit–cost analysis application are numerous and varied. Five categories of concerns are discussed here, after a brief history of the development of benefit–cost analysis. The first area of controversy is defining the questions the analysis will help answer. The decision maker and analyst must agree on the relevant sphere of influence of the policy, as well as the depth to which the analyst will look for consequences of the policy. Once consequences are identified, the second area of controversy is the valuation of those consequences. The dollar estimation of intangible values necessarily involves many methodological choices for the analyst. Third, the analyst must take into consideration the time frame of the consequences of the policy, which implies a choice of discount rate for future benefits and costs. Fourth, the analysis must deal with issues of uncertainty. Last, the decision maker must consider the impact of the distribution of benefits and costs across different segments of society. While the principle of benefit–cost analysis is simple, the application is often intricate. All practitioners of benefit–cost analysis agree that final reports should clearly state assumptions and should contain sensitivity testing of those assumptions.

THE ORIGINS OF BENEFIT–COST ANALYSIS

The origin of cost–benefit analysis is widely attributed to the United States Army Corps of Engineers, which took the first steps toward formulating a consistent benefit–cost analysis methodology in the 1930s. Charged with the task of evaluating various improvements in the nation’s waterways, the Corps generated reports estimating the costs required for each project. The Flood Control Act of 1936 authorized the Department of Agriculture to estimate projects’ benefits “to whomever they accrue” and compare these benefits to estimates of the projects’ costs. After 1936, the history of benefit–cost analysis can be divided into two categories: 1) the search for theoretical foundations by welfare economists and 2) improvements in the application of benefit–cost analysis.

The study of welfare economics has long been dominated by the concept of “Pareto Optimality” that states: A distribution of outcomes/benefits is deemed efficient if no individual or group can be made better off without making even one person worse off. The Pareto Optimality criterion is relatively uncontroversial but also largely inapplicable to public decision making, because nearly every public decision creates winners and losers. The Pareto criterion cannot justify the use of benefit–cost analysis. In 1939, Hicks and Kaldor, working independently, crafted a “modification” to the Pareto criterion, which became known as the Kaldor–Hicks compensation criterion.^[7,8] Under this rubric, a policy will improve public welfare, if the gains to the winners are large enough such that the winners can compensate the losers and still be better off than before the policy. The Kaldor–Hicks compensation criterion is often considered a utilitarian approach to policy making. Critics of the Kaldor–Hicks compensation criterion are quick to point out that compensation does not need to be paid, and, thus, a policy that passes the test may give all the gains to one person and create many losers.^[9]

To facilitate the practice of benefit–cost analysis, agencies using this decision-making tool began publishing guidelines in the late 1930s. In the 1950s, the Rand Corporation also refined the analysis process when drafting rules to guide decision makers in selecting military objectives.^[10] Around the same time, health economists began using cost-of-illness as well as cost–benefit analyses to evaluate various health policies. Over the 1960s and 1970s, the use of benefit–cost analysis spread throughout other U.S. government agencies and overseas to Great Britain, the Organization for Economic Co-operation and Development, United Nations Industrial Development Organization, and the World Bank. In the 1980s and 1990s, the use of benefit–cost analysis spread to environmental, safety, and health regulations. Although most presidential

administrations advocate the use of benefit–cost analysis, not all legislation has embraced this decision-making tool. Most notably, the U.S. Clean Air Act of 1970 makes no mention of the need to evaluate the costs of regulations drafted in support of the act. The courts have upheld the practice of ignoring the costs of clean air regulations. More recently, the Job Creation and Wage Enhancement Act (H.R. 9, 1995), which passed the House of Representatives but failed to reach the Senate floor in the 104th Congress, mandated the use of benefit–cost analysis in all new regulations issued by the federal government.

DEFINING THE ANALYSIS

The first step in conducting a benefit–cost analysis is defining the decision the analysis will evaluate. Sometimes the question at hand is simple, such as whether or not to dredge a harbor. Other decisions are more complex, creating an array of options for the decision maker to choose from. For example, the decision to adopt a new safety regulation for cars may involve deciding how quickly to implement the new safety rule. Should the new device be phased in or be adopted immediately? Immediate adoption may be costly compared to a gradual introduction; if the benefits mostly accrue far into the future, then a gradual approach would be appropriate. The analyst should help the decision maker determine how quickly to implement the new regulations. As a second example, consider the decision to build a flood control levee. Benefit–cost analysis should not simply give an up or down answer; the final report should present a range of conclusions based on different heights for the levee. The more complicated the set of questions the benefit–cost analysis is called upon to answer, the more complicated the final report.

Once the appropriate project is identified, the analyst must construct a list of all the consequences of the decision. This includes a somewhat arbitrary decision about the relevant sphere of influence for the decision maker. Does the decision maker mandate the inclusion of benefits and costs that accrue only to their constituency? Or is there a greater sense of public? For example, a regulation in the United States reducing the emission of sulfur oxides from coal-burning power plants will reduce the incidence of acid rain in Canada. The decision maker must decide if the benefits to Canadians should be included. A good benefit–cost report would compare both options so that the policy maker can make the final decision. But as the number of such issues increases, so will the complexity of the final report. An intractably complicated report containing a myriad of conclusions defeats the lucidity of benefit–cost analysis that makes this quantitative tool so attractive. The analyst should

open this discovery phase to the public and seek a wide range of opinions about the sphere of influence and the potentially relevant consequences.

VALUATION

The analyst must attempt to assign a dollar value to all costs and benefits for each consequence of the policy; if the analyst is unable to do this, then the exceptions must be clearly stated. Many costs and benefits will be easily estimated from existing market data. The direct cost of constructing a dam is readily estimated from previous dam constructions, as are the benefits from power generation, irrigation, or flood control. However, many benefits and costs are not easy to estimate. In the case of dam construction, examples of negative consequences that are difficult to assign dollar values include habitat loss for native animal and plant species, addition of greenhouse gases from rotting vegetation in the reservoir, and lost recreational value. Among the intangible benefits where value is hard to assess are habitat saved from flooding, lives saved, and increased recreational value on the reservoir. Valuation estimation of non-market benefits and costs often occupies much of the analyst's time and creates much of the strongest debate.

There are many established techniques used to value non-market benefits and costs; all of them are anthropocentric, meaning that only those benefits and costs that accrue to humans are included in benefit–cost analysis. One such technique, Hedonic valuation, decomposes the market to extrapolate values for non-market goods. For example, the external costs of industrial pollution can be estimated by the reduction in residential land values surrounding the polluter. Therefore, an environmental regulation that reduces pollution will have a benefit of increased property value. There are other markets that can be dissected by Hedonic valuation. Analysts can examine differences in pay based on occupational hazards, to determine the value of safe working conditions. Another popular application of Hedonic valuation is the travel cost method, where a natural amenity with free access is valued by the money people spend to reach the amenity.

Some intangible values can be inferred from jury awards and political decrees. If benefit–cost analysis is used for a policy that saves lives or potentially costs lives, then the value of a human life is a critical item that must be valued carefully. The analyst may choose to estimate the value of human life by considering the damages awarded in wrongful death lawsuits. In practice, this involves estimating the future earnings of the deceased. The value of a person is equivalent to their value in the labor market. Also known as the human capital approach, this method of valuation

suffers from labor market imperfections such as diminished wages for women and minorities.^[11] The body politic may also provide financial compensation for intangible losses such as those that occur when a person is wrongfully imprisoned.

The benefits of a policy can be estimated by using the private expenditures made by individuals in their efforts to avoid or diminish a problem that will be solved by the policy. This approach is commonly referred to as the aversion-cost approach. For example, a private citizen who lives next to a freeway may construct a large wall to diminish freeway noise pollution. When considering new freeway projects, the cost of noise pollution may be estimated by extrapolating the costs private citizens pay to reduce the noise.

Another approach to evaluating benefits is contingent valuation. This method utilizes a scientific survey that asks people to place a value on policy issues. Contingent valuation should attempt to frame the questions in a meaningful and non-biased manner, so that respondents will answer honestly. Many contingent valuation surveys will propose simple yes or no questions, such as, “is preserving forest X worth an additional cost of \$75 per property per year in property taxes.” Different respondents are given different amounts. The analyst then uses statistical analysis to estimate the average willingness to pay for the potential benefit. The distribution of wealth will influence the estimates produced by contingent valuation surveys, because respondents' willingness to pay for a policy is a function of income and wealth.

Consider a policy designed to reduce crime. The analyst must estimate the intangible costs of crime imposed directly on victims and indirectly through fear created in the general population. This is necessary because the elimination of those costs is the main benefit of crime prevention. There are several methodological approaches the analyst can use. The analyst may choose the Hedonic valuation method to statistically isolate the component of property values attributable to ambient crime rates. However, the analyst must be careful to control for other factors that influence property rates, such as the prevalence of low-wage incomes, proximity to polluters, disappointing school districts, etc. Not all of these data will be easy to obtain. The analyst may choose to extrapolate the costs of crime from jury awards connected to past crime. But, this methodology takes data from a jury award given to a specific victim and applies the data to the valuation of crimes prevented for unknown persons. Because of the face-to-face nature of jury trials, they are likely to award penalties far in excess of what society would be willing to pay to prevent the same crime. The analyst may choose to apply an aversion-cost approach. The analyst could estimate the benefits of increased law enforcement by calculating

Table 1 Present value of a \$1,000,000 future value

Years into the future	Discount rate						
	00.0%	00.5%	01.0%	05.0%	10.0%	15.0%	25.0%
1	\$1,000,000	\$995,025	\$990,099	\$952,381	\$909,091	\$869,565	\$800,000
10	\$1,000,000	\$951,348	\$905,287	\$613,913	\$385,543	\$247,185	\$107,374
20	\$1,000,000	\$905,063	\$819,544	\$376,889	\$148,644	\$61,100	\$11,529
50	\$1,000,000	\$779,286	\$608,039	\$87,204	\$8519	\$923	\$14
75	\$1,000,000	\$687,932	\$474,129	\$25,752	\$786	\$28	\$0
100	\$1,000,000	\$607,287	\$369,711	\$7604	\$73	\$1	\$0

the money spent on private security measures such as barred windows, locks, alarms, gun sales, and private security services. Finally, contingent valuation surveys may be the best way to estimate society's willingness to pay for the prevention of certain crimes, but surveys are costly and time consuming.

DISCOUNTING FUTURE BENEFITS AND COSTS

There are several aspects of benefit–cost analysis of special concern to the policy maker. Many of the benefits and costs of decisions made today will have an impact several years in the future. As a general rule, people do not value the future as much as they value the present. Therefore, any future benefits or costs should not be considered the same as current costs and benefits. Future costs and benefits are discounted. The rate at which future benefits and costs are discounted is a question of considerable debate. Future benefits and costs are discounted according to the following formula:

$$PV = FV/(1 + r)^t,$$

where

PV = the present or discounted value to be included in the sum of benefits and costs,

FV = the estimated future value of the benefit or cost,

r = the discount rate per time period, and

t = the number of time periods into the future until the benefit is realized.

The future values should be estimated in current dollars to avoid inflation complications. As t increases, the impact of future benefits and costs on present value is reduced because of the compounding nature of the formula. Table 1 demonstrates the impact of discounting in biasing benefit–cost analysis against preferences of individuals in the future. As we go down any column with a positive discount rate, we see that the present value of a one-million-dollar future value gets smaller the farther we go into the future.

Many benefit–cost predictions hinge on the value of the discount rate, r . For example, a program to store nuclear waste will have large costs today, but the benefits will be felt for thousands of years. If a low value of r is chosen, then the program has a better chance of passing the benefit–cost test than if a high value is chosen. A good report will present a range of present values that use several different discount rates. Table 1 demonstrates the powerful impact of the choice of discount rate. As we go across any row, the discount rate increases, and the present value for a one-million-dollar future value gets smaller.

Some people argue that the discount rate should be set at the prevailing interest rate, because society can capture the benefits of some future consequence today, by taking out a loan with the future benefit as the eventual amount of the loan repayment including interest charges. However, such a practice makes the analysis subject to the whims of financial markets that may have little relationship to society's time preference. For example, few financial instruments extend beyond a 30-yr maturity, but many public works will last much longer. However, interest rates may be the correct choice for costs, because the present value represents the amount of money that would need to be set aside and invested today so that compound interest would provide enough funding in the future. Some analysts prefer to conduct an internal rate of return analysis, where the objective is to find the discount rate, where the sum of the present value of benefits exactly equals the sum of the present value of costs, and, thus, the decision maker simply decides if the resulting discount rate accurately reflects society's rate of time preference. Some people argue that the discount rate should be set to zero, making future benefits and costs equivalent to today's benefits and costs.

UNCERTAINTY

Many of the consequences resulting from public decisions are not certain. To the extent that benefits and

costs are uncertain, quantitative estimates of those benefits and costs must be adjusted to incorporate the uncertainty. Analysts commonly estimate the riskiness of the consequences. The analyst will, using the best information available, assign a probability to every possible outcome. If a policy is estimated to have two potential outcomes, say a one-million-dollar net benefit occurring with a 10% chance or a zero net benefit occurring with a 90% chance, it would be tempting to simply say the expected net benefit is \$100,000, the mathematical expectation. However, people are generally risk averse, meaning they would rather take a certain net benefit less than \$100,000 instead of the gamble on the risky policy. It may be appropriate to use this certainty equivalent in the summation of benefits and costs. But, the value of a certainty equivalent depends on society's degree of risk aversion, which is difficult to estimate.

The analyst may account for the degree of risk aversion by adjusting the discount rate in the discounting formula applied to the mathematically expected future values.^[12] For risky benefits in the future, the discount rate, r , should be increased, and, for risky costs in the future, the discount rate should be decreased. Increasing r makes the present value smaller to reflect the aversion to risk; the risky future benefit is not worth as much to us today as is a certain future benefit. The analyst should be clear about any risk adjustments to the discount rate. Sensitivity analysis should also be presented so the decision maker can evaluate the impact of adjusting for risk.

It is important to note that incorporating risk does not fully account for the problems of an uncertain future. Risk is a particular subset of uncertainty, where all possible outcomes are known and all of those potential outcomes have a known probability of occurring. However, for most public decisions, there are many outcomes that cannot possibly be known beforehand. Even if all possible outcomes are known, we may have to estimate the probability of each occurrence. Because of these unknown possibilities, benefit–cost analysis may recommend a project that, in hindsight, should not have been undertaken.

DISTRIBUTION CONSIDERATIONS

By using the dollar metric, benefit–cost analysis implicitly weights the estimation of benefits and costs among individuals, based on their wealth. An unadjusted benefit–cost analysis assumes that the distribution of wealth is optimal. If the distribution of wealth is not deemed to be optimal, then the benefits and costs of the policy can be reweighed to compensate for any wealth bias. Reweighing can be based on the overt statements of elected officials, or the new weights can

be estimated in an attempt to align “economic votes” with democratic votes, in which each individual counts the same. In the former case, politicians may clearly state that the preferences of the poor or disadvantaged minorities should be valued more than others. In the latter case, benefits and costs accruing to the wealthy are discounted. The final report should be clear about the distributional aspects of the policy. The analyst should conduct and report on the sensitivity of the reweighing procedure.

CONCLUSION

Benefit–cost analysis is not cheap, easy, or quick. If analysts try to track and quantify every possible consequence, they will never finish. Implicitly, every analyst must conduct a benefit–cost analysis of the benefit–cost analysis. Additional information may not be worth the price of obtaining it or of delaying the report. It is also important to recognize the possibility that benefit–cost analysis can be requested by policy makers to delay policy implementation rather than ascertain the policy's efficiency.

Benefit–cost analysis can successfully aid public decision makers when assumptions and sensitive methodological choices are clearly stated. As benefit–cost analysis gains popularity with policy makers, those affected, such as public administrators, public advocates, and concerned citizen groups, will increasingly need to understand the nuances of this and other quantitative decision-making aids.

ARTICLES OF FURTHER INTEREST

- Budgetary Analysis and Economics*, p. 177.
- Capital Programming and Budgeting: Comparative Local Government Perspectives*, p. 270.
- Decision Making, Rational*, p. 524.
- Economic Efficiency*, p. 614.
- Time Value of Money*, p. 1923.
- Uncertainty in Decision Making*, p. 1984.
- Using Econometric Models for Public-Sector Decision Making*, p. 2017.
- Welfare Economics*, p. 2047.

REFERENCES

1. Nash, C. Cost–benefit analysis of transport projects. In *Efficiency in the Public Sector: The Theory and Practice of Cost-Benefit Analysis*; Williams, A., Giardina, E., Eds.; Edward Elgar Publishing Co.: Brookfield, VT, 1993; 83–105.

2. Pentz, M. Costs, benefits, and cost-effectiveness of comprehensive drug abuse prevention. In *Cost–Benefit/Cost-Effectiveness Research of Drug Abuse Prevention: Implications for Programming and Policy*; Bukaoski, W., Evans, R., Eds.; National Institute on Drug Abuse: Rockville, MD, 1998; 111–129.
3. Pearce, W.; Turner, R. *Economics of Natural Resources and the Environment*; Johns Hopkins University Press: Baltimore, MD, 1990; 1–10.
4. Greenwood, P.; Karoly, L.; Everingham, S.; Houbé, J.; Killburn, M.; Rydell, C.; Sanders, M.; Chiesa, J. Estimating the costs and benefits of early childhood interventions: nurse home visits and the Perry preschool. In *Costs and Benefits of Preventing Crime*; Welsh, B., Farrington, D., Sherman, L., Eds.; Westview Press: Boulder, CO, 2001; 123–148.
5. Gramlich, E. Tax expenditures. In *A Guide to Benefit–Cost Analysis*; Prentice-Hall, Inc.: Englewood Cliffs, NJ, 1990; 185–196.
6. Pearce, W. The foundations of cost–benefit analysis. In *Cost–Benefit Analysis*; St. Martin's Press: New York, 1983; 1–13.
7. Hicks, J. Foundations of welfare economics. *Econ. J.* **1939**, *49*, 696–713.
8. Kaldor, N. Welfare propositions of economics and interpersonal comparisons of utility. *Econ. J.* **1939**, *49*, 549–552.
9. Brent, R. *Applied Cost–Benefit Analysis*; Edward Elgar Publishing, Inc.: Lyme, NH, 1997; 27–49.
10. Hitch, C.; McKean, R. *The Economics of Defense in the Nuclear Age*; Harvard University Press: Cambridge, MA, 1961.
11. Fuguitt, D.; Wilcox, S. *Cost–Benefit Analysis for Public Sector Decision Makers*; Quorum Books: Westport, CT, 1999; 276–295.
12. Brent, R. Risk and uncertainty. In *Applied Cost–Benefit Analysis*; Edward Elgar Publishing, Inc.: Lyme, NH, 1997; 158–183.

Reciprocal Relations Among Peace, Prosperity, and Democracy

Stuart S. Nagel

Political Science Department, University of Illinois, Urbana, Illinois, U.S.A.

INTRODUCTION

The leading controversial policy issues in developing nations like those of South Asia tend to be economic, political, and military in nature. The economic issues relate to making domestic economies more prosperous and facilitating investments, exports, and imports. The political issues relate to promoting democratic institutions, human rights, and self-determination. The military issues relate to advocating non-proliferation of arms and reduction in regional conflicts.^[1]

The purpose of this article is to discuss some aspects of U.S. foreign policy as applied to these policy problems of developing regions and nations.^[2] The context is in terms of win-win thinking where all nations involved are better off as a result of the interaction.^[3] The context is also in terms of three tables, which give useful visual aids to clarify the mutually beneficial interaction.

TEN CAUSAL RELATIONS

The first six causal relations in [Table 1](#) can be interpreted as follows:

1. Reduction in military conflicts is conducive to prosperity and investment.
2. Prosperity is conducive to reduction in military conflicts, especially prosperity based on buying and selling across countries that might otherwise be in conflict.
3. Prosperity is conducive to democratic institutions, human rights, and tolerance of minority ethnic groups.
4. Democracy, human rights, and ethnic peace are conducive to prosperity.
5. Democratic political institutions are conducive to a reduction in military conflicts.
6. Reduction in military conflicts is conducive to democratic political institutions.

The last four causal relations can be interpreted as follows:

7. U.S. policy is concerned with encouraging prosperity, investment, exporting, and importing because doing so is mutually beneficial.

8. U.S. policy is concerned with reducing military conflicts partly because of the favorable effect on prosperity and the economic issues.
9. U.S. policy is concerned with promoting a democratic political environment partly because of the favorable effect on prosperity and the economic issues.
10. U.S. domestic economic policy emphasizes U.S. prosperity and gross national product (GNP) growth, which partly explains why U.S. international policy emphasizes mutually beneficial trade and investment opportunities.

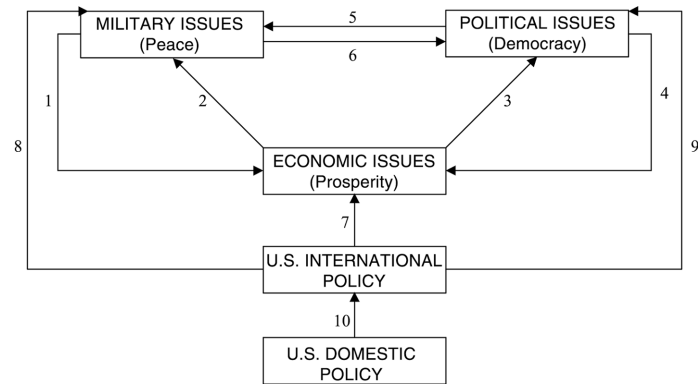
Miscellaneous points include:

1. In the context of South Asia, Pakistan is especially concerned with military security, India is especially concerned with political issues, and Sri Lanka is especially concerned with international economics, but all three countries are concerned with all three sets of issues.
2. The concept of mutual benefit is promoted through regional organizations, such as the South Asia Association for Regional Cooperation, as well as through interregional interaction between South Asia and the United States.
3. Social issues such as poverty and ethnic groups are also important, although they are considered under economic and political issues, respectively. Technology issues are also quite important, but they are discussed in the context of military, economic, and political issues.
4. There are positive relations among all five variables shown in Table 1. Those relations are positive in the sense of upward causation and being desirable relations, especially regarding the promotion of peace, prosperity, and democracy.

HISTORICAL DEVELOPMENT

The past was characterized by colonialism and the Cold War. The present is being characterized by investment, importing, and exporting of funds and foods. The future may be characterized by transfer of technologies and skills from the United States, which

Table 1 Win-win U.S. foreign policy



NOTE: All relations are positive. Thus there is no need for a plus sign on each arrow. For example, Relation 2 says prosperity causes peace because in time of depression, political leaders sometimes look for external enemies to distract people from being unemployed.

results in mutually beneficial investment—returns, buying, and selling. Those basic ideas are summarized in Table 2.

The arrows from the *past* indicate:

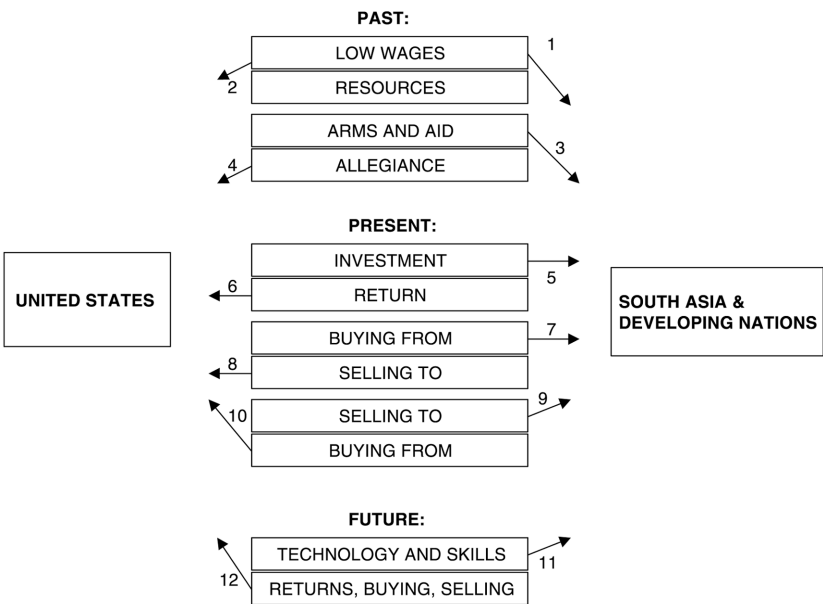
- (1 and 2) Colonialism-involved low wages going to the developing nations, and valuable resources going to the United States or other industrialized nations.

- (3 and 4) The Cold War—involved arms and aid going to developing nations, and allegiance going to the United States or the Soviet Union.

The arrows from the *present* indicate:

- (5 and 6) Capital investment going to the developing nations with a reasonable return going back to the United States.

Table 2 U.S. foreign policy for 200 years



NOTE: Unlike Table 1, the arrows here are not causal arrows. The arrows on the left side show what the U.S. has generally received from its relations with developing nations in the past, present, and future. The arrows on the right side show what developing nations have received from the U.S.

- (7 and 8) Cash or credits going to developing nations in return for their products.
- (9 and 10) Products going to developing nations in return for their cash or credits.

The arrows from the *future* indicate:

- (11 and 12) Technologies and skill go to developing nations, thereby improving their ability to be good places for investment, buying, and selling.

In the past, there was often an imbalance, with disproportionate benefits to the industrial nations and disproportionate detriments to the developing nations. In the present, there is generally mutual benefits from investments, exporting, and importing. In the future, the transfer of technologies and skills may enable all participating countries to exceed their best initial expectations simultaneously.^[4]

ILLUSTRATIVE EXAMPLES

Tables 3–5 show how win–win analysis might be applied to three key issues in South Asia. The issues relate to trade vs. aid, secession, and nuclear arms.

Trade Vs. Aid (Prosperity)

The United States currently tends to favor trade over aid because trade is more mutually beneficial than aid, which tends to mainly benefit the recipient nation unless there is a Cold War return.

Developing nations tend to favor aid with no strings attached, at least in the past, because they are fearful that^[1] buying from the United States will disrupt local industries, and that^[2] they have little to sell to the United States.

Table 3 Trade vs. aid in South Asia

Alternatives	Goals	
	Pro-United States (C)	Pro-South Asia (L)
C Mainly trade	+	–
L Self-determination	–	+
N	0	0
1. Some trade		
2. Some aid		
SOS or win–win	++	++
1. Skills transfer		
2. Technology transfer		

Table 4 Secession in South Asia (Kashmir)

Alternatives	Goals	
	Pro-United States (C)	Pro-South Asia (L)
C Self-determination	+	–
L	–	+
1. Retain rebellious provinces (pro-India)		
2. Release rebellious provinces (pro-Pakistan)		
N Partition	0	0
SOS or win–win	++	++
Autonomy like a U.S. state		

Skills and technology transfer greatly benefit the United States by virtue of improving places for 1) U.S. investment; 2) the buying of American products; and 3) the selling to the U.S. of products needed by the American people. Skills and technology transfer benefit developing nations by enabling them to upgrade their international competitiveness even more than offering them either trade or aid.

Seceding (Democracy)

The United States tends to favor self-determination out of a regard for democratic decision making, emphasizing the majority will within the rebellious provinces.

Developing nations tend to favor retaining their own rebellious provinces, emphasizing the majority will within the larger political entity.

Autonomy, like that of a U.S. state, refers to states having their own constitutions and governors that cannot be removed by Washington. States in India do not

Table 5 Nuclear arms in South Asia

Alternatives	Goals	
	Pro-United States (C)	Pro-South Asia (L)
C No nuclear arms	+	–
L Retain the capability that exists	–	+
N	0	0
1. Reduce		
2. Inspect		
SOS or win–win	++	++
Peaceful conversion		

have their own constitutions and their governors can be removed by New Delhi.

Nuclear Arms (Peace)

The United States tends to favor the removal of nuclear arms from South Asia for fear that their presence may lead to nuclear warfare, which might involve the United States directly or indirectly by way of international disruption.

Countries that have nuclear capability like India and Pakistan are reluctant to weaken their deterrent power against each other.

Peaceful conversion in this context means providing India and Pakistan with the skills and technologies for

converting their nuclear capability into peaceful and safe nuclear energy along with American investment funding.

REFERENCES

1. Stuart, N. et al. Co-effect diagrams and win-win analysis. *Policy Eval.* **2000**, 6, 9–12.
2. Stuart, N. et al. Win-win domestic peace, prosperity, and democracy. *Peace Prosper. Democr.* **2001**, 1, 35–42.
3. Francis, A. Prosperity: the political economy of foreign assistance. In *Multi-National Policy Toward Prosperity*; Stuart, N., Ed.; Kluwer, 2001.
4. Stuart, N. Win-win technology policy. *Dev. Policy Stud.* **1999**, 5, 28–37.

Reduction in Force

Teva J. Scheer

University of Colorado at Denver, Highlands Ranch, Colorado, U.S.A.

INTRODUCTION

This article discusses the requirements and procedures that typically comprise reduction in force systems (RIFs) in government employment. It also briefly reviews the results of employee appeals resulting from RIFs, and closes with a few suggestions for administering an RIF action.

WHAT IS RIF?

An RIF is a systematic process that a government agency can use to reorganize or to reduce its staffing levels. The federal government, as well as many state and local agencies, have published rules and procedures that govern how they must conduct RIFs. These procedures are generally available for review by employees, unions, and even the public, thereby demystifying the process. There are four general reasons why an agency would choose, or be required, to conduct a RIF:

1. Lack of work for the existing work force.
2. Lack of funds to pay the existing work force. Short-term funding problems may be handled using furloughs or use of voluntary leave without pay, but longer furloughs (e.g., more than 30 days) often trigger the RIF mechanism.
3. A change in the number and type of positions needed as a result of reorganization. Reorganizations include Transfers of Function, in which a work unit or department is moved to a new geographical or organizational location.
4. The need to make room for an employee who has reemployment or restoration rights.

Even though most employees fear the negative impact of an RIF, it does provide an equitable mechanism for realigning an agency or reducing its workforce. It also allows an agency and its remaining employees to “move on,” i.e., return to normal recruitment and placement practices immediately. Many agencies decide to downsize or reorganize using attrition rather than RIF, but this strategy has at least two major and negative results:

1. Agencies may be required to suspend normal promotional opportunities during the period of attrition, thereby negatively affecting employee morale and career development.
2. Staff costs will be higher during the attrition period. The agency is likely supporting too many employees, and some of the employees will be in the wrong positions, wrong occupations, and wrong pay levels.

While the specific RIF procedures vary for each government system, typically the procedures require an agency to establish a “competitive area” in which employees will compete for positions. The competitive area may be described geographically, or organizationally, or both. Each competitive area is organized into “competitive levels,” which consist of all the positions within the competitive area that are sufficiently alike in qualification requirements, duties, responsibilities, pay schedules, and working conditions so that the agency may readily assign the incumbent of one position to any of the other positions without changing the terms of the employee’s appointment or unduly interrupting work. All the employees within a competitive level are placed in order on a “retention register.” Their order is determined by such factors as tenure, performance scores, and veteran status. If positions within a competitive level are abolished, the employees who are displaced are those with the lowest scores on the retention register. Employees who are thus displaced may have rights to “bump” an employee of lower retention standing in another competitive level.

RIF APPEALS AND CASE LAW

A review of RIF case decisions issued by federal courts, the Equal Employment Opportunity Commission, and the Merit Systems Protection Board showed that a substantial majority of RIF appeals are decided in favor of agencies.^[1] A notable exception was cases where the adjudicatory authority found that the agency had failed to follow its own guidelines or had committed a technical error in administering the RIF action. Agencies should also keep in mind that RIFs are subject to the requirements of the Uniform Guidelines on Employee Selection, so agencies should

exercise extra caution in following its procedures and in documenting all aspects of the RIF if it appears that the RIF will have adverse impact on members of a particular race, sex, or ethnic group. (See Uniform Guidelines on Employee Selection.)

Some of the RIF aspects that generally have been unsuccessfully challenged by affected employees include allegations that the agency

1. Had failed to prove RIF was based on one of the reasons outlined in the agency RIF procedures (e.g., lack of work, lack of funds).
2. Had established a competitive area that was improperly sized, or that the area's composition was inappropriate (e.g., it should have included additional positions).
3. Had improperly constructed its retention registers.
4. Had failed to satisfy the employees' assignment rights.
5. Had provided inadequate notice to employees as required in the RIF procedures.

MANAGING THE RIF PROCESS

The prospect of an impending RIF action has a chilling effect on employee morale and productivity. Often, management is reluctant to share information with employees in the early stages of RIF planning because the timing and impact are still so uncertain. Once rumors of an RIF begin to spread throughout the workforce, however, the best course of action is to share as much information as possible with employees and/or with the union that represents the employees. Most employees would rather be informed of details, even if they are subject to change, than to hear nothing from their managers.

Employees will naturally turn to their immediate managers for information. If the first-line managers have not been briefed by top management or by human resources, it will contribute to distrust and disaffection on the parts of employees and their managers. Therefore, the first-line managers need to receive briefings and the opportunities to get answers to the questions they will hear from their staff. Human

resources may wish to prepare a briefing document to guide first-line managers in conducting employee meetings. In addition, top management should take advantage of all communication vehicles on a regular basis: memoranda, e-mail, video systems, bulletin boards, employee websites.

If some employees will be separated as a result of the RIF, human resources should establish a counseling system to help them prepare for their lives after leaving the agency. Employees need to be given administrative time to prepare resumes. If possible, a private space should be set up outside the workplace where employees can meet with counselors, search for jobs, and prepare resumes. Employee sessions should be planned to share information about severance pay, health and life insurance benefit continuation, and any placement programs.

Employees who are separated or bumped to another position are subject to the stages of shock and mourning that were first identified by Elisabeth Kübler-Ross in 1969: shock, anger, depression, flight, and acceptance.^[2] Whether they are treated with compassion and respect may well have a substantial impact on the morale of those individuals who remain agency employees.

REFERENCES

1. Merit Systems Protection Board (MSPB). <http://www.mspb.gov/mspbdecisionspage.html> (accessed 17 January, 2005). (While federal, state, and local government agencies are all subject to review by federal courts and the Equal Employment Opportunity Commission, only the federal agencies are subject to review by the MSPB.)
2. Kübler-Ross, E. *On Death and Dying*; Macmillan: New York, 1969.

BIBLIOGRAPHY

1. Regulations governing RIFs in federal government agencies may be found in Title 5, Part 351 of the United States Code of Federal Regulations; Government Printing Office: Washington, DC, 2004.

Regulatory Policy: Role and Importance

Scott R. Furlong

*Public and Environmental Affairs, University of Wisconsin-Green Bay,
Green Bay, Wisconsin, U.S.A.*

INTRODUCTION

Regulatory policy is dominant in the United States. Most of the policies that affect American's everyday lives are probably a direct result of regulations from agencies such as the Environmental Protection Agency (EPA), Food and Drug Administration, or the Occupational Safety and Health Administration. Yet, most people do not think much about this important area of public policy. In fact, until the 1980s and the Reagan administration, regulatory policy was rarely mentioned in discussions of government and politics. We still tend to focus on statutory laws without recognizing that the implementation of those laws is where the proverbial "rubber hits the road." The passage of the Occupational Safety and Health Act did not directly make the workplace safer, but regulations issued by OSHA did and continue to do so. This article examines the evolution and role of regulatory policy in the United States, as well as the obvious importance of this type of policy in our daily lives.

CONTROVERSIES AND DEFINITIONS OF REGULATION

We all have some intuitive sense about what the term regulation means. Often this sense is driven by our ideological beliefs about the role of government in our daily lives. Conservatives, and many Republicans, see regulation as government interference or intrusion upsetting the free market and imposing undue costs on business and society. For liberals, and many Democrats, regulation is seen as government authority to protect the general public from unfair and unsafe practices of corporate America. It is clear that both of these insights are tainted by one's philosophical beliefs.

Eisner et al. define regulation as "an array of public policies explicitly designed to govern economic activity and its consequences at the level of the industry, firm, or individual unit of activity."^[1] This definition obviously is more neutral and describes the actual activity occurring. It also raises a common way to examine types or varieties of regulatory policy. These characterizations of regulatory policy are often referred to as economic regulation and social regulation. Economic

regulation, the more traditional of the two, focuses on the role of a firm in a market. From the general public's perspective, economic regulation tries to protect consumers from the adverse effects of monopolistic practices that can have an adverse impact on consumers. Regulation in this area tends to focus on single industrial sectors, such as broadcasting, railroads, and the stock market. Although economic regulation might be considered the more traditional of the two types, most people when they consider regulatory policy, associate regulation with social regulatory policy. Social regulation tends to focus on the activities of industry in the production of its products. Many of the policies in this area focus on the negative externalities (spillovers) that result from production, such as workplace health and safety issues and environmental protection. In general, many of our social regulatory policies, unlike economic regulation, are broader in their application and do not focus on a single industry. As an example, workplace safety regulations can potentially affect every industry and firm that is trying to meet a particular regulatory criterion set by the agency.

Attacks against regulatory policy often focus on social regulatory policy. In many ways, the Reagan administration typified these attacks, claiming that excessive regulation in this form led to problems with the economy. Simplistically, because businesses needed to install protections of various sorts to comply with these regulations, they were increasing their cost of production and ultimately having a negative impact on the overall economy. Because many social regulatory policies are not industry specific, the impact of workplace safety and health regulations, for example, could be huge. Therefore, to improve the economy, many conservatives believed that there was a need to deregulate in the social regulatory policy area, much like it was done with some economic regulatory policy in the 1970s.

The 1970s did see a strong movement toward deregulation, but as stated, it focused on industry-level regulation. The prime example was the deregulation of the airline industry, around which an interesting coalition developed. Conservatives were supportive of deregulation and of allowing the market to work on its own on the basis of its merits. Liberals believed that deregulation would cause airline ticket prices to

decrease and allow air travel to become accessible to more people. The competition introduced as a result of deregulation (as well as other forces) led to the desired goal in the airline industry—more competition and prices that were likely lower than they would have been otherwise.

Increased competition, though, does not necessarily lead to the desired goals of social regulation. Many social regulatory policies deal with a market failure called a negative externality. A negative externality occurs when two parties interact in a market transaction, and when a third party is adversely affected by this transaction and not compensated for their loss. For example, producing electricity causes air emissions that may adversely affect the health of those downwind. Increasing competition, such as when the airline industry was deregulated, will lead producers to decrease prices to attract consumers. To do so, they will want to reduce their costs of production, which many times are the controls installed to protect society. For example, pollution control technology is installed to reduce pollutants from the processing stream, but this adds to the cost of doing business. Deregulating in this area could mean that businesses do not install these controls (to save money); therefore, pollution emissions are increased.

In general, it appears that the general public supports many of our social regulatory policies. Public opinion polls continually show strong support for environmental and consumer protection policies. There was a significant backlash against the Reagan administration after his attempts to lessen social regulatory policy, especially by the Democratic House of Representatives. A similar backlash occurred against the Republican Congress in 1995 when they attempted to reduce funding in these areas, particularly for environmental protection. Therefore, one can see significant conflict in this public policy arena. Increases and decreases in the extent of regulatory policy tend to generate much debate these days especially at the macropolicy level.

Another issue often raised in regard to economic regulatory policy specifically is the issue of agency capture. Numerous scholars began analyzing the role of regulatory agencies and their relationships with the businesses they were tasked by Congress to regulate. Bernstein's analysis discussed that at some point in an agency's "life cycle," it begins to be more concerned with the health of the industry than with the interests of the general public.^[2] Others also examined the potential issues associated with capture, and many believe that one of the arguments for deregulating the airlines was the concern about capture.^[3,4] Other scholars used the capture argument as evidence and reasons as why extensive delegation of regulatory power was problematic.^[5,6]

Regulatory policy is not without its issues. These issues range from the economic—how these types of

policies can affect economic growth in the United States and health of industries—to the democratic—whether so much power should be given to unelected agency officials primarily responsible for developing regulatory policy. These issues are not new and have been around for almost as long as the existence of regulatory policy (as we currently think of it) itself.

EVOLUTION AND STAGES OF REGULATORY POLICY

Scholars often examine the evolution of regulatory policy in the United States in four distinct stages. The first stage occurred in the late 19th/early 20th century, which saw the creation of the first independent regulatory commission, the Interstate Commerce Commission. Much of this early regulatory policy was strictly economic in scope in that it focused on the private activities of businesses and attempted to control potential monopolistic behavior. Other agencies such as the Federal Trade Commission followed, and again the focus was on the protection of the consumer from unfair labor practices. The second regulatory stage corresponds to the New Deal era. The New Deal era ushered in a much more extensive role for the federal government in our daily lives. This was exemplified by the increasing number of federal agencies, many of them regulatory in nature. Many of the regulatory policies during this time, not surprisingly, were geared toward economic recovery, and many of the policies continued to examine the role of competition in various industries.

The policies of the New Deal and how these policies were created laid the groundwork for the third stage of regulatory policy—the rise of social regulatory policy in the 1960s and 1970s. As mentioned earlier, social regulatory policy is quite a bit different than economic regulatory policy. The goals of social regulatory policy are much more expansive and typically far reaching, and tend to deal with issues of health, safety, or equal opportunity. Wilson classified many of these policies as those that have concentrated costs and diffused benefits.^[7] The costs of the policies are (typically) borne by a few businesses, but the benefits are provided to many. These regulatory policies continued the expansion of the federal role in the United States. Now the government was regulating the safety of consumer products, the workplace safety conditions, and quality of the water among many other areas, and it was not doing so as a way of promoting competition. Rather, the expressed purpose of these policies was the physical protection of the general public from products and industrial processes.

The fourth stage, which we may still be in, encompasses the era of deregulation (discussed earlier) and

various attempts of regulatory reform that followed. The deregulation of many of our industries also fueled similar ideas in the area of social regulation. Numerous regulatory reform efforts promoted and implemented by different administrations have had a profound impact on regulatory policy since the mid-1970s. Often, these reforms have been focused on the management of the rule-making processes used by agencies and the use of analyses to better inform regulatory decisions. President Reagan's Executive Order 12291 is the best example of both these efforts. This executive order initiated elaborate rule-making reviews by the Office of Management and Budget (OMB) and also required agencies to conduct benefit-cost analyses for major rules and take this information into consideration when making decisions. Similar analytical and managerial requirements would follow that affected not only the quantity of regulations issued by federal agencies, but also the nature of the debate within agencies.

The introduction of economic analyses in the development of regulations and the use of these analyses to help in decision making brought to the forefront the idea that regulatory policies may have a detrimental impact on the economy. Different administrations used these arguments in their attempts to decrease the amount of regulatory activity from the agencies. This was especially true during the Reagan administration, which showed major reductions in the number of rules coming out of federal agencies compared with the Carter administration. Managerial changes attempted to centralize the rule-making process, as well as to provide more oversight to the President and his staff, to ensure that agencies were following the administration's goals and agenda. The OMB was used as a clearinghouse for many regulations by reviewing them prior to an agency issuing the policies. The activities of the OMB, and later President George H. Bush's Council on Competitiveness, had the effect of requiring agencies to make changes to their regulations to follow the administration's ideals or having their policies delayed for extended periods of time. Agency behavior changed so that they would give more careful consideration to the development of agency regulations to ensure that they would pass the new administrative hurdles placed in front of them. Economic shops within agencies were expanded and given additional authority to develop and review regulations internally. Now that past presidents have created and used the OMB as a way of monitoring agency rule making, any future president would be crazy to abolish such functions. They may redefine their role or the types of regulations being reviewed, but all would want to keep these mechanisms in place to take advantage of the possibility to influence policy at this level. In addition, presidents also have the ability to change policies dramatically without going through the process of

getting new laws passed by using changes in regulations to pursue their agendas.^[8]

REGULATORY POLICY NOW AND IN THE FUTURE

This article began by stating that regulatory policy is dominant in our daily lives. If you do not believe this, try thinking about your day-to-day activities and how the most mundane activities are likely affected by policies developed by regulatory agencies. The car you are driving must meet certain safety standards imposed by the Department of Transportation. The labels on your groceries are a result of work by the Food and Drug Administration. OSHA governs your workplace in terms of safety and health issues. It is clear that regulatory policy is here to stay, but will we see some changes?

One area where we may continue to see movement in regulatory policy relates to the issue of private or self-regulation. This term has a number of meanings, but one can think of this as a continuum of the amount of control entities have in regulating themselves or ensuring the implementation of regulations placed upon them. This is not a new idea; in fact, there have been some implicit and explicit federal policies to encourage this form of regulation. One example of private regulation is the movie industry's rating system for content. This is an area where the government has allowed the industry and its professional organization to regulate itself. Some agencies have attempted to move in the direction of acting as a consultant for the areas in which they regulate in order to facilitate achieving the desired results. Under President Clinton's National Performance Review, for example, OSHA provided businesses with the option of forming a partnership with the agency; through this partnership, businesses "would receive technical assistance and consultation" from OSHA.^[1] Might we see more of these types of activities and relationships between government and business? This type of relationship suggests the government as a facilitator and a resource to help businesses reach regulatory goals, rather than an adversarial relationship.

A related change that is likely to continue into the future is an increase in public participation methods and the use of alternative dispute resolution programs. Again, these are not new programs. For years, agencies have explored different ways to increase participation in the development of their regulations and policies. One of the primary components of the Administrative Procedure Act are requirements for this participation. Agencies have also explored alternative means to reach consensus or decisions in regard to their policies. Many have lauded the potential of processes such

as regulatory negotiation for a number of years.^[9,10] Agencies have continued to build upon these alternative methods of gathering input and data on their policies. They have also been used to work with the regulated community to attempt to reach mutual decisions. The EPA is one agency that has attempted a number of these process reforms to improve its policies and relationships. The EPA has been one of the leaders in the use of regulatory negotiation processes in their rule development. In addition, it has used numerous alternative dispute resolution processes with varying levels of success.^[11] Its Project XL is another example of a program designed to encourage increased and early participation. Under this program, the EPA works with the major stakeholders in a particular area to develop more cost-effective ways of achieving environmental protection. Recently, the EPA has also experimented with the use of online democracy by using the Internet as a way of increasing the amount of participants dialoging on a specific policy.^[12] One can expect that programs that encourage greater participation and focus on more conciliatory relationships will continue throughout regulatory agencies. They may not always be successful or even engender good will between business and government, but they are potential steps in the right direction. Critics may argue that these types of relationships raise the possibility of a type of capture that needs to be avoided to protect the general public.

Related to these efforts by the EPA are increased efforts by a number of agencies to use the Internet as a way to increase participation through the traditional notice-and-comment rule-making process. These efforts have made it easier for organizations and individuals to learn about potential rule-making efforts through websites such as Regulation.gov, which also provides a simple format in which to send comments to agencies. In addition, a number of agencies such as the Department of Transportation and the EPA are making their rule-making dockets available online and hence providing better access to information and other comments to those interested in participating.

CONCLUSION

This article has discussed numerous issues regarding regulatory policy. It is clear that regulatory policy is

pervasive in our daily lives and will continue to be so in the future. It is also clear that there are different forms of regulatory policy. There is a vast difference between economic regulatory policy, which started the regulatory era, and social regulatory policy, which is now much more common. In addition, people have different beliefs and concerns about the role of this policy. The debate will continue about the impact of regulatory policy, and we can expect numerous reforms to also continue to enter into the debate. The political, economic, and ethical concerns of this important area of public policy will continue to challenge scholars as long as the public continues to demand services and policies from government.

REFERENCES

1. Eisner, M.A.; Worsham, J.; Ringquist, E.J. *Contemporary Regulatory Policy*; Lynne Rienner: Boulder, CO, 2000; 5, 193 pp.
2. Bernstein, M.H. *Regulating Business by Independent Commission*; Princeton University Press: Princeton, NJ, 1955.
3. Stigler, G.J. The theory of economic regulation. *Bell J. Econ. Manage. Sci.* **1971**, 2 (1), 3–21.
4. Peltzman, S. Toward a more general theory of regulation. *J. Law Econ.* **1976**, 19 (2), 211–240.
5. McConnell, G. *Private Power and American Democracy*; Knopf: New York, 1966.
6. Lowi, T.J. *The End of Liberalism: The Second Republic of the United States*; W.W. Norton: New York, 1979.
7. Wilson, J.Q. *The Politics of Regulation*; Basic Books: New York, 1980.
8. Furlong, S. Business and the environment: influencing agency policymaking. In *Business and Environmental Policy: Corporate Interests in the American Political System*; Kraft, M., Kamieniecki, S., Eds.; MIT Press: Boston, 2007.
9. Harter, P. Negotiating regulations: a cure for the malaise. *Georgetown Law J.* **1982**, 71, 1–118.
10. Fiorino, D. Regulatory negotiation as a policy process. *Public Admin. Rev.* **1988**, 48, 764–772.
11. O'Leary, R.; Raines, S.S. Lessons learned from two decades of alternative dispute resolution programs and processes at the U.S. Environmental Protection Agency. *Public Admin. Rev.* **2001**, 61, 682–692.
12. Beierle, T.C. *Democracy On-Line: An Evaluation of the National Dialogue on Public Involvement in EPA Decision*; Resources for the Future: Washington, DC, 2002.

Representative Bureaucracy

Katherine C. Naff

*Public Administration Program, San Francisco State University,
San Francisco, California, U.S.A.*

INTRODUCTION

The term representative bureaucracy refers to the notion that government agencies should mirror the composition of the population they serve. Former President Clinton summarized the concept when he promised a government that would “look like America.” In saying that, he was referring specifically to the idea that his appointed officials would reflect the gender and racial diversity of the American population. That the federal civil service should also reflect this diversity has been the policy of the U.S. Government for nearly three decades.

This entry discusses the origin and development of the concept and the reasons that have been proffered for the need to ensure a bureaucracy that is representative. It further discusses controversies over how such representation should be defined and measured as well as empirical research that has substantiated the notion that representation can make a difference. It concludes with some thoughts about directions for future research.

THE BIRTH OF A THEORY

Original notions of bureaucracies, as developed by Max Weber in his essays, saw them as ideal organizational structures in a democratic environment because they are objective and business is carried out according to rules, in a “dehumanized” way. Thus, all citizens are treated equally. And yet at the same time there is an inherent contradiction between bureaucracies and democracy. They are staffed by civil servants who are responsible for developing and enacting policies within the broad parameters of laws passed by Congress, but these civil servants are unelected and not directly accountable to citizens.

With that concern as a backdrop, scholars came up with various prescriptions as to how to make bureaucrats more accountable, including strengthening executive and/or legislative control. But Kingsley, who coined the term “representative bureaucracy,” was not convinced this was feasible. Writing about the British government during the second World War, he noted that it was the middle class origins of the

civil servants that rendered them able to carry out the government’s policies because the political party holding office shared their social origins. But when the working class Labour Party assumed power, their policies would not be carried out successfully by the middle class civil servants. In his 1944 book he wrote:

[T]he essence of responsibility is psychological rather than mechanical. It is sought in an identity of aim and point of view, in a common background of social prejudice, which leads the agent to act as though he were the principal. . . . No group can safely be entrusted with power that do not themselves mirror the dominant forces in society.^[1]

Kingsley did not apply his concept to American politics, nor did he articulate very well the benefits of a representative bureaucracy.^[2] Other scholars, however, stepped into this breach and further clarified and strengthened his focus on the men and women who comprise the civil service.

THE AMERICAN SCENE

Levitan was the first scholar to take hold of Kingsley’s notion and apply it to the American landscape.^[3] He agreed with Kingsley that external controls were too inflexible to be relied on to control the modern bureaucracy. Other adherents to the theory included Long who suggested that “the rich diversity that makes up the United States is better represented in its civil service than anywhere else,” including Congress.^[4] Others were less enthusiastic about the notion, arguing that it is not necessarily the case that bureaucrats retain the values of their social origins, nor that a representative bureaucracy guarantees that all of society’s interests are taken into account.^[5–7]

Krislov stepped into the debate and suggested that a representative bureaucracy serves other laudable purposes besides serving as a restraint on bureaucratic power.^[8] It:

- Includes a broader array of skills that can augment its ability to address a broader range of social problems;
- Acts as a funnel for very different viewpoints;

- Shares social responsibility, resulting in a broader approval of governmental policies;
- Demonstrates that all have access to government positions; and
- Ensures that even members of disadvantaged groups will be encouraged to gain the education and other qualifications necessary to serve in public office.

“Bureaucracies, by their very structure,” Krislov wrote, “represent truths about the nature of the societies and the values that dominate them” (p. 64).

If a representative bureaucracy provides these other benefits, then, the debate as to whether bureaucrats indeed preserve the values associated with their social origins is less important since a diverse civil service also serves an important symbolic role.

But what characteristics should a bureaucracy display if it is to be considered representative? Most of these early scholars, including J. Donald Kingsley, saw a representative bureaucracy as one that shared the social class origins of the citizenry, and concluded that the federal civil service was indeed representative.^[4,5,9] However, in the United States, with the vocal women’s rights and civil rights movements of the 1960s, that orientation gradually shifted from social class to one where the concern was for sufficient women and minorities to be represented in the bureaucracy. There were a number of reasons for this, including a history of discrimination against these groups in the government (as well as society in general.) Moreover, as government had responsibility for enforcing equal employment opportunity legislation passed by Congress, it had to ensure its own house was in order in this respect. With the passage of the Civil Service Reform Act of 1978, it became the official policy of the U.S. Government to provide a “federal workforce reflective of the nation’s diversity.” The Act further directed agencies to make efforts to eliminate the underrepresentation of people of color and women in all occupations and at all grade levels (5 U.S.C. § 7201).

The federal government has had considerable success in recruiting women and people of color. By the end of fiscal year 2004, women held 43%, and people of color held 33% of permanent federal government jobs.^[10] Whether these percentages mean that the government is indeed representative, however, is not a simple question to answer.

ASSESSING THE QUANTITY AND QUALITY OF REPRESENTATION

Counting the Numbers

The percentages cited above denote what has been called “passive” representation.^[11] Some scholars have

argued that these numbers signifying the overall representation of people of color and women are not good enough. If these groups are to have a real voice in policy making, they must be well represented in senior level jobs. As a result, scholars as well as federal agencies assigned to the task [e.g., the U.S. Office of Personnel Management (OPM) and the Equal Employment Opportunity Commission (EEOC)] also track the proportion of women and people of color in higher level jobs. It turns out that in these jobs, these groups are not as well represented (see Fig. 1). At the end of fiscal year 2004, men held 74%, and whites held 86% of senior pay positions in the federal government.

However, another problem in defining the extent of representation is that there has never been a fully agreed upon way to measure representation. While some early scholars believed the percentage of women and people of color in the bureaucracy should be equal to their representation in the population, the EEOC and OPM employ a more limited benchmark.^[10,13] For these agencies the point of comparison is the civilian labor force (CLF), which includes all people in the United States who are at least 16 years old and not employed by the military. As shown in Table 1, by this measure Hispanic men and women and White women are underrepresented in government. All other groups are fully represented.

However, even these two agencies use different numbers to compute representation. The EEOC includes federal agencies that are not included in OPM’s calculation of government representation, and while OPM includes Puerto Rico in its estimation, the EEOC does not. So, while the EEOC reported that Hispanics represented 7.46% of the

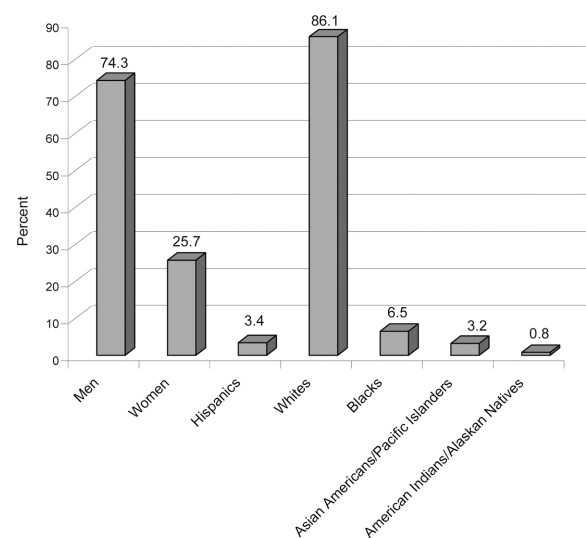


Fig. 1 Representation of men, women, and minorities in senior pay level positions in the Federal Government (FY 2004). (From Ref.^[10]).

Table 1 Representation of racial and ethnic groups in the federal workforce and CLF (%)

	Representation in federal workforce (2004)	Representation in CLF (2000)
Hispanic males	4.48	6.20
Hispanic females	2.98	4.50
White males	40.71	39.00
White females	26.19	33.70
Black males	7.81	4.80
Black females	10.37	5.80
Asian American/Pacific Islander males	3.32	2.10
Asian American/Pacific Islander females	2.46	1.90
American Indian/Alaskan Native males	0.77	0.50
American Indian/Alaskan Native females	0.90	0.50

(From Ref.^[10].)

federal workforce and 10.7% of the CLF in 2004, OPM reported that Hispanics held 7.3% of federal jobs and 12.6% of the CLF in the same year.^[10,13]

Some scholars are even critical of these measures and have proposed others. In a 1973 publication, Nachmias and Rosenbloom suggested a Measure of Variation that compares the total number of differences within a group to the total number of potential differences.^[14] The U.S. General Accounting Office has advocated the use of a “ratio-based approach” that compares the numbers of women and each ethnic group to a benchmark group such as white men.^[15] Even counting the numbers, then, is not a simple matter.

Looking for Action

These measures of passive representation have been criticized for another reason. There are those who argue that passive representation is pointless unless it results in “active” representation.^[11] That is, there must be evidence that these so-called representatives are actually acting in the interests of the groups they are presumed to represent. Some research has indeed found bureaucrats playing such a role, although some has not. For example, Hinderer and colleagues have found that the greater the number of African American and Hispanic field investigators in EEOC offices, the greater the likelihood that complaints brought by African Americans and Latinos, respectively, will be pursued.^[16,17] Selden found a similar positive correlation between the representation of Hispanic, African American, and Asian American county supervisors in Farmers Home Administration Offices and the number of favorable eligibility decisions for loan applications filed by members of these groups.^[18,19] Meier and colleagues have identified an inverse relationship between the proportion of African American and Hispanic school teachers and school board

members and the incidents of disciplinary actions taken against students in these groups.^[20–22] It should be noted, however, that such relationships have not always appeared as expected. For example, Hinderer found a negative relationship between the number of white women investigators with the EEOC and the number of complaints filed on behalf of women.^[23] Selden found neither the representation of women nor Native Americans were related to loan eligibility determinations for members of these groups.^[19]

Clearly it would be beneficial to continue to add agencies to this investigation of the relationship between passive and active representation. But while this research is promising, it is also limited in that it only looks at the subset of government employees who work directly with citizens. Yet there are hundreds of thousands of bureaucrats who hold other significant policy making and implementation roles. Research in a range of fields, including psychology, sociology, and public administration, has revealed subtle ways in which the authority and influence of people in such positions can be limited.^[24] More research is required to ferret out the nature of such constraints and to hopefully identify the means for eliminating them.

CONCLUSION

The exceptions notwithstanding, scholars have found considerable evidence, at least among bureaucrats who are involved in direct client service, that passive representation does provide benefits in the form of additional assistance to those being represented. In addition, as Samuel Krislov has suggested, there are additional benefits served by a representative bureaucracy. These include improved policy formulation because of the diversity of viewpoints and experiences brought to bear and the increased legitimacy of the government as the nation’s largest employer and

enforcer of its equal opportunity laws. Once only a theory advocated by scholars, a federal statute now calls for a civil service “reflective of the nation’s diversity” and agencies such as the EEOC and OPM are charged with monitoring the extent to which this objective is achieved at all grade levels. Future research should continue to examine the ramifications and value of a representative bureaucracy, while also assessing the extent to which those benefits are being limited by barriers to the full authority and influence of women and people of color.

REFERENCES

1. Kingsley, J.D. *Representative Bureaucracy: An Interpretation of the British Civil Service*; The Antioch Press: Yellow Springs, OH, 1944.
2. Krislov, S. *Representative Bureaucracy*; Prentice-Hall: Englewood Cliffs, NJ, 1974.
3. Levitan, D.M. The responsibility of administrative officials in a democratic society. *Polit. Sci. Quart.* **1946**, *61*, 562–598.
4. Long, N.E. Bureaucracy and constitutionalism. *Am. Polit. Sci. Rev.* **1952**, *46*, 808–818.
5. Subramaniam, V. Representative bureaucracy: A reassessment. *Am. Polit. Sci. Rev.* **1967**, *61*, 1010–1019.
6. Larson, A.D. Representative bureaucracy and administrative responsibility: A reassessment. *Midwest Rev. Public Admin.* **1973**, *7*, 79–89.
7. Meier, K.J. Representative bureaucracy: An empirical assessment. *Am. Polit. Sci. Rev.* **1975**, *69* (2), 526–542.
8. Krislov, S. *The Negro in Federal Employment*; University of Minnesota Press: Minneapolis, MN, 1967.
9. Van Riper, P.P. *History of the United States Civil Service*; Row, Peterson and Company: Evanston, IL, 1958.
10. U.S. Equal Employment Opportunity Commission. 2005. *Annual report on the federal work force, Fiscal year 2004*, <http://www.eeoc.gov/federal/fsp2004/index.html>, 2005 (accessed February 2006).
11. Mosher, F.C. *Democracy and the Public Service*; Oxford University Press: New York, 1968.
12. Mosher, F.C. *Democracy and the Public Service*, 2nd Ed.; Oxford University Press: New York, 1982.
13. U.S. Office of Personnel Management. *Annual Report to the Congress: Federal Equal Opportunity Recruitment Program*; U.S. Office of Personnel Management: Washington, DC, 2005.
14. Nachmias, D.; Rosenbloom, D.H. Measuring bureaucratic representation and integration. *Public Admin. Rev.* **1973**, *33*, 590–596.
15. U.S. Government Accounting Office. *Affirmative Employment: Assessing Progress of EEO Groups in Key Federal Jobs can be Improved*, GAO/GGD-93-65, 1993.
16. Hinderer, J.J. Representative bureaucracy: Further evidence of active representation in the EEOC district office. *J. Public Admin. Res. Theory* **1993**, *3* (4), 415–442.
17. Hinderer, J.J.; Young, C.D. Representative bureaucracy: The theoretical implications of statistical interaction. *Polit. Res. Quart.* **1998**, *51* (3), 655–671.
18. Selden, S.C. Representative bureaucracy: Examining the linkage between passive and active representation in the Farmers Home Administration. *Am. Rev. Public Admin.* **1997**, *27* (1), 22–42.
19. Selden, S.C. *The Promise of Representative Bureaucracy*; M.E. Sharpe: Armonk, NY, 1997.
20. Meier, K.J. Teachers, students and discrimination: The policy impact of black representation. *J. Politics* **1984**, *46* (1), 252–263.
21. Meier, K.J. Latinos and representative bureaucracy. *J. Public Admin. Res. Theory* **1993**, *3* (4), 393–414.
22. Meier, K.J.; Stewart, J., Jr. *The politics of Hispanic Education*; State University of New York Press: Albany, NY, 1991.
23. Hinderer, J.J. Representative bureaucracy: Imprimis evidence of active representation in the EEOC district offices. *Social Sci. Quart.* **1993**, *74*, 95–108.
24. Naff, K.C. *To Look Like America*; Westview: Boulder, CO, 2001.

Restorative Justice

Michelle Maiese

University of Colorado at Boulder, Boulder, Colorado, U.S.A.

INTRODUCTION

The dominant understanding of crime and justice in the United States is shaped by a retributive framework. This framework focuses on the establishment of guilt and understands crime as a matter of lawbreaking and an offense against the state. When a law is broken, this creates a debt that must be repaid to society. Offenders must receive their “just desserts” and endure their punishment. The United States has established elaborate legal mechanisms to administer “just” doses of pain to those who have willfully broken the law.

Restorative justice, on the other hand, understands crime as an offense against real people and relationships rather than the state. Making things right requires healing victims’ wounds, restoring offenders to law-abiding lives, and repairing the harm done to interpersonal relationship and the community. Moreover, restorative justice recognizes that victims, offenders, family members, and affected communities are all key stakeholders in the restorative process and should play an active role in deciding how to repair the damage caused by the offense. Restorative justice thus represents a progressive way of thinking about crime and justice. The alternatives to punishment that it suggests can help to empower and heal victims, offenders, and communities (Table 1).

NEEDS OF VICTIMS, OFFENDERS, AND COMMUNITIES

The safety, support, and healing of crime victims are the starting points for any restorative justice process. A primary objective is to attend to victims’ needs: material, financial, emotional, and social.^[1] Victims need to receive compensation for any property they have lost and any physical injuries they have suffered. Other central victims’ needs include security, dignity, social support, and the sense that justice has been done.^[2] They must have the opportunity to tell their story and express their feelings about what has happened. They also need to feel vindicated and to know that others recognize what happened to them was wrong.

Restitution often plays a central role not only in restoring material losses, but also in acknowledging wrongdoing. Restitution agreements symbolize the fact

that offenders accept responsibility for the harm they have caused and wish to make amends. However, rather than focusing on punishing or defeating the wrongdoer, restitution aims to elevate the victim.^[3] Restorative justice therefore advocates restitution to the victim by the offender rather than retribution by the state against the offender. Instead of causing further harm, it tries to restore relationships and stop the cycle of violence.^[4]

Restorative justice also requires that victims be empowered. Because our current system defines the state as victim, this typically means that actual victims “are mere footnotes to the criminal justice process” and do not play an active role in their own cases.^[3] Restorative justice, on the other hand, acknowledges that because part of the harm victims have suffered is loss of power, one dimension of justice is ensuring that this power is returned to them. Victims must have access to information about who the offender is, why they were targeted as victims, and what is being done about the crime. They must also have the opportunity to engage in dialogue with offenders. Victims should play an active role in directing the exchange that takes place and defining the responsibilities and obligations of offenders.

Offenders are likewise encouraged to participate in this exchange, to understand the harm they have caused to victims, and to take active responsibility for it. Both the rehabilitation of offenders and their reintegration into the community are vital aspects of restorative justice. Restorative processes encourage offenders to understand and acknowledge the harm they have caused and assume responsibility for making things right. This means making an effort to make amends for their violations by committing to certain obligations: reparations, restitution, or community work. In addition, offenders should share in the responsibility for deciding what needs to be done to repair the harm they have caused. This may increase their sense of accountability.

While fulfilling these obligations may be experienced as painful, the goal is not revenge, but rather the restoration of healthy relationships between victims, offenders, and communities. Moreover, if offenders can play an active role in the justice process and take responsibility for what they have done, this can contribute to their rehabilitation. Removing offenders

Table 1 Comparison of retributive and restorative justice models

Retributive justice	Restorative justice
Crime is a matter of lawbreaking and an offense against the state.	Crime is an offense against individuals, relationships, and communities.
Those who commit crimes freely and willfully break the law.	Those who commit crimes act within a specific context. In many cases offenders themselves have suffered harm.
The process focuses on the establishment of guilt and is oriented toward the past.	The process focuses on offenders taking responsibility and is oriented toward the future.
To make things right, offenders must receive their “just desserts” and undergo pain.	To make things right, the harm caused to individuals and communities must be repaired.
Responding to crime centers on punishment.	Responding to crime centers on healing.
Victims’ needs are peripheral.	Victims’ needs are central.
Legal personnel process cases. Victims and offenders do not play an active role in the justice process.	Victims and offenders play an active role in determining outcomes.

Source: From Ref.^[3].

from the community, or imposing any other severe restrictions, should be viewed as a last resort. The best way to prevent reoffending is reintegration.

Restorative justice is a forward-looking, preventive response that strives to understand crime in its social context. Although offenders are responsible for the harm they have caused, their behavior arises out of a social, economic, and psychological context. Moreover, crime often grows out of harm that the offender himself/herself has suffered. Restorative justice challenges us to examine the root causes of violence and the social conditions that give rise to crime.^[4] It suggests that communities must take some responsibility for remedying those conditions that contribute to crime and work to promote healing. This means transforming relationships and communities so that crime does not recur.

Restorative justice processes thus aim to strengthen the community and promote changes that will prevent similar harms from happening in the future. Many believe that the implementation of restorative justice processes can improve the quality, effectiveness, and efficiency of justice as a whole.^[3] A growing number of state and county justice systems in the United States have begun to incorporate “restorative” policies and programs (Table 1).

RESTORATIVE JUSTICE INITIATIVES

Restorative justice initiatives provide opportunities for dialogue and problem solving to interested crime victims and offenders so that they can more fully understand the events that have taken place. In short, the initiatives aim to put power and responsibility in the hands of those directly involved in the crime.

Such programs take on various forms. Victim–offender mediation is perhaps the most common and

involves face-to-face dialogues between victims and offenders. Group conferencing is an extension of victim–offender mediation and involves more participants, including family members of the victim or offender, community contacts, teachers, neighbors, or counselors. In addition, community victim-support organizations work to provide victims with material, psychological, and social support to aid in the healing process. Other organizations assist in reintegration for offenders and provide literacy education, employment services, counseling, and housing accommodation.

Research has shown that these restorative justice programs provide higher levels of victim and offender satisfaction than do traditional criminal justice programs.^[5] They are significantly more likely to result in victims’ perceptions that their cases were handled fairly and help to humanize the criminal justice experience.^[2] Below are more detailed descriptions of four restorative practices commonly used in North America and Europe: victim–offender mediation, group conferencing, community boards, and circle sentencing. These processes are non-adversarial, less formal than current criminal justice proceedings, and typically involve community members in planning and implementation.^[6] The continued development of such initiatives indicates that the range of services available for crime victims is expanding (Table 2).

Victim–Offender Mediation

Victim–offender mediation (VOM) provides interested victims the opportunity to meet with their offenders in the presence of trained volunteers who act as mediators. A central goal of VOM is to actively involve both victims and offenders in addressing the emotional and material harm that has been done. Through a face-to-face meeting, victims and offenders can develop

Q-Sexual

Table 2 Comparison of four restorative justice processes

	Victim–offender mediation	Family group conferencing	Community reparative boards	Circle sentencing
Key participants	Mediator, victim, and offender.	Victim, offender, their close family members and friends, police, and social service persons.	Reparative coordinator, community board participants, and offender.	Victim, offender, support persons, and relevant community members.
Prior to meetings	Mediators have face-to-face meetings with victims and offenders to explain the process.	Coordinators contact the offender, the offender's family, and the victim (either by phone or in person).	There is no preparation for individual hearings. Board members receive general preservice training.	Community justice committees work extensively with offenders and victims to prepare them for the process.
Role of facilitator	Mediator manages dialogue and keeps conversation flowing. Victim and offender do most of the talking.	The coordinator calls on each participant to describe the impact the crime has had on his or her life and ensures that all participants have a chance to speak.	Board chairperson dictates the course the meeting takes and calls on participants to speak.	“Keeper” simply initiates the conversation. Dialogue proceeds by passing the talking piece.
Decision of outcome	The victim plays a central role in shaping the restitution agreement. Both victim and offender must consent.	All participants, including offender, work together to determine how the offender can best repair the harm caused. Participants typically sign an agreement.	Board members develop a list of proposed sanctions. The offender agrees to fulfill obligations within a certain time frame.	Participants gather in a sentencing circle to develop a sentencing plan and identify the obligations the offender must fulfill.

Source: From Ref.^[6].

a better understanding of what has happened and work to develop mutually acceptable restitution plans.

This dialogue-driven form of “mediation” differs from traditional mediation insofar as the parties involved are not disputants and do not discuss the subject of guilt. Because the process is not focused on settlement, victims are not expected to compromise with respect to having their needs met and losses restored. Rather than passively taking their punishment, offenders take an active role in figuring out how to restore losses. Together, victim and offender determine the best way to repair the harms suffered.

Victim–offender mediation stresses the importance of extensive preparation for both victim and offender prior to the session. Before a dialogue takes place, the mediator meets with both offenders and victims to inform them about the process and invite them to participate. Preliminary meetings give the mediator a chance to establish credibility and rapport with the victim, provide information about the mediation program, describe the mediation process and its goals, and explain the possible benefits and risks of participation. Should the victim make an informed decision to participate, mediators also help to brainstorm about how losses and needs might be addressed. A dialogue with the offender should not be scheduled until the victim is ready and has realistic expectations about the process.^[7]

Meetings should occur in a structured setting in a location that victims consider safe. The dialogue itself should be structured so as to neutralize differences in status and power and ensure that both parties’ complete stories are heard. Parties are typically seated across from each other at a table, allowing them to have direct eye contact during the session. Support persons often accompany victims to the session to increase their sense of security and comfort. The mediator encourages participants to express their feelings and tries to facilitate an open dialogue in which victims and offenders do most of the talking. Although mediators must be careful not to intervene too frequently, they must also be prepared to step in immediately if a victim feels unsafe.^[7]

In addition to the option to terminate the meeting at any time, victims have many choices throughout VOM. They decide when and where to meet, who should attend, and what restitution plan best meets their needs. In addition to reimbursement of monetary expenses, the victim may request that the offender do community service, perform a personal service, write a letter of apology, participate in treatment or rehabilitation programs, or complete some creative assignment. While the final restitution plan must be negotiated with the offender, the process puts much decision-making power in the hands of the victim.

While VOM sessions typically result in signed restitution agreements, the dialogue itself is the most

important part. Victims have an opportunity to express their feelings, speak about the impact that the crime has had upon their lives, and receive answers to important questions about what occurred. The opportunity to come face-to-face with their offender increases the likelihood that victims will receive an apology and gain emotional closure.^[2] Offenders, on the other hand, come to understand the real human impact of their behavior and develop empathy for victims, which can help to prevent future criminal behavior. Thus mediated dialogue can satisfy the needs of the parties and aid in the healing process without resulting in a written restitution agreement.

At one time, people questioned whether victims would really want to meet with their offenders face-to-face. However, studies show that the majority of victims who are given the opportunity to engage their offenders in dialogue choose to participate.^[7] Victims who meet with their offenders are also significantly less fearful of being revictimized and report that the opportunity to be actively involved in the justice process results in a heightened sense of emotional closure.^[2] Research also indicates that offenders who meet with victims are more likely to fulfill their restitution obligations and less likely to commit future crimes.^[6]

As a result, the American Bar Association has endorsed the use of VOM and dialogue in courts throughout the United States. In 2001, there were approximately 320 victim–offender mediation programs in the United States and more than 700 in Europe.^[6] Many of the programs work only with juvenile offenders and handle mostly cases involving property offenses and minor assaults. However, advocates have begun to challenge the assumption that VOM is not suitable for violent or sexual crimes. Its use in these more serious cases is becoming more widespread.

Family Group Conferencing

Family group conferencing (FGC) involves all the people most affected by the crime: victim, offender, and the family and friends of both. The practice originated in New Zealand and emphasizes the role of family and community in addressing wrongdoing, reintegrating offenders, and holding them accountable. The process encourages the offender’s support system to take collective responsibility for making amends and shaping the offender’s future behavior. In addition to family members, conferences typically involve teachers, peers, and community resource people. Because conferencing involves more community members than VOM, some believe that it can contribute more to the empowerment and healing of the community as a whole.^[5]

Prior to the conference, the facilitator contacts both victim and offender to explain the process and

invite them to a conference. This in-person meeting provides an opportunity for coordinators to build trust, provide information, and prepare parties for the conference should they choose to participate. If victims and offenders are not prepared, they may not feel safe nor be able to participate freely in a genuine dialogue.

Once both parties have agreed to participate, they identify key members of their support systems to accompany them. With the help of the trained facilitator, the parties discuss how they have been harmed by the offense and how this might be addressed. After the offender has described why the crime occurred, each participant speaks about the impact the offense has had on his or her life. The offender is thereby faced with the real human impact of his/her behavior on the victim, those close to the victim, and his/her own family and friends. After this discussion, the victim helps to decide the obligations that the offender must carry out. All participants play a role in determining how the offender might fulfill these obligations and repair the harm he/she has caused. The session typically ends with participants signing a restitution agreement.

Family group conferencing is now being used in many states in the United States, including Minnesota, Montana, Pennsylvania, and Vermont.^[6] It is used primarily with juvenile offenders and in cases of theft, arson, minor assaults, and vandalism. Police people, probation officers, or school officials often act as facilitators. The process is most often used as a diversion from the court process and can often provide a much quicker and more satisfying resolution than traditional criminal justice practices.

Community Reparative Boards

Community reparative boards are primarily used with adult offenders convicted of non-violent and minor offenses. A small group of citizens who have received intensive training meet face-to-face with offenders (and sometimes victims) to discuss what has happened. During meetings, board members constructively discuss the nature of the offense with the offender and point out its negative consequences. A chairperson guides members through their questioning of the offender and their discussions with the other participants. The board members then propose and discuss a list of proposed sanctions. The offender agrees to take specific action within a given time period to repair the harm he or she has caused. A reparative coordinator, usually a State corrections employee, is responsible for monitoring the offender's compliance. This process enables offenders to take personal responsibility for what has happened and actually give something back to the community. It also reduces costly

reliance on formal justice system processes. However, some criticize reparative boards for their minimal involvement of victims.^[6]

Circle Sentencing

The practice of circle sentencing grew out of traditional sanctioning practices of aboriginal peoples in Canada and Native Americans in the United States. The first modern program was implemented in Canada and its use spread to the United States in 1996, when a pilot project was initiated in Minnesota.^[6] It is a holistic integrative strategy that aims both to address the delinquent behavior of offenders as well as to consider the needs of victims and communities. Crime victims, offenders, family, friends, interested community residents, and justice and social service personnel all gather together to develop a better understanding of the event. The process tends to enhance respect and understanding among all involved and often produces innovative solutions.

Circle sentencing demands extensive pre-session preparation. To be admitted to a circle, offenders must petition the community justice committee, visit a respected community member for a conference, and begin work on a reparative plan. This plan typically involves some restitution to the victim and community service. The pre-session procedures act as a screening mechanism to ensure that offenders take the process seriously.

Circle sentencing typically involves a multistep process. First, there are separate healing circles for the victim and the offender. Next, there is a sentencing circle to develop a sentencing plan. After that, there are follow-up circles to monitor the progress the offender has made. Participants are expected to take responsibility for monitoring the offender's compliance with the sentence. Throughout the process, community support groups work to fulfill the needs of victim, offender, and community. A central function of these support groups is to protect victims, provide them a sense of security, and hear their stories.

A trained community member called a "keeper" initiates the dialogue and ensures the process is followed. Participants sit in a circle and pass around a "talking piece." When it is their turn to talk, they explain their feelings about the crime and express support for the victim and offender. Central goals of the process include promoting healing of all affected parties, empowering victims and community members by giving them a voice in the process, and providing an opportunity for the offender to make amends.^[6] Circles also help to build a sense of community and address the underlying causes of criminal behavior.

CONCERNS ABOUT IMPLEMENTATION

Restorative justice is not yet a complete model. It is not entirely clear how it can address certain difficult issues, such as disputes over culpability, consequences of failure to reach an agreement, equity and proportionality in outcomes, or failure of the offender to comply with outcomes reached.^[2] Also, it seems the notion of restorative justice has been developed largely within a white, male, North American perspective and has yet to be tested in diverse cultures and traditions.^[3] Finally, it is unclear how restorative justice meshes with larger questions of social, economic, and political justice.

To prove successful, restorative processes must be highly sensitive to victims' needs. Programs that are overly offender-focused may treat victims simply as tools for rehabilitating offenders. In addition, restorative processes can perpetuate power imbalances already existing between victim and offender. Especially in cases where victim and offender have a prior relationship, power and coercion may operate in these informal structures to revictimize the victim. Facilitators must be careful not to exhibit bias toward the offender and to take the harms that victims have experienced sufficiently seriously. They must also be prepared to intervene should the victim feel threatened or uncomfortable. Processes that fail to adequately address victims' needs may have damaging effects.

Perhaps the most common argument against restorative justice practices is that the process and outcomes vary across programs, so that comparable crimes are not "punished" equally. Instead, restorative justice approaches yield individualized responses to crimes. This lack of uniformity contradicts some of our basic ideas about fairness. However, some suggest that issues of "fairness" could be mitigated if restorative justice initiatives were implemented system-wide. Furthermore, it seems that any worries that restorative processes might lead to random, inequitable, and overly severe sanctions are misguided.^[8] In fact, the emphasis of restorative justice on healing and constructive goals suggests that such initiatives would actually have the opposite effect.

There is some debate about whether restorative justice should remain an unofficial alternative to the criminal justice system or whether such practices should be implemented system-wide. Some maintain that restorative justice is not likely to have lasting, significant effects if it continues to operate primarily at the grassroots level. However, others suggest that even if restorative justice principles cannot transform the justice system, they may improve current criminal justice policies. Still others worry that government agencies will bureaucratize such initiatives and strip them of creativity. Spontaneity and grassroots ties are often what make these initiatives so successful.

Some advocate a double system of punishment, in which offenders are first processed through the traditional system and then move on to informal restorative justice programs to agree to a reparative contract. However, this may mean that they end up being subject to more sanctions. In addition, few resources will be saved if restorative initiatives are only a supplement to traditional punishment or are used only for minor crimes. It seems that for restorative justice initiatives to have a forceful impact, they must be implemented system-wide.^[8]

Without official encouragement and support, restorative justice programs are likely to be concentrated in neighborhoods with relatively minor crime problems and to be used only with fairly minor offenses. However, the empirical data suggest that programs increasingly handle cases involving adult offenders, more serious crime, and disadvantaged urban communities.^[8] Many believe that these experiments in applying restorative processes should continue. This will help to determine whether restorative justice is applicable in cases involving spouse and child abuse, rape, and murder, and what sorts of safeguards are needed. Testing of new programs will require creativity, risk-taking, and hard work.

CONCLUSION

To be sure, a restorative justice approach is not a realistic response in all situations. In cases where offenders are dangerous, it may be necessary to incarcerate and forego restorative tactics. There are also cases in which victims may experience meeting their offenders as threatening. In cases of emotional vulnerability, for example when victim and offender are well known to each other, a court process with its formality and impersonal procedures may be preferable.

However, the restorative justice paradigm shows promise of offering crime victims more justice than they currently receive in the dominant criminal justice paradigm. Such initiatives provide "alternatives to punishment which offer possibilities for accountability, repair, and empowerment."^[3] While traditional justice practices tend to ignore victims' needs and treat crime as an offense against the state, restorative justice recognizes that crime involves real-life individuals. Rather than simply inflicting more harm via punishment, it seeks to heal individuals and repair communities.

REFERENCES

1. Marshall, T.F. *Restorative Justice: An Overview*; <http://www.homeoffice.gov.uk/rds/pdfs/occ-resjus.pdf> (accessed December 2003).

2. Strang, H. *Repair or Revenge: Victims and Restorative Justice*; Clarendon Press: Oxford, 2002; 48, 54 pp.
3. Zehr, H. *Changing Lenses: A New Focus For Crime and Justice*; Herald Press: Scottsdale, PA, 1990; 82, 221, 223 pp.
4. Hutchison, P.; Wray, H. *What is Restorative Justice?* <http://gbgm-umc.org/nwo/99ja/what.html> (accessed December 2003).
5. Umbreit, M. *Family Group Conferencing: Implications for Crime Victims*; Center for Restorative Justice and Peacemaking United States Department of Justice: Washington, DC, 2000; 1–10.
6. Bazemore, G.; Umbreit, M. *A Comparison of Four Restorative Justice Conferencing Models*; Juvenile Justice Bulletin United States Department of Justice: Washington, DC, 2001; 1–19.
7. Umbreit, M.; Greenwood, J. *Guidelines for Victim-Sensitive Victim–Offender Mediation: Restorative Justice Through Dialogue*; Center for Restorative Justice and Peacemaking United States Department of Justice: Washington, DC, 2000; 1–63.
8. Kurki, L. *Incorporating Restorative and Community Justice Into American Sentencing and Corrections*; Sentencing and Corrections 3 United States Department of Justice: Washington, DC, 1999; 1–11.

Revenue Forecasting and Revenue Estimation

Marilyn Marks Rubin

*Department of Public Management, John Jay College of Criminal Justice,
City University of New York, New York, New York, U.S.A.*

Nancy Mantell

Rutgers University, New Brunswick, New Jersey, U.S.A.

INTRODUCTION

Revenue prediction is a critical component of the budget-making process. Decision makers must know how much tax revenue they can anticipate in the upcoming budget period under existing tax laws, and how much revenue they can anticipate if there are tax law changes. This knowledge is especially critical for governments that are operating under balanced budget laws requiring equivalence between expenditures and revenues. Two types of revenue prediction are revenue forecasting and revenue estimation. Predicting baseline revenue collections under existing laws is known as “revenue forecasting.” Predicting the impact of specific tax law changes is known as “revenue estimation.” Revenue estimates are generally reported in “fiscal notes” accompanying specific proposed tax law changes.

RESPONSIBILITIES FOR REVENUE FORECASTING AND REVENUE ESTIMATION

Federal Government

In the federal government, both the executive and legislative branches prepare revenue forecasts and revenue estimates. In the executive branch, revenue forecasts and estimates are prepared by the Office of Tax Analysis (OTA)—in the Treasury Department’s Office of Tax Policy—for the Office of Management and Budget, the agency responsible for all executive branch budgetary matters.

In Congress, the Congressional Budget Office (CBO) prepares the baseline revenue forecast and estimates (“scores”) the revenue impacts of specific spending proposals. The House and Senate Budget Committees use CBO’s revenue forecasts to develop both their annual budget resolutions and their directives to other congressional committees. The Joint Committee on Taxation (JCT) staff is the official

Congressional “scorekeeper” in determining the revenue impact of proposed tax law changes. Congress will not consider tax law change legislation unless it is accompanied by a JCT revenue estimate.

State Governments

Among the states, the institutional responsibility for revenue forecasting varies. According to the National Association of State Budget Officers,^[1] in 17 states, the executive branch has the primary responsibility for developing the official state revenue forecast; in 22 states, a consensus process is used to arrive at the forecast. The consensus group generally consists of representatives of the executive and legislative branches; private sector and university representatives may also join the group. Many studies have indicated that consensus forecasting is quite advantageous in producing accurate forecasts.^[2] In the states, revenue estimates are presented to the legislature as fiscal notes appended to tax bills. The notes may be prepared by the executive agency responsible for revenue forecasts or by other agencies in the executive branch. In some states, legislative staff members may also prepare revenue estimates.

Local Governments

Information about institutional responsibilities for revenue forecasting in the more than 87,000 substate governments in the United States—of which close to 20,000 are classified by the U.S. Census as municipal governments—is primarily anecdotal. It indicates that the executive branch of government is generally responsible for the forecasts. However, in large cities such as New York, legislative staff may also prepare forecasts, and in some municipalities, academics or other consultants from outside the government may provide them.

REVENUE ESTIMATION AND FORECASTING METHODS

The four primary approaches currently used by governments to forecast revenues are 1) qualitative methods; 2) time-series methods; 3) econometric modeling; and 4) microsimulation modeling. Functional forms and examples of approaches 2 through 4 are shown in the appendix at the end of this article.

Qualitative Methods

Qualitative (or subjective) methods to predict revenues rely on expert judgment and opinions. Although few governmental units depend solely on judgment to forecast or estimate revenues, it is often used to inform and modify the results of more formal quantitative forecasting methods. A common application of the subjective approach is consensus forecasting in which information from several forecasts is used to arrive at a final revenue projection.

Time-Series Methods

In time-series methods, data on revenues in a past year or years are used to forecast revenues for the upcoming year(s). There are four general approaches to time series forecasting: 1) simple trend analysis; 2) the autoregressive approach; 3) the mixed technique approach; and 4) the approach that uses data on seasonal fluctuations of a revenue source to predict and adjust current year revenues. The first three methods are used to forecast revenues for the upcoming year(s); the fourth approach is used to predict revenues within a budget year based on collection data for the first few months of the year.

The simplest type of trend forecasting assumes that revenues in the forecast year will grow by the same absolute amount as in the past year, or by the same percent as in the past year (appendix, Eqs. 1 and Eqs. 2). Other trend forecasts assume that revenues in the forecast year will grow at the same average percent as in the past several years, or by the same average amount as in the past several years (appendix, 3 and Eq. 4). The latter method is equivalent to estimating a trend line.

More sophisticated time-series approaches to revenue forecasting use autoregressive techniques in which the prediction is generated by a weighted average of past observations going back “n” periods with a random disturbance term for the current period (appendix, 5). Many autoregressive models use regression techniques that incorporate corrections for autocorrelation of residuals and/or for moving average processes into the models. The Box-Jenkins ARIMA (autoregressive integrated moving average)

model is an example of this approach. Trend analysis may be modified to take rate or rule changes into account (appendix, Eq. 6). The historical revenue data can be adjusted to take the change into account, or if a time trend regression is used, the change can be explicitly included.

The main advantage of time-series models is their limited data requirements relative to other forecasting approaches (e.g., econometrics, microsimulation) because they require only a data series of the revenue being forecast. However, time-series models do not incorporate changes in the economic or demographic base underlying the tax structure, and this is seen as their primary disadvantage.

Econometric Models

An econometric model is essentially a mathematical representation of an economic phenomenon. The econometric approach to revenue forecasting differs from time-series analysis in that the former is based on assumed relationships among the variables rather than just on changes over time. An econometric model may consist of a single equation, a set of equations that are solved recursively, or a set of equations that are solved simultaneously. In a recursive model, forecast data for the independent variables are fed into the system; there is no feedback among the forecasts of the dependent variables; in a simultaneous model, there is feedback among the forecasts.

The simplest econometric model that might be used to forecast revenues posits that tax revenues in time period “t” depend on the tax base in “t” and the tax rate in “t.” The model involves only one equation (appendix, Eq. 7), and both the tax rate and the tax base are exogenous to the model. In a recursive system, there would be at least one equation besides the revenue equation, and the second equation would be used to forecast the tax base. There would be no feedback from the tax revenue to the tax base variables. In a simultaneous model, there would be feedback between the tax and economic variables. Although this type of system is more complex both to estimate and to forecast, it provides a more realistic look at the tax system than either a trend model or a recursive model.

Numerous issues must be resolved when using econometric modeling techniques, including forecasting the underlying economic variables included in the model, solving the liability/collection timing problem, defining the tax base, and adjusting for legislative changes during the modeling period.

The first issue relates to the fact that econometric revenue forecasting models are driven by the variables fed into them. Predicting these variables is especially difficult in uncertain economic times because many

are sensitive to, and reflect, rapidly changing macro-economic conditions. For example, among the macro-economic variables that feed into federal revenue forecasting models are wages and salaries, employment, and several business condition indicators, such as corporate profits—all variables that are highly sensitive to a changing environment. Subnational revenue models incorporate these national indicators, as well as variables that relate to the state and local economies, making them even more vulnerable to changing conditions. To the extent that the underlying forecast of the economic variables errs, so too will the revenue forecast.

A second issue in using econometric models is the underlying relationship between economic activity and tax collections. In econometric models, relationships are generally posited between economic activity and tax liability (i.e., how much tax is owed relative to the level of economic activity). However, economic activity and tax liability in a given year do not necessarily translate into tax collections for that year. Current tax year liability may reflect economic activity in the past year and/or accounting practices.

Connected to the liability-collection problem is the timing of collections and the proper temporal relationship between economic variables and collections. For example, when building an econometric model to forecast business taxes, there are a number of factors that confound the results. Corporations may not use the same fiscal year as the tax collecting entity. Corporations also have the ability to extend payment periods so that their economic activity in a given year may not relate to their tax payment in that year. These same factors may apply to personal income taxes although to a lesser extent.

A third issue in econometric revenue forecasting results from the lack of appropriate data to measure the tax base as defined in tax law. This is especially difficult for state and local governments. For example, lack of appropriate data is a major difficulty in forecasting sales tax revenues because the federal government stopped publishing regional retail sales data (usually a close approximation to the sales tax base) in 1996. Since then, it has been necessary to estimate state retail sales data prior to specifying the sales tax revenue forecasting equation. One way to do this is to “share” national retail sales data down to the state level. However, this methodology will produce inaccurate state sales data if state sales patterns differ widely from national patterns.

A fourth issue relates to changes in tax rates. The basic econometric forecasting model includes tax rate(s) and the tax base. If the rate remains stable, it can be excluded for purposes of revenue forecasting. However, if the model is used for estimating the impact on revenue caused by changes in tax rates (or tax rules), an assumption must be made

about how the change in the rate will affect collections. This will depend, to some degree, on the elasticity of the tax with respect to rate changes. Non-“rate” legislation such as caps, minimum taxes, and tax holidays are also considered, usually through dummy or other variables.

Microsimulation Models

The microsimulation model uses data from a probability sample of computerized individual taxpayer returns (or averages of individual taxpayer returns within tax brackets), along with a “tax calculator” to predict taxpayer behavior. This “tax calculator” is basically a computer program that manipulates the sample of tax returns using a set of input parameters that define the tax laws. For example, an equation to predict the income tax for a taxpayer in a specific tax bracket would be specified as follows:

$$\begin{aligned} \text{Tax revenue}(i) &= \text{tax rate}(i) \\ &\quad \times (\text{income} - \text{exemptions} \\ &\quad - \text{deductions}) \end{aligned}$$

where i is the tax bracket. There would be separate equations used for each tax bracket. The results for individual taxpayers in each tax bracket in the sample would then be weighted and applied to the entire set of taxpayers with the summation constituting the forecast of baseline tax revenues for the upcoming budget year. Unlike an econometric model, a microsimulation model is not estimated from historical data using a statistical method [e.g., ordinary least squares (OLS)]; thus, equations in microsimulation models do not include error terms.

Using the microsimulation model to estimate the impact of specific tax law changes can be accomplished in three ways: by changing the tax parameters, by changing the coding of the tax calculator, or by changing both. The tax calculator is changed by deleting certain code lines, by adding new code lines, or by doing both.

Some of the issues to be addressed when using a microsimulation model are similar to those that must be addressed in using econometric models [e.g., the temporal relationship between taxpayer liability and collections]. Taxpayers may not necessarily pay their taxes in the year for which the simulation is run. Other issues are unique to the microsimulation model itself. For example, the selection of the sample and the sample size will affect the outcome of the model. The use of large samples is expensive and time consuming, but reducing the sample size may reduce the reliability and validity of the results. The microfile must also be kept up to date with respect to the taxpayer base and must be modified to incorporate changes in the tax law.

Multiple Methods

Each of the four principal approaches to revenue forecasting and estimation may be (and is) used separately. However, given the limitations associated with each approach, many governments use multiple methods to forecast and estimate revenue more accurately. For example, New York City uses multiple methods to forecast its personal income tax revenues and a microsimulation model to forecast total tax liability. Withholding, estimated payments, and settlement payments are forecast using historical relationships, and estimated payments are forecast using an econometric model.

SELECTION OF FORECASTING AND ESTIMATION METHODOLOGIES

Federal Government

The OTA in the executive branch, and the CBO and the JCT in Congress, rely primarily on the Individual Income Tax Simulation Model (a microsimulation model) and the Corporate Tax Model (an econometric model) for revenue forecasting and estimation. When these models are not appropriate, others models such as the Treasury's Depreciation Model, as well as models for the estate tax and the foreign tax credit, are used.

The OTA's counterparts at the CBO and the JCT use similar models and estimating conventions. Their forecasts and estimates will, however, differ from those of the OTA to the extent that their underlying macroeconomic forecasts differ. The CBO's economic forecasts serve as the basis for the revenue estimates in the Congressional budget resolution; the CEA's macroeconomic forecasts provide the basis for the revenue forecasts in the President's budget proposals. When the forecasts differ, Congress can accept either because there is no formal mechanism for reconciliation.

Regarding estimation of specific tax law changes, during congressional consideration of tax legislation, the JTC provides the official revenue estimates. However, the executive branch may rely on Treasury estimates in deciding whether to support or oppose a particular proposal. Treasury estimates, when they differ from the JTC estimates, are generally not made public during the markup, but there have been situations when the OTA's numbers have been released.

State Governments

States use various methods to forecast revenues, with close to 40% relying on some type of consensus to make their final revenue forecasts.^[3] Time-series

analysis is generally used to forecast revenue for "non-economic" taxes, taxes whose collections are relatively stable, or taxes whose collections have grown steadily over time. Time-series analysis may also be used to forecast taxes for which there are no rate or rule changes in the forecast period. Where rate or rule changes have occurred, trend analysis may be used if the historical revenue data are adjusted to take the change into account. If a time trend regression is used, the change can be explicitly included.

In 1980, 10 states used complex econometric models for revenue forecasting and revenue estimation.^[4] By 1996, that number had increased to 34, primarily due to the availability of microcomputers that are capable of performing complex econometric equations. At least two-thirds of the states use econometric modeling, microsimulation, and regression analysis, or some combination thereof. Most states use three or four of these methods in some combination to forecast revenues. The methodologies that states use for revenue estimation are generally quite similar to those that they use for revenue forecasting.

Local Governments

Prior to 1999, there were no data available on the revenue forecasting methods used by cities. In 1999, the National League of Cities (NLC) included its first-ever question on revenue forecasting approaches in its annual survey of the fiscal conditions of U.S. cities.^{[5]a} The survey revealed that the predominant forecasting approach that cities use depends on the tax. For the property tax—the main tax revenue source for cities—46% of cities reported that they use only judgment to forecast revenues.

For the sales tax—the second most important tax revenue source for cities—56% of cities reported using only trend analysis to forecast revenues, whereas 21% said that they rely solely on judgment. For the income tax—the third major source of tax dollars for cities—45% reported using only judgment to forecast revenues, whereas 21% said that they relied solely on trend analysis. Local government reliance on subjective approaches for all taxes is not surprising given the paucity of data for local governments. In fact, some cities that have experimented with an econometric approach abandoned it primarily due to the unavailability of data.

^aThe NLC mailed questionnaires to 1056 U.S. cities, including those with populations greater than 50,000 and a random sample of 520 cities with populations between 10,000 and 50,000. The survey asked which of seven revenue-forecasting methodologies they used to forecast tax revenues. The choices were judgment, consensus, trend analysis, time-series analysis, econometrics, microsimulation, and other.

ERRORS IN FORECASTING

Because forecasts are made within a climate of uncertainty and are based on assumptions regarding this uncertainty, there is always the possibility of making a forecasting error (i.e., of under- or overestimating revenues). The magnitude of this error can be determined by measuring forecast bias and forecast accuracy.

Forecast Bias

Forecast bias is the extent to which the forecast differs from actual revenues. The accuracy of a forecast is the degree to which forecast values are narrowly dispersed around actual outcomes. The measure for bias most widely used is the mean error of the forecast that will equal zero over time if there is no bias. This measure is easy to calculate, but may be misleading because the mean error of two forecasts could be identical if, for example, one had small errors and the other had large, balancing errors. Other measures for bias include the calculation of a track ratio that keeps a running total of forecast errors divided by the absolute value of the average error. If the absolute value of this ratio exceeds a predetermined positive number, the model is biased. A third measure used to test for forecast bias is based on linear regression analysis of forecast vs. actual values.

Forecast Accuracy

Accuracy, the primary criterion for judging the performance of a forecast, will more clearly reflect forecasting performance than will the mean error. Two of the most commonly used accuracy measures are the mean absolute error of a forecast and the root mean square error. The mean absolute error is the average of the absolute difference between the actual and predicted values of the forecast over several time periods. The root mean square error is the square root of the average of the sum of the squared differences between the actual values and the forecasted values.

For all levels of government, the size of the forecast error tends to be larger when the economy is unstable than during periods of economic stability and is generally greatest at turning points. Among the states, most studies have found that revenues tend to be underestimated, thus providing a “cushion” for unforeseen economic events and/or expenditure demands.^[6] Research has also indicated that revenue forecast errors in states with a split government (where the leadership in the executive branch, senate, and house are not of the same party) are likely to be larger than errors made in states with a unified government (in

which the leadership in all branches is from the same political party).

Among local governments, the limited information available indicates that revenue underestimation is persistent and substantial.^[7] Results from the NLC’s annual survey provide corroboration of revenue underestimation by local governments. From 1987 to 1996, the approximately 300 governments responding to the NLC survey underestimated budgeted revenue changes in all years, except 1990 and 1991.

ISSUES IN PREDICTING REVENUES

Several problems associated with specific revenue forecasting and estimation techniques have already been discussed. The issues summarized below are more general and, except for the first—the political climate—relate to revenue estimation rather than to revenue forecasting.

Political Climate

There is a rather substantial body of literature on the impact of politics on public budgeting.^b Much less empirical research has been conducted on how politics affects revenue forecasting and estimation. Anecdotal evidence, however, suggests that at all levels of government revenue forecasting and estimation are subject to political considerations, especially with regard to the assumptions made in the underlying economic forecasts.

Indirect and Inducted Effects of Tax Law Changes

The direct impact of a proposed tax law change is what is generally considered in “scoring” tax law changes; rarely are the induced and/or indirect effects of proposed changes included due, primarily, to the complications involved in their estimation.^c At the present time, neither the CBO, the JCT, nor the OTA incorporate indirect/induced effects in their revenue estimates. Among the states, only Massachusetts regularly incorporates indirect and/or inducted effects into its revenue estimation process, and California requires that feedback effects be considered if the tax

^bIrene Rubin summarizes much of this literature in Ref. [8].

^cThere are four major approaches to categorizing these direct and indirect effects: 1) as partial equilibrium or general equilibrium effects; 2) as primary, secondary, and tertiary effects; 3) as microeconomic and macroeconomic effects; and 4) as static and dynamic effects. Most discussion about behavioral assumptions that has taken place in recent years uses the static–dynamic dichotomy, with dynamic effects sometimes referred to as “feedback” effects.

bill is over a specified amount. There is no empirical evidence available suggesting that local governments systemically include anything other than the direct effects of tax law changes in their revenue estimates.

Interaction Among Multiple Tax Rule Changes

Another controversial issue in revenue estimation concerns “stacking,” or the order in which tax law changes are entered into the estimate. For example, the 1986 Federal Tax Reform Act reduced marginal tax rates and broadened the tax base. When estimating the impact of the legislation, the OTA stacked base broadeners first and then rate cuts. Accordingly, the estimates for proposals to narrow the number and level of exclusions or deductions assumed that the existing marginal tax rates would apply to the broadened defined base. The JCT, in contrast, stacked the changes in marginal tax rates first and base broadeners second because most amendments to the tax code were to change tax rates rather than the tax base. Stacking the rate reduction first resulted in a lower revenue estimate than stacking the broadening first. There is no “correct” order in which to enter the proposed changes, but the selection may affect the revenue estimate.

Time Horizon

Still another controversial issue in revenue estimation relates to the time horizon used to calculate the budget impact of the estimation. On the federal level, revenue estimates are subject to the same 5-year time horizon as are all other budget items. This raises the issue of how to “score” changes in tax collections that will result from proposed tax law changes. Scoring is defined as the way in which budgetary actions, including those related to tax rule changes, are tracked. For example, suppose there is a proposed federal tax rule change that would exempt businesses from the corporate income tax starting in year six, if they paid an additional \$50,000 in taxes for years one through five. In years one through five, there would be an increase in revenue resulting from the tax rule; in year six and thereafter, revenues would fall. Because of the 5-year time horizon, the change in tax collections would be scored as a gain, and there would be no scoring for the loss in the later years.

An Intergovernmental Issue

If the federal tax rate structure changes, state and local government tax systems are not affected. However,

changes in the federal tax base may have a considerable impact on state and local taxes if they are “tied” to the federal structure (e.g., if the first entry on their tax form is federal taxable income). For example, when the federal government broadened the individual income tax base in 1986, states tied to the federal structure saw their tax bases broaden and would have experienced what was referred to as a “windfall” had they done nothing to prevent the pass-through. States with relatively high tax rates, such as New York, saw this windfall, not as a benefit, but as a threat to their already unfavorable position vis-à-vis states with lower tax rates. Federal government tax policy experts do not necessarily consider the impacts of such changes on lower levels of government.

APPENDIX A

Illustrations of Time-Series Analysis Equations Used in Revenue Forecasting and Estimation

Simple Trend Analysis

$$\text{Revenue}_t = \text{revenue}_{t-1} + (\text{revenue}_{t-1} - \text{revenue}_{t-2}) \quad (1)$$

revenue grows by the same amount as in the previous year.

$$\text{Revenue}_t = \text{revenue}_{t-1} \times (\text{revenue}_{t-1} / \text{revenue}_{t-2}) \quad (2)$$

revenue grows by the same percent as in the previous year.

$$\text{Revenue}_t = \text{revenue}_{t-1} \times \text{average}[(\text{revenue}_{t-1} / \text{revenue}_{t-2}), \dots, \text{revenue}_{t-n+1} / \text{revenue}_{t-n}] \quad (3)$$

revenue grows by the same average percent as in the previous several years.

$$\text{Revenue}_t = f(\text{time}) \quad (4)$$

revenue grows by the same average amount as in the previous several years. This is equivalent to estimating a trend line and using it to forecast with.

$$\begin{aligned} \text{Revenue}_t &= a + a_1 * \text{revenue}_{t-1} + a_2 \\ &\times \text{revenue}_{t-2} + a_3 \\ &\times \text{revenue}_{t-3} \dots + a_n \\ &\times \text{revenue}_{t-n} + e_t \end{aligned} \quad (5)$$

revenue is estimated using an autoregressive model that goes out for n periods and includes an error term.

$$\text{Revenue}_t = f(\text{time}, \text{tax rate}_t) \quad (6)$$

This is a modification of Eq. 4 to take changes in the tax rate into account.

APPENDIX B

Illustrations of Econometric Models Used in Revenue Forecasting and Estimation

Illustration of the simplest econometric model

$$\text{Revenue}_t = f(\text{tax base}_t, \text{tax rate}_t) \quad (7)$$

This equation says that tax revenues in time period t depend on the tax base in t and the tax rate in t .

An example of the simplest model used to forecast the Personal Income Tax

$$\begin{aligned} \text{Personal income tax revenue}_t \\ = \alpha + \beta * \text{personal income}_t \\ + \theta * \text{personal income tax rate}_t + \varepsilon \end{aligned} \quad (8)$$

α , β , and θ are estimated coefficients, and ε is the error term. The tax rate and personal income forecasts are exogenous to the model.

An example of a recursive model used to forecast the Personal Income Tax

$$\begin{aligned} \text{Personal income tax revenue}_t \\ = \alpha + \beta * \text{personal income}_t \\ + \theta * \text{personal income tax rate}_t + \varepsilon \end{aligned} \quad (9)$$

$$\begin{aligned} \text{Personal income}_t = a + b * \text{employment}_t \\ + c * \text{consumer price index}_t + e \end{aligned} \quad (10)$$

In addition to Eq. 7, there would be at least one other equation—that used to forecast personal income—but there would be no feedback from the tax revenue to income.

An example of a simultaneous model used to forecast the Personal Income Tax

$$\begin{aligned} \text{Personal income tax revenue} \\ = f(\text{personal income}, \text{personal income tax rate}) \end{aligned} \quad (11)$$

$$\begin{aligned} \text{Personal income} = f\left(\sum \text{wage rates}_i \right. \\ \left. \times \text{employment}_i\right) \end{aligned} \quad (12)$$

$$\text{Employment}_{\text{gov't}} = f\left(\sum \text{revenues}_t\right) \quad (13)$$

Personal income tax revenue Eq. 11 is a function of personal income Eq. 12 that, in turn, depends on employment Eq. 13. One employment sector is government employment; it depends, to some extent, on the size of tax collections. Thus, assuming positive relationships among revenues, income, and employment, if the tax rate is increased, revenues rise, which leads to an increase in government employment. This, in turn, results in growth in total wages and personal income, which leads to further increases in personal income tax revenues.

APPENDIX C

Illustration of a Microsimulation Model using the Personal Income Tax

$$\begin{aligned} \text{Average taxable income}_i = (\text{gross income}_i \\ - \text{deductions}_i - \text{exemptions}_i) / \text{taxpayers}_i \end{aligned} \quad (14)$$

where Exemptions_i = value of a personal exemption $\times$ average number of exemptions $_i \times$ taxpayers $_i$, and deductions_i = average deductions $_i \times$ taxpayers $_i$

$$\begin{aligned} \text{Avg. tax}_i = \text{tax base}_i + \text{tax rate}_i \\ \times (\text{average taxable income}_i \\ - \text{lower income bound}_i) \end{aligned} \quad (15)$$

$$\text{Tax liability}_i = \text{average tax}_i \times \text{taxpayers}_i \quad (16)$$

$$\text{Total tax liability} = \sum (\text{tax liability}_i) \quad (17)$$

In this model, there are three income classes, with corresponding tax bases, tax rates, and income bounds. When changes are made in the tax rate structure, or in deductions, exemptions, or credits, the effects on each income class and tax bracket (as well as on total

liabilities) can be predicted. The model can be used to forecast tax liability by income class, as well as to estimate the impact of tax law changes.

REFERENCES

1. National association of state budget officers. *Budget Processes in the States*; 1999.
2. Voorhees, W.R. The impact of political, institutional, methodological, and economic factors on forecast error. In *Dissertation*; Indiana University: Bloomington, IN, 2000.
3. Rubin, M.; Peters, J.L.; Mantell, N. Revenue forecasting and estimation. In *Handbook on Taxation*; Hildreth, W.B., Richardson, J., Eds.; Marcel Dekker, Inc.: New York City, NY, 1999.
4. Bahl, R.W. Revenue and expenditure forecasting in state and local governments. In *State and Local Government Finance and Financial Management*; Peterson, J.E., Spain, C., Lafferty, M., Eds.; Finance Research Center: Washington, DC, 1980.
5. Rubin, M.; Mantell, N.; Pagano, M. Approaches to revenue forecasting by state and local governments. In *Paper Presented at Annual Meeting of Association for Budgeting and Financial Management*; 2000.
6. Rodgers, R.; Joyce, P. The effect of under forecasting on the accuracy of revenue forecasts by state governments. *Public Adm. Rev.* **1996**, 56, 48–56.
7. Frank, H.A. *Budgetary Forecasting in Local Government*; Quorum: Westport, CT, 1993.
8. Rubin, I. *The Politics of Public Budgeting*, 3rd Ed.; Chatham House Publishers: Chatham, NJ.

Risk Management

David L. Baker

School of Public Affairs, Arizona State University, Tempe, Arizona, U.S.A.

INTRODUCTION

Public agencies operate in a risky environment. Risk relates to the chance of injury, damage, or loss.^[1] It represents a set of circumstances in which multiple outcomes are possible and the probability of each possible outcome can be estimated.^[2] Losses affect the human, financial, physical, and natural resources under the care of a public agency.^[3] Risk management seeks to conserve and to protect public resources from accidental loss. It proactively mitigates governmental risk through awareness, avoidance, assessment, containment, funding, and management oversight.

The following review discusses risks and defines exposures, perils, and hazards. It explicates six primary risk management concepts (risk awareness, exposure avoidance, risk assessment, risk containment, risk funding, and risk management). After highlighting macrotreatment and microtreatment strategies, common elements of governmental risk containment programs are described and the manners in which they minimize loss are examined. The review outlines the major tasks of risk management administration and introduces the notion of pervasive risk management. Following an appraisal of the organizational benefits arising from risk management, the review concludes with an overall summary of the topic.

RISKS, EXPOSURES, PERILS, AND HAZARDS

Two primary types of risk may be distinguished: speculative risk and pure risk. Speculative risk refers to situations where both loss and gain may occur. Pure risk involves unexpected loss without the opportunity of some gain. Public agency risk management deals with pure risk.

An understanding of exposures, perils, and hazards provides a common framework for analyzing pure risks.^[1] Exposures have to do with risk circumstances that can result in a loss. It refers to vulnerabilities resulting from a public agency's existence and its operation. There is no exposure if there is no likelihood of loss. Perils are defined as the actual causes of loss.^[4] For example, perils consist of death, injury, natural catastrophes, litigation, and illegal activity. Hazards boost the probability and/or severity of loss. Actions or

conditions create hazards. For example, faulty brakes on school buses are a hazard that increases the likelihood of the peril of injury, property damage, and litigation.

Risk management seeks to reduce or eliminate losses by reducing hazards. Consequently, it is useful to evaluate hazards by segregating them into three categories: physical, moral, and morale (or behavioral) hazards.^[1,4] Physical hazards are defined as conditions that increase the probability of loss. For instance, the absence of secure exterior doors on a public building constitutes a physical hazard. Moral hazards consist of dishonest behavior by individuals, either observed or unobserved, that affects the likelihood or magnitude of a loss^[5] (e.g., a public employee stealing gasoline from a public agency for the employee's personal vehicle). Morale hazard comes from personal carelessness. An employee recklessly driving an agency's vehicle denotes an example of morale hazard. Within this framework of understanding, the review now turns to explaining the core concepts of risk management.

RISK MANAGEMENT CONCEPTS

Six risk management concepts capture the underpinnings of a comprehensive risk management program for governmental entities. First, "risk awareness" concerns identification of public agency resources that could be affected in unintended, adverse ways and to the detriment of the public interest. Resources face natural events (e.g., monsoons, volcanoes, and earthquakes) and human actions (e.g., terrorism, fires, and lawsuits) with serious consequences for governmental entities. Many exposures present indirect impacts involving expenses to meet essential operations and service responsibilities.^[3] Wide-ranging exposure calls for persistent vigilance to discover and to address public agency risks.

Second, "exposure avoidance" means to avoid or to eliminate a given exposure.^[6] The goal is to remove the likelihood of loss totally. A completely avoided exposure requires no further resource investment for risk containment or risk funding. Consequently, public agencies often attempt exposure avoidance when they first discover vulnerabilities.

Third, after establishing risk awareness and attempting exposure avoidance, “risk assessment” encompasses the gray area between certainty (knowing exactly what will occur) and uncertainty (not knowing what will occur). It confronts Simon’s^[7] “bounded rationality” inasmuch as estimates are made about possible occurrences based on imperfect information.^[8] Risk assessment entails measurement of risk by analyzing past loss frequency and severity probabilities. However, it is forward looking.^[9] Exposures that present the greatest potential impact on an agency command the highest attention. Past loss history primes analysis as a public agency evaluates the extent of possible future losses and the relative likelihood of those losses.

Fourth, “risk containment” follows assessment. It refers to strategies designed to minimize risks and losses. This typically involves policy development and personnel training in many areas including security, safety, environmental protection, and emergency planning.^[3] Risk containment activities require a well-coordinated effort through a designated risk manager. The risk manager functions under a defined system of accountability involving all levels of management, supervision, and front-line employees. Risk management administration is discussed more fully later.

Fifth, some risk losses may be inevitable despite a public agency’s best efforts. As a result, part of a comprehensive risk management program involves “risk funding.” The imperative to preserve a functioning public agency in the face of a significant risk obligates a governmental entity to maintain adequate funds and funding approaches to cover incurred losses. This calls for a mix of tactics designed to shield public assets at the least cost. The “least cost spectrum” ranges from 1) some risk retention by way of self-insurance; 2) risk sharing through some form of multiagency pooling arrangements; 3) the means to borrow from a financial institution; and 4) risk transfer with the purchase of insurance products.^[3] Insurance solutions frequently provide a variety of intermixed methodologies to contain costs. For example, a public agency may be self-insured for the first \$250,000 per incident, have a pooling arrangement for losses involving over \$250,000 but less than \$1.0 million, and purchase insurance to cover losses that exceed \$1.0 million per incident.

Sixth, “risk management” combines governmental risk awareness, exposure avoidance, risk assessment, risk containment, and risk funding through a compelling, coordinated, and communicated effort to protect public assets. This usually involves various levels of policy development, training, staffing, organizing, and ongoing evaluation with periodic public reports to the community served. Conscientious risk management fosters the notion of “pervasive risk management.” This review introduces and explores that subject later.

RISK TREATMENT STRATEGIES

Risk treatment strategies commence with management fundamentals at the macrolevel. After analysis of an agency’s strengths, weaknesses, threats, and opportunities, management initiates a risk management vision. This usually includes a statement of exactly what the agency seeks to accomplish, identification of resource allocations to support the desired accomplishments, measurable outcomes, and timelines for monitoring activities. Periodic performance review should include a regime of internal and external critique.

Moving to the microlevel, risk treatment strategies generally include many implementing measures that flow from macrolevel activities. For example, successful public agency risk treatment strategies require; 1) briefing elected officials and public policy approvals; 2) communicating, coordinating, and educating the organization; 3) discussing and involving staff throughout the organization to identify and assess risk areas; 4) carefully constructing risk management policies; 5) implementing procedures and providing constructive assistance to organizational units in policy applications;^[3] 6) monitoring and evaluating reports to review compliance, tackle problems, measure trends, and exploit emerging opportunities; and 7) periodic external critique through timely audits and review of data from comparable jurisdictions known for outstanding practices and performance.

RISK CONTAINMENT PROGRAM ELEMENTS

Risk containment is a central focus of risk management. It meets the objective of minimizing risks and losses through initiating activities that positively alter the probability that a loss incident occurs. When an incident does happen, the goal is to reduce the severity of the loss and provide a means for recovery. These activities, or program elements, are designed and modified based on specific risk exposure and assessment analysis. Common elements of governmental risk containment programs include the following.

- **Security:** Protecting people, possessions/property, information,^[3] and the means to use information (information and communication technology) from illegal or unauthorized internal and external threats.
- **Personal safety:** Meeting all employee and citizen health and safety regulations,^[3] taking preventative measures, maintaining records of concerns and incidents, and training individuals as needed.
- **Equipment safety:** Developing and enforcing use standards on all equipment to assure safe operations. This may include checking weapons on law

enforcement personnel, needle-stick and lifting protocols in public hospitals, and specialized forklift training in governmental warehouses.

- Emergency preparedness: Developing contingency plans to address natural and environmental disasters as well as human-induced actions. Such preparedness aims to safeguard human, private, and governmental loss.

Beyond these risk containment program elements, there are several techniques designed to minimize loss.^[1] First, loss prevention hunts for ways to avoid harm before something occurs. For example, security guards may serve as a deterrent to theft. Second, loss reduction strives to cut the consequence of a loss. For instance, fire detection alarms and extinguishers may lessen a loss once a fire starts. Third, segregation relates to tactics taken to trim down potential harm. This is illustrated in the practice of insisting that several elected officials attending a conference use more than one aircraft for transportation. Fourth, and finally, duplication or redundancy protects an agency by assuring the availability of a replacement asset in critical situations should the primary asset be lost or malfunction. A case in point would be the provision for a backup generator to maintain computer services to law enforcement and fire suppression personnel during an electrical power shortage.

RISK MANAGEMENT ADMINISTRATION

Risk management programs require continuous direction and careful attention.^[3] The staffing component dedicated to the effort varies depending on organizational size, complexity, and available resources. Effective programs administratively focus on the following activities.^[1]

- Developing the overall mission and adaptive strategies, including periodic elected official briefings, generation of written policies, and constant performance monitoring and improvement.
- Defining responsibilities and organizing resources for risk awareness, exposure avoidance, risk assessment, risk containment, risk funding, and risk management.
- Communicating throughout the organization and to the public about the overall agency's challenges, objectives, and progress.

The job of risk manager requires a multistep process. The designated individual works as staff to the chief administrative officer in developing an organizational vision, mission, and objectives for the risk management effort. A comprehensive program is developed

and implemented to contain and to manage exposures. Internal self-audit as well as external evaluation hone practices and refine strategies while identifying areas for continuous improvement and modification.

PERVASIVE RISK MANAGEMENT

Risk management requires deliberate attention throughout governmental activities. This does not mean that risk adversity should stifle governance and delivery of services with creativity crushing inertia. It does suggest that risk management needs to be addressed as part of everyday governmental activities. The notion of "pervasive risk management" calls for integrating macrorisk and microrisk management thinking into the daily activities of government. It brings systems thinking to bear on risk management holistically. Proactive problem solving emanates from examining the context of the system where the concerns arise.^[10] Conserving and protecting public agency resources should not be an afterthought.

Public agencies need to push risk containment program elements back into program planning and strategic thinking at the earliest point possible to achieve risk management at the lowest cost. The pervasive risk management notion requires that the classical functions of planning, organizing, staffing, directing, coordinating, reporting, and budgeting^[11] consider all these activities with sensitivity to risk management.

RISK MANAGEMENT BENEFITS

Several public agency benefits arise from risk management activities.^[1,9,12] The more compelling involve; 1) control of insurance costs; 2) enhanced communication, coordination, and record-keeping; 3) active case, or claims, management; 4) finely tuned risk sensitivity and training to curb injuries and claim expenses; 5) more sharply focused insurance decisions; 6) heighten security for citizens, employees, and public agency assets; 7) minimal operational disruptions; 8) immediate resources to address some losses and funding strategies to address larger losses; and 9) more proactive administrative leadership and an enhanced organizational culture to support it. Individually and collectively, these benefits boost the effectiveness of a public agency while conserving and protecting public resources.

CONCLUSION

All public agencies function with some level of risk.^[13] Risk management seeks to conserve and to protect

public resources from loss. It does this through a vigilant campaign of awareness, exposure avoidance, assessment, containment, funding, and overall management. There are macrorisk and microrisk treatment strategies, and public agencies are well served through implementing effective security, personal safety, equipment safety, and emergency preparedness programs. Risk management administration critically focuses on risk management leadership, responsibilities, resources, and communications. Execution of a comprehensive effort involves the notion of pervasive risk management where systems' thinking about risk management permeates the organization. The many benefits of a sound risk management program boost the effectiveness of a public agency while guarding public resources.

REFERENCES

1. Hampton, J.J. *Essentials of Risk Management and Insurance*; American Management Association: New York, 1993.
2. Parkin, M. *Microeconomics*, 5th Ed.; Addison-Wesley: Reading, MA, 2000.
3. Phelus, D.G. Risk management: a developed discipline. In *Risk Management Today: A How-to Guide for Local Government*; Wasserman, N., Phelus, D.G., Eds.; International City Management Association: Washington, DC, 1985; 7–21.
4. American Institute of Certified Public Accountants. In *Guide to Risk Management and Insurance*; American Institute of Certified Public Accountants, Inc.: New York, 1992.
5. Pindyck, R.S.; Rubinfeld, D.L. *Microeconomics*, 5th Ed.; Prentice-Hall: Upper Saddle River, NJ, 2000.
6. Head, G.L., Ed.; *Essentials of Risk Control, Volume I*; Insurance Institute of America: Malvern, PA, 1986.
7. Simon, H. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*, 4th Ed.; The Free Press: New York, 1997.
8. White, M.J.; Clayton, R.; Myrtle, R.; Siegel, G.; Rose, A. *Managing Public Systems: Analytic Techniques for Public Administration*; University Press of America, Inc.: Lanham, MD, 1985.
9. Borodovsky, L., Lore, M., Eds.; *The Professional's Handbook of Financial Risk Management*; Butterworth-Heinemann: London, 2000.
10. White, D. Application of systems thinking to risk management. *Manage. Decis.* **1995**, 33 (10), 35–45.
11. Gulick, L. Notes on the theory of organization. In *Classics of Public Administration*, 4th Ed.; Sharfritz, J.M., Hyde, A.C., Eds.; Harcourt Brace College Publishers: Fort Worth, TX, 1997; 81–89.
12. Wasserman, N.; Phelus, D.G. Case studies in risk management. In *Risk Management Today: A How-to Guide for Local Government*; Wasserman, N., Phelus, D.G., Eds.; International City Management Association: Washington, DC, 1985; 159–168.
13. Borgsdorf, D.; Pliszka, D. Manage your risk or risk your management. *Public Manage.* **1999**, 81 (11), 6–10.

Risk Pools

Peter C. Young

Graduate School of Business, University of St. Thomas, Minneapolis, Minnesota, U.S.A.

INTRODUCTION

The emergence of pooling may be one of the more significant developments in the field of American public sector risk management (pools also exist in Japan, Australia, and a handful of other countries). Pools developed in the United States in response to two episodes of insurance unavailability and unaffordability in the 1970s and 1980s. Since that time, they have grown in size and number to a point where they are a dominant provider of risk financing for local governments. Further, they have had a profound impact on the practice of risk management in, particularly, small local authorities.

This article introduces both the concept and practice of pooling. After briefly discussing the concept of pooling, a history of pooling is presented, followed by a description of pool forms and functions. The article concludes with an overview of practical issues, challenges, and problems facing pools today.

POOLING AS A RISK MANAGEMENT TOOL

The idea of pooling is rather simple in one sense, and quite technically complex in another. In the simple sense, pooling is a means of protection; if a group of individuals pool their time and efforts, they can provide for a common cause that is more effective than could be provided through individual initiative. This is an example of *resource* pooling, which commonly has the goal of extending capabilities.

In the more complex sense, the pooling of *risks*, although offering some of the capacity-expanding features that resource pooling provides, focuses more on the matter of risk or uncertainty reduction. Pooling can improve predictability, stabilize budgeting and forecasting, and reduce fear and worry. The subtle distinctions between risk and resource pooling are not always visible to the naked eye, and in fact, they overlap to a certain degree. However, readers need to recognize that two distinct things can occur in pooling: 1) the pool participants can achieve more than is possible as individuals and 2) uncertainty and risk can be reduced.

Risk pooling almost invariably is a means of financing risk and in this way is related to the concept of

pooling in insurance. Relying on actuarial principles, risk financing pools can structure financing arrangements to provide a stable flow of funds into the pool—funds that are adequate to cover the sporadic outflows that occur due to losses.

Resource pooling tends to apply to a broader range of risk management (indeed, general management) activities where capacity enhancement is more critical than is risk reduction. Mutual aid agreements, group purchase programs (for school books, insurance, equipment, and machinery), and regional cooperatives all represent activities that seek benefit from enhancing capabilities, not necessarily from reducing uncertainty.

Interestingly, risk pools mainly began as risk financing pools, but have been evolving into risk management pools. Or, to put it in the language of the preceding discussion, pools initially focused on risk pooling, but have begun to broaden their scope of activity into resource pooling.^[1]

HISTORY OF RISK POOLING IN THE UNITED STATES

Since 1970, state and local governments have experienced two “insurance crises,” the first occurring in the mid- to late 1970s and the second in the mid-1980s. These downturns in the insurance market were the result of cyclical market problems, the state of the economy preceding each episode, insurance industry management practices, and inherent insurance pricing problems. The basic riskiness (or perceived riskiness) of the insured—state and local governments—also was a factor. The result has been a 25-year period of time in which state and local governments have been able to afford commercial insurance (particularly liability and workers’ compensation insurance) only intermittently, if insurance has been available at all.^[2]

In response to this problem, local governments, government associations, and even state governments acted to create alternative risk financing mechanisms to provide coverage for local governments. These mechanisms, commonly referred to as self-insurance pools, enabled local governments within a state to pool together risks and resources to finance the costs of fortuitous losses. Initially, the pools provided mainly liability and workers’ compensation coverage for

members. However, pools today provide a wide range of coverages, ranging from property coverage to employee benefits.

A 1988 study of pools found that approximately 200 (see discussion below on counting pools) public risk financing pools operated in the United States.^[3] Those pools, with an estimated aggregate membership of between 15,000 and 20,000 entities, were found in almost all regions of the country—although over one-half of the pools were located in California. Almost all pools were found to be homogeneous as to membership; that is, school districts pooled with school districts, counties with counties, and so on. The pools also tended to be homogeneous as to member population/size. Almost all pool participants could be characterized as small local governments. The pools themselves, however, varied widely in terms of size; the smallest pools reported three to five participating governments, with the largest reporting 1100 to 1300 members.

In 1994, a follow-up study was conducted to find out the degree to which the pooling environment had changed.^[4] It had, dramatically. The findings indicated that there were at least 400 pools in operation, and that these pools had an aggregate membership of 30,000 to 35,000 local entities. Pools still remained the preserve of small local authorities, with membership consisting largely of entities in communities with populations under 10,000.

By 1994, pooling had spread throughout the United States and California no longer dominated the pooling world in quite the same way. Over one-half of the pools were sponsored by associations (e.g., leagues of cities, counties, school board associations), and the estimated aggregate contributions were believed to be about \$5.5 billion.

Since 1994, evidence suggests that pool growth has moderated. No updating studies have been undertaken as of 2001, but experienced observers suggest that the number of pools has plateaued just above 500. (There is some controversy about how pools are counted because some pools are “clusters” of funds—for instance, a pool consisting of four individual coverage programs with segregated funds could be identified as one pool or four pools. Using clusters as the basis for counting, there probably are about 250 pools).^[5] The reasons for the slow down in expansion is identified below.

Looking back on the history of pooling since 1970, we can say that pooling is a significant public sector innovation. Some analysts believe that pooling engages nearly one-half of the local governments in the United States, and may have become a \$6 to \$7 billion “industry.” Given this significance, it is important to look more fully at pooling forms, functions, and practices.

POOLING TODAY

Legal Forms of Organization

Risk financing pools were difficult, if not impossible, to form until state legislatures took steps to permit a general exemption to the prohibitions on the comingling of public funds. The state of California established a “joint powers authority” provision in its state statutes, which had an immediate and profound effect—by the late 1980s, over one-half of the 200 pools in existence were in California. Because this was the case, the joint powers authority model (or sometimes an interlocal cooperation act) was adopted by many other states.

Consequently, nearly 60% of all pools are organized as joint powers or interlocal cooperation act agreements. Other approaches are used, however. Over 20% of pools are legally organized as non-profit trusts, whereas another 5% assume a corporate form. A few are legally constituted as mutual or reciprocal insurance companies, or as risk retention groups.

Sponsorship

Related to the matter of legal form is the question of “sponsorship” or pool ownership. About one-third of all pools are organized by a sponsoring association, such as a state municipal league or association of counties. Of the non-association pools, the majority were formed by voluntary collaboration among member entities, although there also are instances of specific individuals or organizations (e.g., a local insurance agent) serving as a catalyst for pool formation.

Pool Types and Functions

Conceptually, one might imagine that pools could approach the task of risk financing in several different ways. A pool could be organized to do just one thing—say, provide risk financing for workers’ compensation risks. Or a pool could offer several types of coverages; for example, general liability, property coverages, and police officers’ liability protection. Single coverage pools are referred to as “monoline” pools and multiple coverage pools are “multiline.” Due to enabling statute language in many states, some pools offer multiple coverages, but must offer them through distinct and separate pooling arrangements. For instance, a pool management organization might have responsibility for a workers’ compensation pool, employee benefits pool, and a property risk pool. Such arrangements are known as “fleet pools.”

A slightly different way of looking at pooling types is to consider the mechanics of the risk financing arrangements. Most commonly, pools are structured

on a risk-sharing basis, which is to say that they operate more or less like an insurance company. Funds are collected and pooled, and losses are paid out of those pooled funds. The risk sharing may occur on a primary-level basis (all losses, small or big, are financed) or on an excess-level basis (only large losses are financed). Other pools, although admittedly not many, are purely group insurance buying pools. They only pool premium dollars used to purchase commercial coverages for members. Perhaps not surprisingly, many pools are risk financing at the primary level, and insurance buying at the excess loss level (roughly approximating the reinsurance arrangements that insurance companies rely on).

Other pools, although again not many, assume what is called a banking form. They do not commingle funds, but maintain separate accounts for members. They can loan funds to members if losses are high, but they are not risk-sharing arrangements in the conventional sense of the term. Finally, a handful of pools are not involved in risk financing, but provide administrative services to members.

It should be noted that another form of pool has emerged in recent years, the “risk management pool.” This form engages the pool as the risk manager for its member entities. In that capacity the pool not only provides risk financing arrangements, but also provides risk assessment, control, and management services. The risk management pool, in some sense, appears to be an evolutionary stage for pools—meaning that many risk financing pools realized that successful pooling required that efforts be undertaken to manage the underlying risks. It must also be said that the move toward broader risk management services has been somewhat spurred by customer demand and also by competitive pressure (from the commercial insurance industry).

Before departing the subject, a recent development is the emergence of excess pools—really, “pools for pools.” Such pools provide varying forms of large-loss protection for pools in a manner similar to excess insurance and reinsurance. The Schools Excess Liability Fund in California, the County Reinsurance Limited, and the National League of Cities’ National Public Entity Excess program are examples of this new innovation.^[5]

Pool Governance

Almost all pools have a governance structure that is organized around a board of directors/chief executive form of organization. The boards, which typically consist of representatives of pool members, are elected boards about one-half the time and appointed boards in about one-quarter of the cases. Remaining pool boards are partially elected and partially appointed.

The relationship of the board to the pool varies in numerous key ways, but the distinction between sponsored and non-sponsored pools suggests one way in which board/pool relationships may differ. Non-sponsored pool boards derive from the membership and are more or less directly responsive to members. Sponsored pool boards also derive from membership, but may also derive from the sponsoring organization. Frequently, for instance, the chairperson of a pool board is the executive director of the sponsoring organization (a league of cities or an association of counties, for example).

Abstractly, the issue of conflict of interests or purposes arises in a sponsored pool setting. Association interests and member interests (even pool management interests) do not always coincide, and the governance processes by which these conflicts are resolved is not widely known or understood. As just one illustration of this point, risk financing pools rely on the principle of rate and risk relationships—that is, higher risks pay more for coverage than do lower risks. This rate “discrimination” is a convention of pooling, but often associations view pools as a “benefit” of association membership, suggesting that members should not be discriminated against.^[6]

Pool Management

The chief executive reports to the pool board, but otherwise has responsibility for the management of the pooling operation. This includes staff management.

There is a wide degree of variation and size in pool staffs. Research suggests that, on average, pools have a staff of six. However, the largest pools may have a staff of 100, whereas the smallest may have only a single part-time administrator. Size of staff is predicated not only on pools’ size, but also on the degree to which functions are outsourced or kept in-house. Historically, pools have relied heavily on third-party providers to supply services, such as:

1. Claims management
2. Underwriting
3. Loss control
4. Actuarial and financial audit
5. Legal counsel
6. Investment management
7. Banking services
8. Brokerage
9. Management consultancy
10. IT management
11. Sales and marketing

However, as pools have grown in size and experience, there has been a tendency to bring such functions

inside the pool, meaning staff size expands. Increased control or the hope of reduced costs are cited as the primary motivations for such moves. However, there is also some evidence that a few pools are being managed by brokers, agents, or outside management firms—in a sense, pool management has been completely outsourced.

Coverages Provided

There are two ways that one might understand the different coverages provided by pools. First, pools could be sorted out according to the range of coverages offered by individual pools. In this regard, findings would suggest that most pools offer quite limited ranges of coverages. Monoline pools, by definition, offer a single line of coverage—say, workers' compensation. Fleet pools, also by definition, are collections of pools that each offer a single line (or, occasionally, a narrow cluster of lines) of coverage. Even the multiline pools tend to focus on a rather narrow band of related coverages. It is the rare pool that offers an extremely wide assortment of coverages.

Second, one could ask what is the range of coverages offered by the pooling community as a whole, and the answer would be impressively wide ranging. For example, in the 1994 study, all the following lines of coverage were found to be offered by at least one pool:

1. Building and contents
2. Inland marine
3. General liability
4. Auto property and liability
5. Crime
6. Employee dishonesty
7. Workers' compensation
8. Unemployment compensation
9. Boiler and machinery
10. Public official's liability
11. Law enforcement liability
12. Medical malpractice/hospital liability
13. Mass transit property and liability
14. Liquor legal liability
15. Airport liability
16. Umbrella liability
17. Excess loss
18. Reinsurance
19. Pollution
20. Criminal defense reimbursement
21. Land use takings

As a point of reference, the most frequently provided coverages (in order) are general liability, auto property and liability, building and contents, public official's liability, and inland marine and workers' compensation.

Pool Membership

Estimates suggest that 30,000 to 35,000 local authorities participate in pooling; a significant number representing around 40% of all governmental entities in the United States.^[7]

Also pooling is a small government phenomenon. Fully 94% of pool participants are under 10,000 in population. Having noted this, it is also true that many entities with populations up to 100,000 have participated in pooling. So, perhaps it is somewhat more accurate to say that pools have proven useful for local entities with populations up to 100,000, but that, for several reasons, smaller entities have a greater likelihood of participate in pooling.

Pool memberships almost always are homogeneous as to state, member type, and member size. Although some experimentation has begun to occur with interstate arrangements, the authorizing legislation, as well as practical and political considerations, limits most pools to single-state operation. Because homogeneity is such a critical criteria for pooling (simply put, the less alike pooled risks are, the more likely the inequities and aberrations), one should not be surprised to find that most pools consist of similarly sized members of a common entity type.

Surveys show that pool membership in the United States includes municipalities, counties, townships, school districts, and special districts, such as county appraisal districts, housing authorities, municipal utility districts, regional special education consortia, port authorities, hospitals, special services cooperatives, water districts, drainage districts, zoological societies, nutrition councils, river authorities, and, most recently, non-profits such as transit and hospital organizations.

The pools themselves reveal a wide range of sizes with respect to membership numbers. The smallest pools have three to five members, whereas the largest are approaching 4500 members.^[5]

Pool Financial Management

It is estimated that member contributions to pools are approaching \$7 billion. Given such a sizable level of economic activity, it is useful to understand the basic mechanics of pool financial management.

Rate and Premium Development

Historically, pool rates were derived from insurance industry data (the Insurance Services Office, a rating bureau, in particular). This was natural because pools had no past data on which to estimate claims, and also because the consulting actuaries who assisted pools in rate development had knowledge of and access to such

information. Over time, however, pools have demonstrated a strong move away from insurance industry-based pricing, and now rely to a larger degree on their own loss experience.

As a point of reference, insurance premiums consist not just of the expected losses that will be paid, but also of the expenses and overhead charges necessary to administer the insurer. Pool contributions also contain expense loadings, but it is also important to realize that the components are not identical to insurance company expense loadings. The principal differences are: 1) most pools do not have to pay federal or state income taxes; 2) most pools do not have to pay insurance premium taxes; 3) most pools do not have to make participation commitments to state guaranty funds, high-risk pools, or other arrangements mandated by the state insurance commissioner; and 4) most pools do not have to make allowances for agent and broker commissions. These differences are not inconsequential and could, everything else being equal, account for 10% to 20% spread between a pool's premium and a commercial insurance company's premium. The premiums charged to members are payable annually (in over one-half the cases), or monthly, quarterly, or semiannually (in nearly one-half the cases).

Assessments and Refunds

In conventional insurance companies, the premium that is paid is the sum total of the insured's consideration under the contract. However, some cooperative (or "mutual") types of insurers reveal an assessment feature to their pricing scheme, which means that the insurer can require additional contributions to the pool if losses exceed some predetermined level. Assessability may be limited or unlimited.

Most pools are assessable pools (over 8 out of 10); of those assessable pools, nearly 70% have unlimited assessment features.

Refunding, rebating, credits, and dividend distributions are other means by which premiums paid may be adjusted. A majority of pools use one or more of these tools to offset the contribution costs if loss experience is favorable. Another widely used measure is discounting, whereby the pool discounts that initial contribution based on the supposition that the pool will be able to prevent losses from rising to a level that would justify the undiscounted contribution.

Other Financial Management Issues

In 1992, the Governmental Accounting Standards Board (GASB) implemented Statement 10, which serves as a broad directive on pool accounting practices. The GASB 10 (as it is called) is almost universally

adopted as the basis for issuing financial reports today, which has gone a long way toward standardizing pool reporting and is beginning to allow for greater transparency of pooling performance.^[8]

The GASB 10 has not only promoted standardization, but also has naturally had a significant effect on how pools reserve for losses, how they report losses, and on matters related to investment returns, discounting, maintenance of surplus/contingency funds, and other financial issues.

Because most pools are not taxable entities, financial analysis of pooling performance is not influenced by state or federal income tax considerations. This "advantage" over commercial insurance companies is somewhat offset by the restrictions that exist for investment of funds. Although insurance companies are subject to regulatory restrictions with respect to the riskiness of investments, public entity restrictions are even more stringent, and the standards that apply to public sector investment practices apply almost universally to pools.

Finally, a word should be said about alternative risk financing schemes. The phrase "alternative risk financing" used in relation to pooling is rather ironic because pooling is a classic alternative risk financing scheme, when compared with commercial insurance. However, in recent years, methods of financing pools have begun to alter somewhat. Most frequently (and an irony in its own right), pools have begun to purchase more commercial insurance, typically in the form of excess loss and reinsurance coverages. Less frequently seen alternative approaches include debt and banking arrangements, such as letters and lines of credit, bond issues, and finite risk insurance programs.

POOLING: A LOOK AHEAD

Pools now are well into their third decade of existence, and the general evidence suggests that they will continue to be an important presence in the world of public sector risk management. But this is not to say that there are no difficulties or challenges.

Pool administrators naturally argue that their pools are doing better than commercial insurers who operate in the public sector market. Their explanations seem to suggest the heart of the matter for pools. The overriding perception is that pools are more effective than insurance companies because the pool members are the owners and have greater control over the pool than they would over an insurer. In addition, administrators argue that pools understand the public sector better because they are "of the public sector." True proof of pool superiority will ultimately be established through financial performance. To date, no research has been undertaken to explore this issue. Arguably,

the adoption of the GASB 10 will ultimately have the effect of allowing for such analysis, but it has not yet been done.

When looking to the future, pools tend to report concerns about increasing competition from commercial insurers, challenges to maintain positive member relations, and possible legislative or regulatory changes that could adversely effect the ability of pools to operate as they presently do. More generally, many observers are concerned about the future of pooling in light of a re-emergent commercial insurance marketplace. Does the government have a role competing with private business? Although the answer at present seems to be “yes,” analysts agree that the nature of that market-based relationship is not known at the present time. Pooling likely will continue, but there are many unanswered questions that remain.^[9]

One point that should not be overlooked when considering the role/legacy of pooling is their impact on risk management practices in the public sector. At previous points in this entry, the concept of a risk management pool has been cited as both an abstract idea and as a trend in pooling practices. Certainly, an argument can be made that small local authorities can benefit greatly from the introduction of risk management practices, as well as risk financing services. Certainly, it seems that resource pooling can enable such small entities to develop fairly sophisticated practices—well beyond their individual abilities to do so. But is this actually happening?

The answer seems to be “yes.” An investigation of pools today shows that a diversification of risk management services is the most obvious event occurring within the pools. These services range from training and education to the promulgation of manuals, workbooks, risk checklists, and other risk control tools, to political lobbying to state legislatures, to resource pooling, to strategic planning, and a host of services intended to make members better managers of risk.

One innovation, not widely seen, but harkening back to the prepooling era is the circuit-riding risk manager, and some pools have adopted this practice as a way of getting a “ground-level” risk management presence in the member entities.

Regardless of the specific innovations, there is little doubt that pools have introduced risk management into thousands of local authorities that otherwise would not have had such an exposure. Even were pooling to disappear over time, it is likely the imprint of the pooling movement would remain.

REFERENCES

1. Young, P.C. Local government risk management pools: theory and practice. In *Unpublished Doctoral Dissertation*; University of Minnesota, 1988; Chapter One.
2. Young, P.C. Local government risk financing and risk control pools: Understanding their form, functions, and purposes. *Public Budg. Finance* **December, 1989**, 9 (4).
3. Young, P.C. *Risk Pooling: Scope and Practices—1987–1988*; Public Risk Management Association: Arlington, VA, 1988.
4. Young, P.C. *1993 Pooling Survey: Scope and Practices*; Public Risk Management Association: Arlington, VA, 1994.
5. Association of Governmental Risk and Insurance Pools (AGRIP). *Membership Statistics*; AGRIP Publishing: Prague, OK, 2001.
6. Young, P.C.; Fone, M. An evaluation of public risk financing pools: financing local government risks in a post-municipal mutual world. *Public Risk* **October, 1996**.
7. Young, P.C.; Reed, B.J. Government risk financing pools: assessing current practices. *Public Budg. Finance* **Winter, 1995**, 16 (3).
8. Young, P.C. *The GASB 10 Primer for Pools*; Risk Management Research Services: Minneapolis, MN, 1991.
9. Fone, M.; Young, P.C. *Public Sector Risk Management*; Butterworth Heineman: Oxford, 2000; Chapter Nine.

Rulemaking

Cornelius M. Kerwin

American University, Washington, District of Columbia, U.S.A.

INTRODUCTION

The official definition of rulemaking found in the Administrative Procedure Act (APA) is admirable in its simplicity, but does little to convey the weight and substance of this most critical element of the American public policy process. The act, at section 551, states that “rulemaking means agency process for formulating, amending, or repealing a rule.”^[1] More instructive is the definition of “rule” that is contained in the subsequent section of the APA: “rule means the whole or part of an agency statement of general or particular applicability and future effect designed to implement, interpret or prescribe law or policy.”^[2] We begin to comprehend the role of rulemaking in governance when we understand this definition grants rulemaking the authority to act on our law with a great range of impact.

RULEMAKING

When implementing law, rules provide nothing more than vehicles to carry the content of statutes to those affected by their provision. Rulemaking in these cases simply builds the conduit. Interpretation of law in a rule or regulation suggests a more substantial effort because it suggests that the language in the legislation is adapted—in some cases stretched and in others contracted—to achieve the intent of the Congress of the United States, in light of the circumstances that the agency perceives in the affected populations or activities. Rulemaking in such instances involves more judgment and knowledge of the external environment than simple implementation. In those cases when agencies prescribe law or policy, they are acting to transform general goals and authority granted them in a Congressional statute into specific requirements. Rule making in these cases fills a largely clean slate, at least considerably more so than when they implement or even interpret. But what is important to note is that in the case of interpretation and most certainly in the case of prescription, the rule-making process effectively supplements or substitutes for the legislative process. This fact is the root of both the genius of rule

making as an instrument of public policy, and the concern in some quarters that its mere existence poses a fundamental threat to the integrity of our democratic system.

DELEGATION OF RULE-MAKING AUTHORITY

Rulemaking enables the existence of a democratic public policy process on an American scale of size and complexity. It is so central and necessary that no less an authority than Kenneth Culp Davis has described a particular form of rulemaking as “one of the greatest inventions of modern government.”^[3] Without the capacity to “implement, interpret and prescribe” law and policy housed in institutions outside itself, Congress would be unable to respond effectively, if at all, to the public’s desire for action. If, as some would have it, our elected representatives were required to hone every law they enact into a form that could be put in effect with no subsequent adjustments by responsible officials in executive branch or independent agencies, there would be precious few laws of any import.^[4] One example will illustrate the magnitude of the actual law-making task. When it was last amended more than a decade ago, the Clean Air Act was widely considered one of the longest and most detailed pieces of legislation ever crafted by Congress. However, once enacted, the Environmental Protection Agency was required to write literally hundreds of rules and regulations to give more precise, predictable meaning to the language of the statute for enforcement staff and regulated parties. This pattern, if not the scale of the rule-making task, is repeated with virtually every law enacted by the Congress.

Rulemaking allows our elected representatives to respond, with varying degrees of specificity, to the will of the people, routinely expressed in a blizzard of expectations presented to them by constituents and supplicants every term. The process by which Congress transfers the law-making power from itself to executive or independent agencies is “delegation.” When Congress delegates, it does so for reasons other than the sheer crush of public and private calls for legislative action. The technical complexity of the issues and

problems that are the object of Congressional legislation—ranging from stem cell research to the intricacies of telecommunications—routinely pushes decision makers to the edge of human knowledge and comprehension, or beyond. The necessary research and analysis is largely entrusted to the career civil service and contractors they manage, and it is the results of these inquiries that ultimately become the substance of rules and, therefore, our law. With workload and complexity, Congress also transfers political pressure. It has been long hypothesized that Congress frequently delegates responsibility for decisions to agencies when their consequences entail considerable political risk. When controversial issues are completely or partially unresolved by Congress and inherited by those who write rules, the affected interests do not disappear. They move to the agencies, monitoring their actions and drawing on a wide range of tools and tactics to influence the outcome of rulemaking. As we will note below, when the rulemakers issue their decisions, they are subject to oversight by the very Congress that initiated the process. So, when delegating the authority to write rules, Congress transfers a great deal more than a heavy workload.

The status of rulemaking as a device for public policy is rivaled in importance only by the profound threat it poses in the minds of some to the very fabric of our constitutional democracy. It is certainly true that Article 1, Section 1 of the Constitution of the United States, the first pronouncement made by the founders about the structure and process of our particular democracy, states that “All legislative powers herein granted will be vested in a Congress, consisting of a House of Representatives and a Senate.” The reader will note that the Founders did not add to the components of the Congress, for example, the Nuclear Regulatory Commission or the Environmental Protection Agency. So, the Constitution would appear to foreclose the process of delegation and the vesting of the law-making power in the form of rulemaking. Clearly, what is missing to the critics of rulemaking is the strong and direct link between the writing of law and the people affected by it that is provided by the electoral process. There is no doubt that the Founders considered lawmaking the most important power and vested it in a body at least somewhat resistant to the suspected antidemocratic tendencies of executives and judges known at the time. Obviously, the Republic has survived with rulemaking firmly in place for most of our history, and the Supreme Court, despite the apparently absolute language in the Constitution has, on all but one occasion, rejected arguments that this form of delegation is illicit and a corrosive violation of basic democratic principles. We survive as a democratic system and rulemaking has emerged from the harsh light of the judicial scrutiny because the

Congress, more than a century and a half after authorizing rulemaking for the first time, fashioned an acceptable process to write rules that Professor Davis contends “can be, when the agency so desires, a virtual duplicate of the legislative process.”^[3] That approach to rulemaking is now embedded in the basic architecture of the administrative process.

THE RULE-MAKING PROCESS

Before turning to that process it is important to review what rulemaking currently produces to convey the magnitude of its importance to our political and legal system. The traditional way to measure the extent of rule-making activity is to simply count the number of rules published yearly in the *Federal Register*. The *Federal Register* was created by Congress to serve as a daily compendium of actions taken by agencies of the federal government.^[5] By this yardstick, rule-making activity has waxed and waned somewhat over the past, but the volume has been high by any standard. For example, the high water mark for final and proposed rules, until the end of the Clinton administration, was in 1980, the final year of the Carter presidency. In that year, more than 13,000 rules appeared in the *Federal Register*. Since that time, the fewest number of rules published was in 1997. Throughout the period, the number has mainly remained in the 7000 to 8000 rule range. The number of rules written by agencies in a given year tells us little about the impact they have on the lives and activities they affect. Among other things, rules establish the requirements that are enforced in our many and varied regulatory programs. The actual costs and benefits arising from the operation of regulatory programs are determined to a considerable extent by the rules, and for the last several years, we have had official estimates of these forms of effect. In a report on their findings, the Office of Regulatory and Information Affairs in the Office of Management and Budget set the ranges of costs and benefits in dollars incurred as a result of compliance with regulation at between 254 billion and 1.8 trillion and 146 and 229 billion, respectively.^[6] Although the uncertainty in the analysis is considerable and the ranges are obviously quite broad, if one takes the lowest estimates in both categories the effects of rules, positive and negative, are still enormous. When the stakes are this high, the process that establishes the winners and losers is obviously important, and will be the focal point of much attention.

The contemporary rule-making process has its origins in the APA of 1946. Debates about the need for and design of a statute that would provide operating instructions to administrative agencies for the conduct of their most important functions began well before the

World War II. Over time the deliberations took numerous twists and turns, and generated considerable controversy.^[5] As enacted, the APA considers various forms of rulemaking, but by far the most important is outlined in Section 553, and has come to be known as “notice and comment” or informal rulemaking.

In this basic design, the responsible agency is expected to publish its intentions with regard to a new rule or regulation in the *Federal Register*, allow the public to comment on the proposed action in writing and publish its final decision, also in the *Federal Register*, with an effective date. The APA provided little guidance beyond these skeletal requirements, but subsequent Congresses, presidents, and courts have put plenty of meat on its bones. The space allotted here is insufficient to even scratch the surface of the mountain of detail in statutes, executive actions, and court opinions that may affect the conduct of a given rule-making. In all this subsequent legal authority, the fundamental operating principles of the APA have been preserved. Virtually every pronouncement by a Congress, a president, or judges relate in some way to one or more of the three fundamental rule-making principles established in the APA: information, participation, and accountability.

Information, Participation, and Accountability

The conception of information found in the original APA was quite limited. The act did require prospective notification of the public of the rule-making action the agency was contemplating in the *Federal Register*, but a general description could substitute for the actual language being proposed. Since that time, the clear expectation expressed repeatedly by those with political and judicial influence, and the requirements of the agencies themselves and the *Federal Register*, is that the full text of the proposed rule be published in the *Federal Register*, accompanied by a preamble that outlines authority under which the action is taken and the basis for the actions. Following a period of public comment, a crucial element in the second basic operating principle of the rule-making process, the agency again informs the public through publication of the final rule in the *Federal Register*, again with a preamble that provides not only essential background information, but also summarizes the results of the public input that the earlier notice generated. The vehicles that agencies themselves use to transmit this essential information to the public have also diversified and increased in number, fueled by the rise of interest groups, as well as the recent revolution in electronic communication and the web.

Developments since passage of the APA with regard to the second principle of rulemaking, participation,

have been even more dramatic. This is both understandable and appropriate because participation by the public in the writing of rules is, in effect, the surrogate for the democratic function served by the electoral process in legislation. It is what has led to Professor Davis's conclusion, expressed above, that rulemaking can be a more than an acceptable substitute for the legislative process. Like information, the level and type of participation established in the APA was limited, and there were no expressed guarantees that agency personnel would take what the public was expressing seriously. The act required agencies to solicit and receive oral or written comments at their discretion; it did not specify (or even mention) what the responsible officials were to do with the information once it was in hand. Since then, the forms of participation, and the expectations with regard to agencies' uses of what the public provides, have changed dramatically, and again the rise of interest groups has fueled many of these developments.

In addition to written comment, we know from survey work that interest groups closely monitor rule-making activity and engage in a variety of activities to influence agency decisions. Among the more prominent is the development of grassroots campaigns to increase the volume and intensity of public comments, coalition formation to develop allies in the campaign to influence the rule, and “informal” contacts with agency personnel prior to publication of proposed and final rules.^[5] It should also be noted that interest groups report relatively high levels of success in getting what they want from rule-making processes. This not surprising since agencies are now expected to take what they hear from the public seriously, to the extent that in the mandatory preambles of final rules the agency is expected to summarize the data and views received from the public, as well as explain how the original proposal has changed in light of the new information. The most extensive form of public participation, and one where the input of the public can be most influential, is negotiated rulemaking, which is discussed in the subsequent section.

The third fundamental principle that anchors the entire rule-making process is accountability, or the expectation that those who write rules may be questioned about and held responsible for their actions. The sole form of accountability mentioned in the APA is judicial review that grants the public the right to sue an agency if it violated procedures during the course of a rulemaking or issued a rule that constitutes an “arbitrary or capricious abuse of discretion.”^[7] Since passage of the act, the courts have virtually opened their doors to those seeking relief from rules and this form of accountability remains robust. The act does not explicitly mention the many mechanisms of accountability, some long-established, others more

recent, that are exercised by Congress and the president. The president has always appointed the leadership of agencies that write rules, submitted most of their budgets to Congress and supervised the agencies through executive orders and general management authority. The traditional powers of the purse, oversight, legislation, and approval of appointments have always been the purview of Congress. However, the president and the Congress have also fashioned instruments of accountability more finely tuned and focused on the rule-making process.

The Carter administration instituted a system of senior management involvement in the setting of rule-making priorities. Far more aggressive and significant was the system of White House review of all proposed and final rules instituted by President Reagan, which has been refined by each president since then. In addition, presidents have routinely issued statements of "regulatory principles" to guide those writing rules. Congress came late to this party. They enacted the Regulatory Flexibility Act as a new section in Title 5 of the U.S. Code where the APA is contained. It required agencies to analyze the effects of proposed rules on small businesses.^[8] Subsequently, they found the president's approach of microlevel review appealing. In 1996, they passed the Congressional Review Act that also amended Title 5 of the U.S. Code, requiring submission of rules to the General Accounting Office for a determination of its compliance with authorizing legislation with the option of enacting legislation to declare the rule null and void.^[5] The Congressional Review Act was a much delayed response to the Supreme Court's repudiation of the "legislative veto" that existed in various forms until it was invalidated in 1983.^[9] Although the powers of the Congressional Review Act are substantial, they have been used only once since its enactment in the early months of the Bush administration, when legislation passed and signed by the president undid the ergonomics rule issued by the Occupational Safety and Health Administration of the Department of Labor. Used or unused, the mechanisms of accountability currently available to the Congress, the president, and the courts are now so extensive that those writing rules must anticipate issues and concerns arising in these branches if the products of their rulemakings are to survive.

The rule-making process that has evolved from both the development and interaction of these three fundamental principles is now highly complex. Scholars have established that a contemporary rulemaking can pass through as many as a dozen stages of development, ranging from the authorization to begin a rule-making project, planning and securing the resources needed to complete the work, proceeding to the collection of essential information and policy guidance to develop a draft proposal and final rule, and the

multiple levels and types of review to which the product will be subjected.

RECENT RULE-MAKING REFORMS

In an effort to improve the process and content of rules, and to increase chances that it will succeed in accomplishing their varied purposes, agencies now have numerous tools at their disposal. Among the most prominent of these is the use of various outreach and consensus building techniques to improve both content and acceptability; the application of the principles of cost-benefit and risk analysis to determine the need for, and best approaches to, regulations under development and employment of information; and advanced communications technology to provide the public with essential information and to facilitate participation in the process.

Negotiated rulemaking is a form of consensual decision making that essentially transfers the concept of collective bargaining and the principles of consensual process to the development of rules. It has been used only sparingly, and there is significant controversy over its effectiveness and efficiency, but available research does show that those who participate in negotiated rulemaking rate the process and its results as superior to other techniques. Participants also report a deeper understanding for the complexities of the public policy process and a better appreciation for the positions of their adversaries.^[10]

The use of cost-benefit and risk analyses in rule-making spans several decades, but they have been given renewed emphasis in the current Bush administration. Risk analysis promotes an understanding of the dangers associated with various activities and situations and a strong priority-setting process for agencies. Cost-benefit analysis can build on risk analysis by providing estimates of the price to be paid under various potential interventions to solve perceived problems, as well as provide a value for the public and private good that the regulatory action will produce.

Electronic tools have revolutionized the creation, transmission, storage, retrieval, and use of rule-making information. Virtually all agencies are currently engaged at some level with the use of electronic resources to facilitate and make for efficient or effective participation and management of rulemaking.

A good case can be made that rulemaking is the most important source of lawmaking. Until the tragedy and outrage of September 11 was visited on the United States, the major public policy issues on our agenda were all domestic, and with no exception, they would ultimately be resolved through rulemaking. From stem cell research to arsenic in water to reform of social security, rulemaking will be the primary

source of guidance to the public on their rights and obligations under the law. These are but the latest in a virtually endless stream of issues and problems that have been delegated to agencies. Given the wellsprings of our politics and the inherent limitations and proclivities of our institutions, rulemaking will be ascendant until the policy process changes in some fundamental way.

REFERENCES

1. 5 U.S.C. 551 (5).
2. 5 U.S.C. 551 (4).
3. Davis, K.C. *Discretionary Justice: A Preliminary Inquiry*; University of Illinois Press: Urbana, IL, 1976; 65 pp.
4. Schoenbrod, D. *Power Without Responsibility: How Congress Abuses People Through Delegation*; Yale University Press: New Haven, 1993; 65 pp.
5. Kerwin, C. *How Agencies Write Law and Make Policy*, 2nd Ed.; Congressional Quarterly Press: Washington, DC, 1998; Chapter 5, 10, 67, 68, 220 pp.
6. Office of Information and Regulatory Affairs. *Report to Congress on the Costs and Benefits of Federal Regulations*; Office of Management and Budget: Washington, DC, 2000.
7. 5 U.S.C. 706 (2)(a).
8. 5 U.S.C. 601–612.
9. 462 U.S. 919.
10. Langbein, L.; Kerwin, C. Regulatory negotiations versus conventional rulemaking claims, counter claims and empirical evidence. *J. Public Adm. Res. Theory* **2000**, *10*, 599–600.

Rural Economies, Developing

Terry F. Buss

*Center for International Affairs, National Academy of Public Administration,
Washington, District of Columbia, U.S.A.*

Marcela Tribble

*School of Policy and Management, Florida International University,
Miami, Florida, U.S.A.*

INTRODUCTION

Rural economic development remains an elusive field to all but a handful of experts and specialists. Novices and generalists often harbor misperceptions about rural places and their redevelopment. A few misconceptions are as follows: 1) rural places are economically, socially, or politically homogeneous, or nearly so, so that they share common circumstances, problems, and prospects; 2) rural places have a few comparative advantages—such as inexpensive labor and extensive undeveloped land resources—but have major disadvantages or deficits—such as remoteness from markets, sparse populations, uneducated workforces, and shortages of capital—that narrow economic development options; 3) widespread consensus about how to help rural places exists, but carefully executed, targeted policies are needed; and 4) rural economic development policy and practice placed in ways that benefit people most.

at economic growth from 1980 to 1993 in an effort to assess diversity in the rural heartland (see Table 1).

Federal Reserve researchers found that winners (high-growth) and losers (decline) were composed of very different county types. Losers likely had concentrations of farming and trade-dominated counties, whereas winners were more likely to be government, retirement, and mixed-use counties. In part, this reflects the remnants of the farm crisis in the late 1980s and early 1990s. Adjacent, larger population, rural counties tended to grow at greater rates than non-adjacent, smaller population rural counties. Adjacent counties likely shared economic growth factors with urban areas without becoming, as yet, urban areas.

Even though it may make sense to talk about differences in prosperity between county types, there are also considerable differences within county categories. From a public policy perspective, this makes targeting federal/state economic development aid or programs problematic. Each county should be examined on its own merits.

HETEROGENEITY OF RURAL PLACES

The U.S. Department of Agriculture's (USDA's) Economic Research Service defines rural areas as places, incorporated or unincorporated, with fewer than 2500 residents and open territory. About one-fourth of the U.S. population lives in rural areas. About one-fourth of rural counties are dominated by farming (see Table 1). One-third are trade intensive, and only 3% represent mining. Manufacturing, government, retirement, and mixed-use counties each account for about 6% of concentrations.

COMPARATIVE ADVANTAGES AND DISADVANTAGES

The Federal Reserve Bank of Kansas City combined economic use, population, and proximity indicators for 779 rural counties in 12 states, then they looked

ABSENCE OF A CONSENSUS ON POLICY

National Rural Policy

Farm interests are effective in articulating and realizing their agendas nationally and regionally. A national farm policy has guided agriculture throughout the century. Farmers want and receive credit usually through the Farm Credit Bureau and crop price subsidies through the USDA. Only under the Clinton administration in the 1990s have farmers been less successful in winning policy battles with environmental interests, concerned about wetlands, fertilizer runoffs, and land preservation.

With the exception of a coherent farm policy, no consistent voice represents rural America's interests in the national rural development policy arena. Beth Honadle, policy maker at USDA for nearly 20 years, offers an example. In the late 1980s, four powerful associations—National Association of Towns and

Table 1 Economic winners and losers in rural counties, 1980–1993, 10th Federal Reserve District

County type	Winners		Losers	
	No.	%	No.	%
Farming	23	15.5	73	30.9
Manufacturing	10	6.8	17	7.2
Mining	1	0.7	12	5.1
Government	26	17.6	16	6.8
Retirement	19	12.8	9	3.8
Trade	35	23.6	97	41.1
Other	16	10.8	8	3.4
Mixed	18	12.2	4	1.7
Total	148	100.0	236	100.0

Source: Mark Drabenstott and Tim R. Smith. (1996) "The Changing Economy of the Rural Heartland," in *Economic Forces Shaping the Rural Heartland*. Kansas City: Federal Reserve Bank.

Townships, National Association of Counties, National Association of Regional Councils, and National Association of Development Organizations—tried to come together to lobby Congress and the administration for a national rural economic development policy, but nothing came of the effort because the associations could not agree on much.^[1]

Successive congresses and administrations have been divided on rural policy.^[2] Although it might appear that the USDA and its respective Congressional oversight committees determine rural economic development directions, they do not. A myriad of alphabet executive branch agencies—Small Business Administration, Rural Electrification Administration, Farmers Home Administration, Extension Service, Soil Conservation Service, U.S. Forest Service, Economic Development Administration, Department of Interior, and Environmental Protection Agency, not to mention the Appalachian Regional Commission, Tennessee Valley Authority, and Corps of Engineers—control bits and pieces of a larger pie. Each competes with overlapping jurisdictions, contradictory programs, and diverging visions for rural America.^[3–5] The U.S. Agency for International Development, for example, bolsters competitiveness of foreign firms, while the U.S. Commerce Department try to make American firms more competitive in world markets.

Locally, conflicts over policy objectives can be intense. At the heart of the issue are two, often conflicting, factors: preserving whatever makes rural life good and different from other places, while trying to address needs of poor people, flagging economies, declining populations and tax bases, and plant closings or military base shutdowns, on the one hand, or taking advantage of economic opportunity, on the other hand.

Ascendancy of one vision over another is not without consequence. Elderly retired people on fixed incomes in small towns suddenly find that rapid economic development so raises their property values that they cannot afford to live in the community, but are unable to leave. Some development of virgin land destroys it forever. Conversely, people in America's poorest regions—Indian reservations and the Mississippi Delta, for instance—may never benefit unless economic development is channeled their way, or once undeveloped rangelands become sites for gambling casinos developed by out-of-towners or Native American tribes. All conflicting visions and values.^[6]

EVOLVING ECONOMIC DEVELOPMENT POLICY AND PRACTICE

Federal, state, and local economic development policy and practice have evolved over the past three decades to address three epochs in an evolving rural America. The 1970s witnessed the "Rural Renaissance," a period of economic development and prosperity. The 1980s ushered in the farm crisis and decline of many rural places through population loss, declining income, increased poverty, and distress. The 1990s represented an era of differential prosperity and decline, depending on the region and local circumstances. Economic development strategies evolved from a focus on farming to industrial recruitment to entrepreneurship, although now most communities pursue all three simultaneously much as a stock investor diversifies a portfolio. Unhappily, policy and practice have yielded sparse results in many cases.

Industrial Recruitment

Rural industrial recruitment strategies focus on attracting branch plants of larger corporations, light manufacturing, distribution centers, and back office functions (e.g., low skilled data entry for banks), where officials believe they have a comparative advantage. Some rural communities chase high-technology companies and service businesses. The eight-county Mount Rogers Planning Commission in Virginia, for example, created a revolving loan fund (see section below) to attract high-tech companies—"brains over brawn" strategy. Officials in the Appalachian Regional Commission boast that every rural and distressed community is covered by a community college. Having access to colleges allows rural policy makers to replicate high-technology programs common at large, urban universities.

Public and private initiatives to attract outside industries are now seen by many to be wasted efforts.

There are just not enough industries willing to relocate or expand operations in rural places to make recruitment an effective strategy.

Entrepreneurship and Business Climate

Rather than devote so much effort to industrial recruitment, some rural places are trying to grow their own economies from within through entrepreneurship programs, small business incubators, entrepreneurship training, business technical assistance, export promotion, import substitution, business financing, and community marketing, especially in communities that believe they must help themselves or no one else will. Studies generally show that entrepreneurs participating in the programs value them, but many even in the smallest communities are unaware of their existence and do not seem to need them (e.g., Ref.^[7]). In earlier days, in the rush to foster entrepreneurship, organizations with little experience in the field began to work with entrepreneurs. Agricultural extension agents, community colleges, and even church groups (this is referred to as faith-based development) began to carve out a role. Although the assistance providers have, over the years, become more sophisticated in their approaches, in the late 1970s and early 1980s, many assistance programs were that in name only, handing out advice that was correct only by accident. Organizations have become more sophisticated. Now organizations—public, non-profit, and private—offering assistance in some rural places find themselves competing with one another, likely improving service quality.

Other initiatives focus mostly on improving business climate. This strategy creates a supportive environment for all entrepreneurs. It includes *deregulation*—eliminating needless or counterproductive regulations, *tax reduction*—lowering taxes businesses find oppressive, and *public service provision*—ensuring that education, health, safety, infrastructure, and utilities are effective, efficient, and of high quality.

Subsidizing Capital

Capital, in the form of debt or equity, drives rural economic development, especially entrepreneurship and local growth generally. However, analysts disagree about whether sufficient capital exists to meet demand.^[8] Those who hold that capital gaps or credit crunches widely exist in rural America believe that capital is scarce, costly, or problematic. Banks are conservative and lack competition. Equity investment is waiting.

The preponderance of empirical evidence suggests otherwise: *credit* markets effectively serve the needs of nearly all businesses in rural areas, or at least as well

as those in urban areas.^[9] Consider results of one representative study of entrepreneurs seeking bank financing across Idaho, Iowa, and Montana.^[10] Nearly three-fifths of entrepreneurs had bank loans. One-in-10 were turned down for loans, but only a handful of entrepreneurs were unable to eventually obtain bank financing.

Some analysts believe that as larger banks acquire smaller ones, credit availability in rural places will decrease. Results of studies on the impact of this trend are mixed. Some find that banks will take capital out of rural areas, reallocating it to more productive investments in cities. However, others see larger banks looking for growth and competing for more rural business, not less. Larger banks, having much more capital than small independents, actually make more credit available, not less. Banks now use computer credit scoring to evaluate loans, in effect making loans based on merit rather than “good ‘ole boy” networks that characterize rural lending by small banks.

Likewise, the few existing studies suggest that equity investment is widely available in most rural places. Brophy,^[11] in a study of equity investment in the form of venture capital and initial public stock offerings, found that rural areas do have access to venture capitalists and stock markets, even though these financial institutions may not be physically located near them. Entrepreneurs access equity capital in larger cities, serving rural areas. Using the Internet sites such as the Small Business Administration’s ACE-net, entrepreneurs and investors can link up in ways not possible even a few years ago.

Even though there appears to be ample capital, state and federal credit and equity capital programs abound.^[12] State and federal programs offer access to capital where it is unavailable and low interest or favorable terms where it is available—USDA’s Fund for Rural America, Business and Industry Program, Intermediary Re-lending Program, Rural Business Enterprise Grants, Rural Technology and Cooperative Development Grants, and Rural Economic Development Grants and Loans; SBA’s section 7a and 504 Programs; EDA’s Economic Adjustment Assistance Program; and Appalachian Regional Commission’s programs similar to USDA’s (see www.ag.ers.gov or www.arc.gov). In vogue are revolving loan funds in which a lender receives public monies to lend to business. When entrepreneurs repay loans, funds are re-lent. Lenders extract a fee for fund administration.

Another trend in public finance of private business are microcredit programs. Microbusinesses employ one or two people and generally have little growth potential, but they do employ those not employable elsewhere. Public programs provide small loans at favorable terms. Most operate as revolving loan funds. Many programs are subsidized, being unable to turn a profit. Because they focus on low growth, often

subsistence-based microbusinesses, programs do not much affect local economies. For many public officials, they may be more symbolic than economic.

Networking

Some rural economies may lack growth and development opportunities because they cannot achieve critical mass to bring together financial, technical support and educational institutions, and, most important, clusters or concentrations of similar industries that share these institutions. Although numerous networking and clustering initiatives are underway, assessments suggest that they have not “completely overcome the weaknesses of inadequately trained or financed entrepreneurs, the disadvantages of distance from market centers, or the limitations of small-scale entrepreneurs” (Sommers, 1998: 66).

Home-Based Businesses

Economic development strategists consider promotion of home-based businesses as a viable revitalization initiative. Policy makers provide incentives for people in rural areas to work out of their homes, including rezoning; influencing credit unions, banks, and other lenders; and creating small business “incubators without walls.” The “discovery” of home-based businesses seems rather strange. In fact, rural economies rely and have always relied on home-based businesses to a large extent: “old wine in new bottles.”

Telecommunications

The telecommunications revolution, stimulated by the Telecommunications Act of 1996 and revised regulations from the Federal Communications Commission in 1997, may eventually lead to development opportunities for rural places. Firms can cluster or network (see above) without having to be in the same geographic location. Buying and selling goods and services is greatly enhanced by the Internet. Having access to telecommunications of all kinds leads to innovation and new product development. Rural populations may become part of a workforce electronically, staying at home rather than leaving the community for work. Yet, many rural places will be slow to seize opportunities. Lack of competition arising out of small markets inhibits the development of first-rate telecommunications infrastructure and services, causing some commentators to call for publicly owned telecom enterprise. Given the track record of public enterprises (e.g., TVA teetering on insolvency), this may not be a viable solution.

Enterprise Zones/Enterprise Communities and Community Development Financial Institutions

Economic strategies discussed thus far intersect in the new \$1 billion Enterprise Zone/Enterprise Community (EZ/ECs) program and Community Development Financial Institutions. The Clinton administration designated three rural EZs and 30 rural ECs in 1994 from among 227 applicants (Reeder, 1997). The Bush administration has continued the program. Designation as an EZ or EC makes communities eligible for all types of federal aid, in addition to whatever states may channel to designees. Unlike EZs under past administrations, new ones combine economic development with social services. The three EZs are eligible for \$40 million in Social Service Block Grants each, and ECs for nearly \$3 million each. The USDA also targeted an additional \$50 million annually to the EZ/EC areas for infrastructure and business assistance. Each area is eligible for federal tax breaks, relief from government regulation, and other incentives.

In 1997, the Clinton administration added a program comparable to the EZ/EC initiative, the Rural Economic Area Partnership (REAP) to assist rural places in the Northern Great Plains. Recently, North Dakota received REAP designations for two zones. Funding under REAP is used to develop plans, then implement them. North Dakota created a revolving loan fund. Earlier in 1996, the Clinton administration created the Community Development Financial Institutions (CDFI) program for distressed communities, funded at about \$50 million annually. Selected financial institutions—banks, credit unions, loan or venture capital funds—interested in community development receive subsidies and incentives to invest in distressed areas: first-time home buyer mortgages, rental housing rehab, startup business loans, and bank services for low-income people. Ten CDFIs serve rural places.

These programs in turn parallel the 1994 Northwest Economic Adjustment Initiative, a \$1.2 billion, 5-year finance and technical assistance program for businesses, workers, tribes, and communities hurt by reduced federal timber harvesting in California, Oregon, and Washington. This program also promotes development partnerships.

Although focusing on distressed communities is laudable, evaluations of similar programs show only isolated success. Zones performed no better, and in some cases worse, than comparable areas without zones. Business is not attracted from outside distressed communities, leading to new jobs and employment. Rather, local businesses simply move into zones to take advantage of incentives. Zone businesses tend to hire workers, then take advantage of incentives; in other words, incentives compensate business for what

they would have done anyway. Combining economic and social policy has not worked in many places. Small businesses are jeopardized when forced to hire some zone residents who do not want to work, have no work experience, are uneducated, and may have a host of social problems. In addition to poor performance, rural EZ/ECs have been slow in getting underway because of bureaucratic red tape, unclear federal regulations, and contradictory policy directions.^[13]

PARTNERSHIPS AND COLLABORATION

In a previous section, fragmentation of rural economic development interests was presented. It surprises no one, therefore, that a major thrust of rural economic development policy would be for government—at all levels—to try to undo negative effects of fragmentation by creating organizations to overcome it.^[12] The Reagan administration created Rural Enterprise Teams as a way of partnering the USDA and other alphabet agencies—about 60 in all—with state-level economic development agencies to pursue pilot projects. The Office of Rural Development Policy coordinated rural economic development, but was eventually defunded for ineffectiveness. Under the Bush administration in 1990, State Rural Development Councils, having an executive director and staff, work as liaisons between governor's offices and federal agencies, as part of the National Rural Development Partnership (Radin, 1997). As of 1997, 39 states created councils. Councils working with the partnership formulate “action plans,” not policy, to be pursued in a consensus-building process. Bush also appointed a Council on Rural America, whose report, *Revitalizing Rural America Through Collaboration*, recommended creating an advisory council, a senior advisor, and a policy coordinating groups to foster collaboration. The White House staff ignored the report, fearing creation of yet another organization.^[1]

The Clinton administration also supported the National Rural Development Partnership and state councils, and even added another variation in 1993—President's Council on Sustainable Development. Although the council is not exclusively rural, it has a rural taskforce, the Pacific Northwest Council, imbedded in its structure. In its report, *Sustainable America*, the council calls for economic development activities that protect the environment, such as ecoindustrial parks. The council employs a multistakeholder process, but has little program or money. The council wants to create a National Sustainable Development Extension Network to educate, train, publicize, and transfer technology for sustainable development.

The track record of these approaches is mixed, at best (Radin, 1997). Many initiatives added a new layer

of bureaucracy to an already crowded field. Initiatives duplicated existing organizations in some states, furthering, rather than reducing, competition and fragmentation. Initiatives focused so much on process that they sometimes failed to do anything substantive, either because the parties could not agree or the process went on so long that participants lost interest. Initiatives have few resources, minimizing impact. The most salient criticism is that these approaches simply assume a problem and then provide funding to entice states and regions to participate. They do not allow states and regions to define problems and devise their own solutions, then provide funding in support. Consequently, states and regions are independently developing their own innovative approaches, such as Oklahoma Futures; Kansas, Inc.; or Greater Minnesota Corporation of the bilateral Red River Trade Corridor, serving Minnesota, North Dakota, South Dakota, and Canada's Manitoba Province, to solve coordination and resource allocation problems.

PEOPLE, PLACES, AND ECONOMIC DEVELOPMENT

At the heart of rural economic development is whether to invest scarce resources in places or people. Nearly all rural economic development policy is place based. Its intent appears to be to keep people in rural locations, hoping that they can grow their economies, rather than abandon them. The National Commission on Agriculture and Rural Development Policy made this explicit. Two goals of rural policy are: “Rural areas and people must be economically self-reliant,” and “Rural areas and people must be able to adapt to change” (1990: 6). Building industrial parks, recruiting industry, supporting entrepreneurship, developing infrastructure, and improving housing and the like are place-based strategies because they help only if people reside in rural places where investment are made. Even training or retraining programs are place based in the sense that they train people for possible jobs in the area.

Some place-based strategies may work well to be sure, but others clearly have not. There are still large pockets of rural poverty—poverty counties—in the Mississippi Delta, throughout many states in Appalachia, and on Indian reservations, in spite of the billions spent over the decades to make them better places to live. One reason why pockets of poverty remain is that even after continuing heavy public investment, these rural places still lack sustainable competitive advantages over other places. So public subsidies hold many people in place without making their lives better.

There are some exceptions worth noting. Lower transportation costs originating from containerization,

deregulation, and computer linkages; telecommunications allowing instant data transfer; and multisourcing of materials—many rather than one source of supply—have reduced the “rurality” and remoteness problem in rural economies.

REFERENCES

1. Honadle, B.W. Rural development policy. *Econ. Dev. Q.* **1993**, 7 (3), 227–236, August.
2. Committee on Small Business. *Hearings on the Federal Role in Rural Economic Development*; 1993. 103rd Congress, 1st Session, July 21, US Senate.
3. General Accounting Office. *Rural Development: Patchwork of Federal Programs Needs to be Reappraised*; GAO: Washington, DC, 1994. RCED-94-165, July.
4. General Accounting Office. *Rural Development: Patchwork of Federal Water and Sewer Programs is Difficult to Use*; GAO: Washington, DC, 1995. RCED-95-160BR, April.
5. General Accounting Office. *Economic Development: Limited Information Exists on the Impact of Assistance Provided by Three Agencies*; GAO: Washington, DC, 1996. RCED-96-103, April.
6. Rowley, T.D. The value of rural America. *Rural Dev. Perspect.* **1997**, 12 (1), 2–4.
7. Greenberg, E.; Reeder, R.J. *Who Benefits from Business Assistance Programs?* Economic Research Service, US Department of Agriculture: Washington, DC, 1998. #736-04.
8. Rural Finance Task Force. *The Adequacy of Rural Financial Markets*; RUPRI, University of Missouri: Columbia, MO, 1997.
9. Buss, T.F. *Capital, Emerging High Growth Firms and Public Policy*; Praeger: New York, 2001.
10. Buss, T.F.; Yancer, L.C. Bank financing of entrepreneurs in rural America. *Pub. Budg. Account. Financ. Manag.* **1999**, 11 (3), 427–448, Summer.
11. Brophy, D.J. Developing rural equity capital markets. In *Financing Rural America*; Federal Reserve Bank of Kansas City: Kansas City, 1997.
12. Bonnett, T.W. *Strategies for Rural Competitiveness: Policy Options for State Government*; Council of Governors' Policy Advisors: Washington, DC, 1993.
13. General Accounting Office. *Economic Development Activities*; GAO: Washington, DC, 1997. RCED-97-193.

Rural Health Policy

Esther M. Forti

Department of Health Professions, Medical University of South Carolina,
Charleston, South Carolina, U.S.A.

INTRODUCTION

Health care in America has undergone tremendous changes since the 1980s and continues to change. Rural areas are still struggling with inequitable payment systems for services, instability of many small rural hospitals, continued shortage of health care professionals, and low population density to support many needed services and managed care systems. Thus, the health care needs of rural residents are still not being met at the same level as those in more urban areas. Given the scarcity of resources in rural areas, no rural community can be totally self-sufficient in meeting the health care needs of its population. Demographic, political, social, cultural, and economic changes are different for rural than for urban populations, yet health policy does not always recognize these differences.

Policy decisions frequently address urban needs and problems and are usually informed by urban research. The voices of rural policy makers have a tendency to be lost in national policy. When policies are implemented that do not consider diversity in rural, and rural and urban, differences, problems in rural areas remain unsolved with the potential of urban-based policies being detrimental to the rural health care system. For example, lower Medicare payments to rural hospitals led to increases in hospital closures in the 1980s. Moreover, the national policy focus on lowering costs of health care through competition has not been favorable to the ongoing needs of rural health. In light of these events, it is crucial that the policy community be cognizant of and employ available and relevant rural research findings to inform health policy debate processes on the national and local levels.^[1] The goals of this article are to provide an overview of the issues that affect the rural health care system, describe selected rural health policy priority issues, and propose policy recommendations to create a viable rural health care system.

BACKGROUND

In the 1980s, rural advocates talked about a health care “crisis” in rural America. Policies and programs administered by federal agencies did not consider differences in rural and urban areas. The federal government is developing mechanisms to address rural

health concerns. The National Advisory Committee on Rural Health, comprised of nationally recognized rural health experts, was chartered in 1987 to advise the secretary of the Department of Health and Human Services (DHHS) on strategies to address the health care in rural America. The Office of Rural Health Policy (ORHP), also established in late 1987, provides a voice for rural health policy advocacy, information, and grant programs. The office works within the federal government, and in the private sector with associations, providers, and community leaders to solve problems and seek assistance in meeting rural health needs. One focus of the ORHP was the initiation of several grant programs to address rural health deficits. Two of these programs, Outreach and Network Development grants, focus on increasing access to care, as well as coordinating and improving the quality of health care in rural communities. Grants to states include the State Offices of Rural Health (SORH) program and the Rural Hospital Flexibility program. The SORH program, initiated in 1991, provides matching funds to states to help rural communities build viable health care systems. The Rural Hospital Flexibility program was initiated in 1999 to provide funding to state governments to strengthen rural health by allowing small rural hospitals the flexibility to reconfigure their operations and be licensed as critical care hospitals.^[2]

Another program under the direction of the Health Resources and Services Administration is the 1996 Health Centers Consolidated Act. This program provides funds for community health centers (CHCs), migrant health centers, health care for the homeless, and primary care for public housing residents. The CHCs, known as one of the safety net provider systems, assume responsibility for the primary and preventive health care needs of the increasing number of uninsured and vulnerable populations that reside in rural and medically underserved areas.^[3]

SELECTED POLICY ISSUES AND RECOMMENDATIONS

Policy issues surrounding rural health care needs are linked closely to national health care concerns, especially with regard to resources such as access to care, caring for the uninsured, health professions

workforce, defining rural for public programs, and reimbursement structures.

Access to Care

Rural health policy has historically focused, and continues to focus, on problems with access to care. Factors associated with the rural environment that affect access to health care are scarce resources, diversity among rural communities, fragile economies, high poverty rates with high unemployment and underemployment, lack of adequate transportation systems, higher percentage of uninsured or underinsured, fragile rural hospitals, an increase in the number of elderly people (e.g., high dependency ratio), and shifts in population with outmigration. Population shifts are important to understand because they result in a low unstable population base that contributes to an inability of rural communities to sustain adequate health services and infrastructures. Small rural hospitals are more likely to survive by becoming diverse in providing essential services to their communities. Outpatient services, Medicare swing-bed programs, long-term care facilities, home health care services, hospice care, urgent care centers, satellite clinics, multi-hospital arrangements, managed care systems, and critical access designations are several alternate delivery options available to small community hospitals.^[4]

Policy recommendations are to continue to work to ensure that rural people have access to quality health care by strengthening the rural infrastructure through public programs for the safety net provider systems, creating rural health networks, workforce development, and equitable reimbursement systems. Also, it is important to continue to provide funds and technical assistance to small rural hospitals to enable them to diversify services and join health care networks, and to devise innovative community-driven strategies to strengthen rural transportation systems.

Rural Uninsured

There has been a substantial rise in the number of uninsured over the past several years. Rural persons are more likely to be uninsured than those living in more urban areas. Moreover, there are notable differences between the rural and urban uninsured. Rural persons are more likely to rely on private insurance plans because they work for small employers, are self-employed (e.g., farmers), or are underemployed. In addition, rural workers are more likely to have incomes less than 100% below the poverty level than urban workers. Policy intervention will most likely require public investment.^[5]

Policy recommendations are to remember that policy formulations need to be sensitive to the low incomes of rural residents, and strategies need to be devised to aggregate large groups of similarly insured persons because large employers are not common in rural settings.

Shortage of Health Professionals

Despite the fact that rural residents report poorer health and have higher rates of chronic diseases, there exists a shortage of health professionals to serve these health needs. Initiatives developed to address health manpower shortages in rural underserved areas include the National Health Service Corps (NHSC) and federally designated Health Professional Shortage Areas (HPSAs). The NHSC was established in 1972 in response to the shortage of health manpower and inadequate access to health services in rural and urban poverty areas. Legislation was passed to enable students to obtain scholarships for training costs in return for field service or loan repayment to already trained primary care clinicians and matching these candidates to communities in need of their services.^[3]

Also, the federal government designated HPSAs and Medically Underserved Areas (MUAs) to address the shortage of health professionals in rural areas. The DHHS recommends a ratio of one primary care physician to every 2000 residents and designates medically underserved as those falling outside this recommendation. In many rural areas, this ratio translates into one physician for every 3500 residents. These designated shortage areas receive federally subsidized monies in an attempt to correct such shortages.^[6] However, it is difficult to retain physicians in rural practices because of professional isolation, lower salaries and reimbursement structures, poor school systems, inability for spousal employment, and longer work hours.^[7]

Policy recommendations are to continue and increase funding for the NHSC program; to train and educate health professionals in rural community-based settings to increase the potential of health professionals choosing and practicing in these communities; and to use non-physician providers (physician assistants, nurse practitioner, nurse midwives) to address the shortage of primary care physicians. State Practice Acts need to be examined and revised to allow reimbursement to non-physician providers for providing primary care health services in rural communities.

Defining Rural for Public Programs

One of the policy issues of major interest is the continued lack of a uniform definition for what is rural. The need for precision in defining rural has escalated

because today's rural America is much more diverse than yesterday's and the needs continue to grow. The topologies most commonly used are rural-urban or metropolitan-non-metropolitan dichotomous groupings. Data are often collected at the county level that bypass small, remote rural communities. Further, rural and non-metropolitan are not synonymous. For example, non-metropolitan areas include urban populations (small cities), and Metropolitan Statistical Areas (MSAs) contain portions of rural populations. Such crude dichotomous coding of residence makes identification of diversity across rural difficult and complicates the implementation of federal policy designated to meet the health care needs of rural residents. Adding further to the confusion, values defining areas as rural are diverse and varied from one government agency to another.^[8]

One policy recommendation is to continue to identify various typologies to address diversity in rural settings and to assist policy makers in choosing a definition that will match need with program.

Rural Reimbursement Structures

Rural physicians rely more heavily than urban physicians on rates of reimbursement from public entitlement programs, such as Medicaid and Medicare. From the inception of these entitlement programs, the system of payment was lower for rural areas based on the assumption that rural areas had lower costs and a lower wage index. These lower payments have had a negative impact on the survival of rural hospitals and retention of physicians. In addition, the dependency ratio is high in rural areas because of the increasing number of elderly people. Medicare reform is a current topic of debate, and much activity is being put forth to redesign the system. Medicare proposals are spread across a continuum from a government-financed system at one end to a system whereby the government supports private health plans. Medicare payment policies are designed to reward market-based efficiency and cost containment. These policy proposals do not consider that rural areas have low-volume providers with fixed costs, decreased access to health care professionals, and persons with poorer health conditions. These rural characteristics place beneficiaries at risk in a market-driven Medicare redesign system.^[9]

Medicaid is a government-sponsored health insurance program for eligible persons. States determine the eligibility requirements and receive federal dollars based on indicators of need. Many states use Medicaid managed care programs. Medicaid managed care reimbursement models offered to providers range from a case management fee and a fee-for-service basis to a

reimbursement system that provides a fixed monthly per capita rate. The type of program offered in rural areas depends on its proximity to urban centers. Urban centers were much more likely to receive fully capitated programs. However, there has been a reluctance on the part of managed care organizations to provide services to rural areas because of low population density and a lack of appropriately credentialed health care providers such as non-physician practitioners. However, some states have offered flexibility in program eligibility. For example, a wide range of providers were allowed to function as primary care case managers. They also allowed emergency rooms to provide care after hours in rural communities with limited practitioners to share on call.^[10]

Policy recommendations are to redesign the Medicare reimbursement structure to ensure that Medicare beneficiaries have access to needed care in rural settings where payments need to approximate the costs that providers would incur in furnishing high quality care; to redesign Medicare to ensure that quality of care is monitored through standardized measures that are used across managed care and fee-for-service systems; and to have states continue to show flexibility in program requirements.

CONCLUSION

This article provided an overview of the relevant issues related to rural health care and describes trends and recommendations to address several of the problems plaguing rural America. It is essential that federal and state public programs consider rural in their policy decisions. The political milieu, demographics of rural settings, population density as related to services, and diversity among rural communities must be considered in health policy decisions to adequately address the health care needs of rural residents.

The future of rural health relies on a system of networks staffed by health professionals who have been trained in an interdisciplinary manner to meet the unique needs of a diverse rural America. Survival mandates that rural communities accept ownership, develop innovative alternate delivery systems, telecommunication systems, and integrated rural health networks to enhance their health care systems and to provide comprehensive health services to its populations. Solving the problems in our rural health care system is a challenge of balancing broad public policy with focused local needs. As Tilly, a rural sociologist, contended, when threatened communities do act because rural people have a long-standing tradition of coming together to solve their problems.^[11] Finally, a report on rural health defined one vision statement as, "Rural Communities share in the responsibility

for a healthy nation and the nation shares in the responsibility for a healthy rural America.’’^[12]

REFERENCES

1. In *Linking Rural Health Services Research with Health Policy Conference Report*, Proceedings of National Conference, Washington DC, June 15–16; George Mason University: Washington DC, 2000; RUPRI 24.
2. <http://www.ruralhealth.hrsa.gov/policy.htm> (accessed September 2001).
3. *Selected Federal Programs Meeting the Health Care Needs of Rural Americans*; <http://rhr.gmu.edu> (accessed September 2001).
4. Ricketts, T. Federal programs and rural health. In *Rural Health in the United States*; Ricketts, T.C., Ed.; Oxford University Press: New York, 1999; 61–69.
5. Pol, P. Health insurance in rural America. *Rural Policy Brief* **2001**, 5 (11), 1–3.
6. <http://bphc.hrsa.gov/dsd> (accessed September 2001).
7. Forti, E.M.; Martin, K.E.; Jones, R.L.; Herman, J.M. Factors influencing retention of rural Pennsylvania family physicians. *J. Am. Board Fam. Pract.* **1995**, 8, 469–474.
8. Ricketts, T.; Johnson-Webb, K.; Taylor, P. *Definitions of Rural: A Handbook for Health Policy Makers and Researchers*; Federal Office of Rural Health Policy: Washington, DC, 1998; 13 pp. HRSA 93-857(P).
9. Andrew, F.C.; Charles, W.F.; MacKinney, A.C.; Timothy, D.M.; Keith, J.M.; Rebecca, T.S.; Mary, K.W. Redesigning medicare: considerations for rural beneficiaries and health systems. *Rural Policy Brief* **2001**, 6 (6), 1–3.
10. Slifkin, R.; Casey, M. Medicaid managed care in rural areas. In *Rural Health in the United States*; Ricketts, T.C., Ed.; Oxford University Press: New York, 1999; 95–100.
11. Tilly, C. Do communities act? *Sociol. Inq.* **1973**, 43, 209–240.
12. The Federal Office of Rural Health Policy and the National Rural Health Association: Washington DC, 1998; 13 pp.

Science of Administration

Mark R. Rutgers

Department of Public Administration, Leiden University,
Leiden, The Netherlands

INTRODUCTION

The science of administration can be characterized as the systematic or methodologically guided attempt to describe, understand, and/or explain the workings of administration. Administration concerns the mobilization and organization of men and materials to develop and execute specific social goals (i.e., policies). This description can be further specified by indicating that the science of administration is concerned with *public* administration. It is an applied field of study that brings together or integrates knowledge on public administration developed in a wide array of other fields (i.e., it is an interdisciplinary science). There is, however, no consensus on the meaning of “science of administration” because the meaning of “science” and “administration” is a topic of dispute.

DEFINITION OF ADMINISTRATION

Science of administration is used in three different ways: historical, ontological, and methodological. The three different references to a science of administration are intertwined, but can be at odds.

When used to indicate an *historical* episode, science of administration refers to the approach to studying public administration as practiced by American authors in between roughly 1910 and 1940. The idea was that a science of administration would discover universal principles of administration. This period can also be referred to as pre-World War II public administration, or administrative orthodoxy. In this meaning, the science of administration is opposed to post-World War II developments in the study of public administration.

In an *ontological* sense, science of administration is used to indicate that the object of study—administrative reality—is a uniform, monomorphous phenomenon. The science of administration is then put forward as the general study of administration. It opposes the convention that the science of administration has as its object of study (just) public administration. From a scientific perspective, it is considered irrelevant or secondary to distinguish between business (profit) and public (non- and not-for-profit) administration.

Debates on limiting or broadening the object of study of the science of administration have resulted in, for example, arguments to focus on public policy or public management, or to expand the scope to human decision making or governance.

As a *methodological* statement, the reference to a science of administration is used to claim a specific approach (i.e., method) for studying administrative phenomena. Usually it implies that a distinction can be made between a scientific approach and the art, study, practice, and/or profession of administration. Again and again, authors have claimed to provide the proper (i.e., scientific) approach to studying (public) administration contrary to their forerunners. In the second half of the twentieth century, the reference to a science of administration is generally associated with the ideas of Herbert Simon, who argued a logical-positivistic approach. The question as to whether the study of (public) administration is, should be, or can be regarded as a science ultimately depends on the ideas on what constitutes a (social) science. The views on science and on human knowledge in general also determine the meaning (and acceptability) of characterizations of the study in terms of paradigms, discipline, practical relevance, interdisciplinarity, and so on. Proponents and opponents of a scientific approach to administration can be grouped in accordance with the more general divide in the philosophy of the social sciences between nomothetic (or positivistic) and hermeneutic (or interpretative) methodologies.

Cultural Aspects

The debates on the scientific status of the study of public administration are influenced by more linguistic or cultural aspects. An example is the double use of public administration in English for denoting the study (science) and the object of study (social practice). This makes it difficult to keep the two apart in English, and authors are inclined to mix up the two. A common attempt to avoid this is to use capitals when referring to the study of public administration. Another example is the meaning of science. For instance, history, social studies, and the like are regarded as sciences in German because the meaning of *Wissenschaft* is not limited to the approach in the natural sciences.

Cultural differences are even more prominent in the conceptualization of the object of study. Public administration is the product of sociohistorical developments. Notably, the constitutional and legal context is of prime importance as to the specific nature and meaning of public administration in a social reality. For instance the nature of the relation between politics and administration can be very different in many respects. The precise meaning of public administration is very much culture and language specific. Thus the nature of “the separation of powers” may be interpreted differently, and there may be no equivalent concept available for policy, agency, or even administration in another language. Whether this is important for a science of administration again depends on the philosophy of the social science adhered to.

HISTORICAL DEVELOPMENTS

Cameralism to Positivism

The attempt to develop a specific science of administration dates back to the seventeenth century's development of council studies: *Cameralism* and *polity science*.^[1] The appointment of two university professors in Prussia in 1729 signals its formal beginning. Veit Ludwig von Seckendorff can, however, be considered the first administrative scientist due to his publication of *Teutscher Fürsten Staat* (1656), or *German Principality*.^[2] He described topics relevant to the administration of a small principality. Seckendorff called his work scientific because it is an attempt to transcend the uniqueness of everyday practice. The French author Delamare took a similar position in *Traité de la Police* (1703–1738), or *Treatise on Polity*.^[3] He crafted a science about public administration under a prince. Both books were primarily collections of historical, practical, and legal knowledge on the running of a state's administration. It was scientific according to contemporary standards.

Eighteenth-century scholars such as Justi and Sonnenfels were much more theoretically oriented. They took a contract theory of the state as their starting point and argued that the king or prince was supposed to provide happiness for its inhabitants by means of the state's administration or polity: “The polity is a science dealing with organization of the home administration of the state.”^[4]

With the rise of a more modern concept of science in the Enlightenment, we encounter the first attempts to establish a positivistic science of administration. The French author Bonnin's *Principe d'Administration Publique* (1812), or *Principle of Public Administration*, is one of the earliest.^[5] According to Bonnin,

public administration was an application-oriented science, balancing law and social studies. He wanted to model it on the positivistic natural sciences.

In the second half of the nineteenth century, the most prominent attempt to create an encompassing administrative science is by the German scholar Lorenz Von Stein (1815–1890). He regarded *Verwaltungslehre*, or the *Study of Administration*, the crown on all the sciences because it brings together all knowledge available to improving society by means of the state. Stein was one of the first to characterize the modern state as an administrative state.^[6] Despite Stein, a specific administrative science disappeared almost entirely from Europe. This was partly due to changing ideas about science, resulting in the removal of applied fields from the universities, as well as to the rise of liberal ideas on state and society, causing the old polity or police sciences to become reviled. The meaning of administration was narrowed to the execution of law and regarded a topic for lawyers, not for social scientists.

American Developments

In the United States, a science of administration developed in the early twentieth century in the slipstream of nineteenth-century ideas on the improvement of public ethics (the moral reform movement). Wilson's *The Study of Administration* (1887) is generally considered to be its first herald.^[7] Historically, Goodnow's *Politics and Administration* (1900), and the establishment of the New York Bureau of Municipal Research in 1906, can be considered the actual starting points of the science of administration in the United States.^[8] The ideal was a politically neutral, professional, morally impeccable, and performance-oriented state bureaucracy. Early American administrative thought was characterized by a strict normative distinction between political and administrative activities. This made it possible to regard public administration as the business of government and to treat public and business administration alike. There was a strong influence of Frederick Taylor's scientific management in the development of the science of administration. Also influential is Henry Fayol, who developed a general theory of administration by formulating principles of administration. Contrary to the natural sciences, there can be no laws of administration, he argued, as there will always be an amount of uncertainty involved.^[9] The search for universal administrative principles flourished in the 1920s and 1930s. There was the strong conviction that there are general principles of administration, and that their validity is universal and can be empirically validated.

Postwar Developments

In the post-World War II period, the American discourse became more divided. There was much criticism on the advancements, and the search for principles was broadly criticized. In particular, Herbert Simon and Dwight Waldo are associated with the search for new directions. In his *Administrative Behavior* (1947), Simon argued a logical-positivistic empirical science of decision making. The science of administration should focus on facts and refrain from value judgments.^[10] Almost to the opposite, in his *The Administrative State* (1948), Waldo stressed the specific political (i.e., valuation) nature of public administration and called for an interpretative and philosophical approach.^[11] He rejected the reduction of science to the methods of the natural sciences. Waldo started to refer to a “crisis of identity” in the study, a characterization that became widely acknowledged and is still influential.

In the second half of the twentieth century, the development of an autonomous science of (public) administration was picked up again in Europe. It relied heavily on U.S. ideas, but maintained (to the present day) a strong legal orientation and generally identified public administration as the state administration in a *Rechtsstaat*.

In the 1960s, public administrationists became split roughly along two dimensions: academic versus practice and research versus educational ideals. The academic-research-oriented administrationists pursued a positivist approach in the spirit of Herbert Simon. The focus was on decision making and organization studies. The academic-education-oriented group was primarily concerned with the development of the study in terms of a generalist curriculum, combining social sciences, history, comparative studies, and political theory. Here we find followers of Dwight Waldo. Although advocates of both camps were present in Europe, the prime concern was still the establishment of a social science perspective on public administration.

In the 1970s and 1980s, the multitude of alternative approaches to administration made a homogeneous view of the science of administration almost impossible. Nevertheless, the empirical study of public organizations and policies dominated the research agenda. More or less independent schools emerged. Some are characterized by a specific focus on (parts of) social reality, for instance, policy making or public management; others took a more specific (i.e., limited, theoretical, or methodological) orientation, such as rational choice theory or communicative theory. In Europe, this differentiation of the field was less because of the smaller number of scholars involved. At the same time, there was (and still is) a relatively

strong differentiation in national traditions in studying (public) administration.

In the 1980s and 1990s, the mainstream continued to focus on a practically relevant empirical study of public policy making and public management. The proliferation of alternative approaches to a science of administration, however, persisted to the degree that a diverse “refounding movement” established itself.^[12] Debates on the scientific status of the study resurfaced. Authors identified numerous sources for the “identity crisis,” and called for alternatives for understanding and improving public administration. Criticism on the scientific scrutiny of the study remain a recurring theme. The differences between authors still hinge on their perceptions of what constitutes a proper science, how study and practice should be related, and last but not least, how to conceptualize public administration. The latter topic has taken a new turn late twentieth, early twenty-first century as some authors argue “public administration” is too limited a concept and/or too closely associated with the machinery of government. They suggest that “governance” perhaps better captures the contemporary intertwinement of societal “steering” by both public and private organizations and administrations.^[13]

CONCLUSION

Since the mid-1600s, the need for scientific study of (public) administration has been acknowledged repeatedly. Both differentiation and integration of knowledge were advocated, normative foundations were disputed, and the need for a practitioner and/or academic orientation was (re)stated. There is an ongoing flip-flopping between limiting and specializing within the science of administration and attempts for synthesis and interdisciplinarity. Despite general acknowledgment of the existence of an impressive body of knowledge on public administration, the science of administration still does not have a well-defined identity or self-image in the early twenty-first century.

REFERENCES

1. Dunsire, A. *Administration: The Word and the Science*; Martin Robertson: London, 1973.
2. Von Seckendorff, V.L. *Teutscher Fürsten Stat*; Reprint 1656 Edition; Glashütten: Taunus, Germany, 1976.
3. Delamare, M. *Traité de la Police: Où l'on trouvera L'Histoire de son établissement, Les fonctions et les prerogatives de ses magistrats; Toutes les loix et tous les reglemens qui la concernent*; Brunet: Parijs, Vol. 4, 1703–1738.
4. Von Justi, J.H.G. *Die Grundsetze zu der Macht und Glückseligkeit der Staaten; oder ausführliche*

- Vorstellung der gesamten Policy-wissenschaft*; Verlag seeliger Johann Heinrich Hartungs Erben: Königsberg, Germany, 1760.
5. Bonnin, C.-J. *Principe d'Administration Publique*; Renaudiere: Paris, 1812.
 6. Rutgers, M.R. Beyond Woodrow Wilson: the identity of the study of public administration in historical perspective. *Admin. Soc.* **1997**, 28, 164–188.
 7. Wilson, W. The study of administration. In *Classics of Public Administration*; Shafritz, J.M., Hyde, A.C., Eds.; The Dorsey Press: Chicago, 1887; 10–25.
 8. Goodnow, F. *Politics and Administration: A Study in Government*; Reprint; Russell & Russell: New York, 1900.
 9. Fayol, H. La doctrine administrative et les services publics. *Rev. Int. Sci. Admin.* **1929**, 2, 9–15.
 10. Simon, H. *Administrative Behavior*, 3rd Edn.; The Free Press/Collier Macmillan: New York, 1947.
 11. Waldo, D. *The Administrative State: A Study of the Political Theory of American Public Administration*, 2nd Ed.; Holmes & Meier Publishers: New York, 1948.
 12. Stillman R.J., II. Public administration in the United States. In *The Modern State and Its Study: New Administrative Sciences in a Changing Europe and United States*; Kickert, W.J.M., Stillman, R.J., Eds.; Edward Elgar: Cheltenham, UK, 1999; 53–77.
 13. Frederickson, H.G.; Smith, K.B. *The Public Administration Theory Primer*; Westview: Cambridge, MA, 2003.

Seasonality

Daniel Williams

School of Public Affairs, Baruch College, City University of New York,
New York, New York, U.S.A.

INTRODUCTION

Seasonality is a term frequently associated with forecasting, but it is important to the analysis of any data in time. However, if the time element of the data is not recorded along with other characteristics, it is likely not possible to analyze or correct for seasonality. This entry addresses seasonality for both the forecaster and the non-forecaster. First, seasonality is defined. Next, the effects of seasonality are explained. Methods for removing seasonality from seasonal data are discussed, including complete procedures for implementing classical decomposition and a robust description of differencing, along with a brief introduction to more complex methods. Methods for including seasonality in statistical models are discussed, including Holt-Winters exponential smoothing, ARIMA with seasonality, and seasonal fixed effects regression.

WHAT IS SEASONALITY?

Seasonal patterns repeat over a fixed period of time, usually a year.^[1] In Fig. 1, data are plotted on the Y axis and time is plotted on the X axis, with the left representing the more distant past and the right representing the more recent past. We observe obvious peaks in the data in July and troughs in December. If there is uncertainty, it may be easier to observe seasonality if the X axis of the plot is limited to the length of the suspected period of the seasonal cycle, with sequential cycles are plotted separately as shown in Fig. 2. We call Fig. 2 a *seasonality plot*. Data that tend toward the same pattern of ups and downs over each segment are seasonal. If the overlapping series appear random or have conflicting peaks and troughs, the data are not seasonal.

While seasonality is commonly thought to be an annual phenomenon, it is also possible to have seasonality within other time segments. Fig. 3 uses a seasonality plot to demonstrate seasonality within weeks. These data reflect hours of operation at branches of the New York Public Library. Traffic may reflect a within week pattern, as well as a within day pattern. Fig. 4 uses a seasonality plot to demonstrate seasonality within quarters.

WHY CARE ABOUT SEASONALITY?

The reason we care about seasonality differs between forecasting and other statistical use of time-ordered data. For research in which seasonality cannot be examined due to failure to consider it in design, there is a third sort of concern.

Forecasting with data can be divided into explanatory and extrapolation methods. Seasonality is primarily a concern for extrapolation methods, including, for the purpose of this discussion, any method that uses a time index. Seasonality can impact the variance of the errors in extrapolation methods to the degree that the seasonal variance is a principal source of variance. The math of most extrapolation methods exploits the variance of the errors to produce reasonable forecasts. So unaccounted for seasonality prevents extrapolation methods from providing reasonable forecasts.

For other statistical uses, the problem can be that seasonality confounds the use of summary statistics in determining the fit of the statistical model. If both an independent variable and a dependent variable are seasonal in the same cycle, then standard methods of fitting statistical models will find the data to be correlated, regardless of whether there is any other association. Finding data correlated in this case is not necessarily meaningful. Typical diagnostics such as R^2 or t -tests are likely to be misleading.

For data that incorrectly fail to collect adequate seasonal information, the difficulty has to do with generalizing. Data collected at a particular time of the year may reflect special circumstances that are not representative of the entire year. For example, the type of patients seen at pediatric clinics at the beginning of the school year may be very different than those over the rest of the year, so patient satisfaction information collected at that time may not be generalizable to the whole year. If seasonal data are collected without awareness of the seasonal impact and without including seasonality in the design, research objectives may not be met.

WHAT TO DO ABOUT SEASONALITY?

Let us begin with the research design problem. Seasonality cannot be evaluated in research results, if data are not collected for that purpose. To collect such data, three

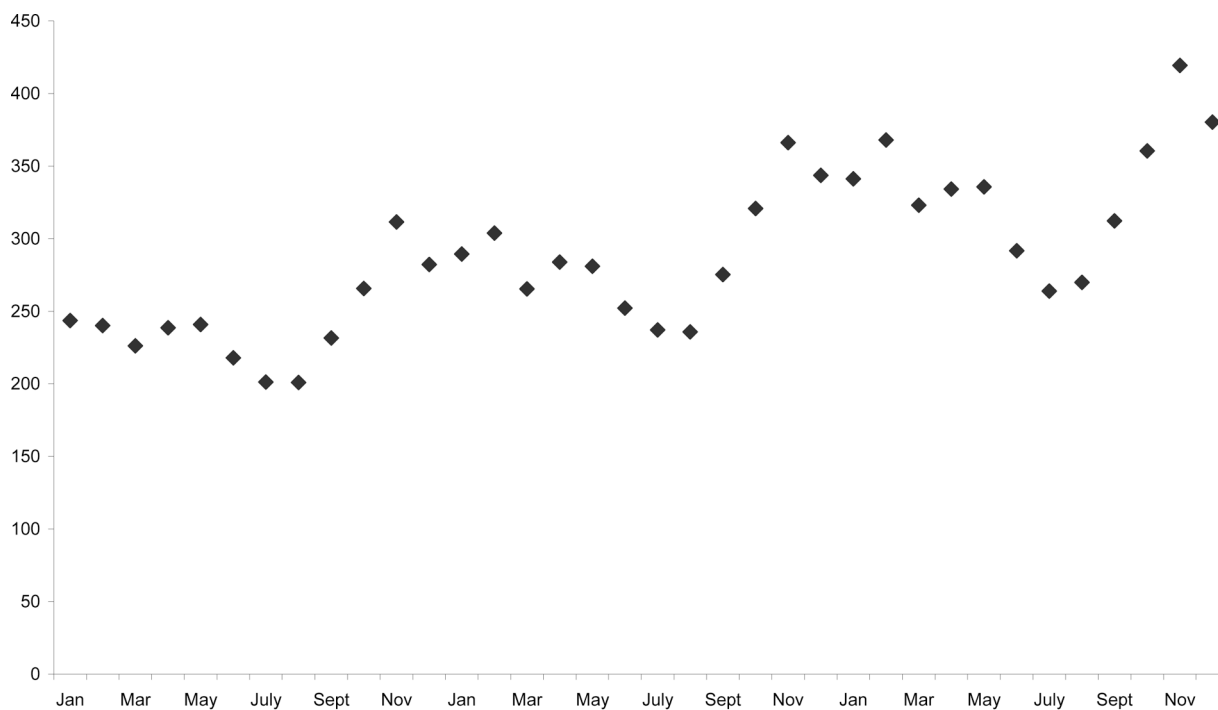


Fig. 1 Seasonal data.

conditions must be met. First, the study must extend over a long enough period of time for the seasonal effect to be observed. Second, the seasonal aspect of the data, that is the time at which the data are collected, must be recorded.

Third, the data collection should not be disproportionate with respect to seasonal activity. One cannot collect a handful of observations all year long, then a great mass of observations at a particular convenient time, and

Overlapping Years

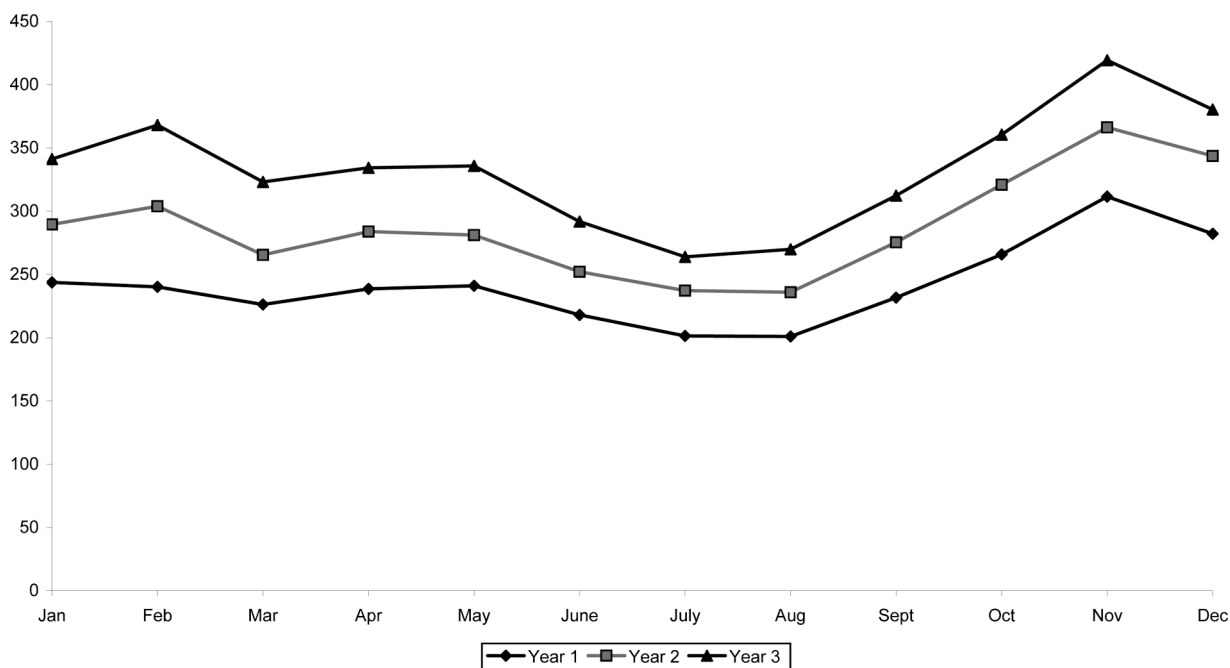


Fig. 2 Seasonal plot.

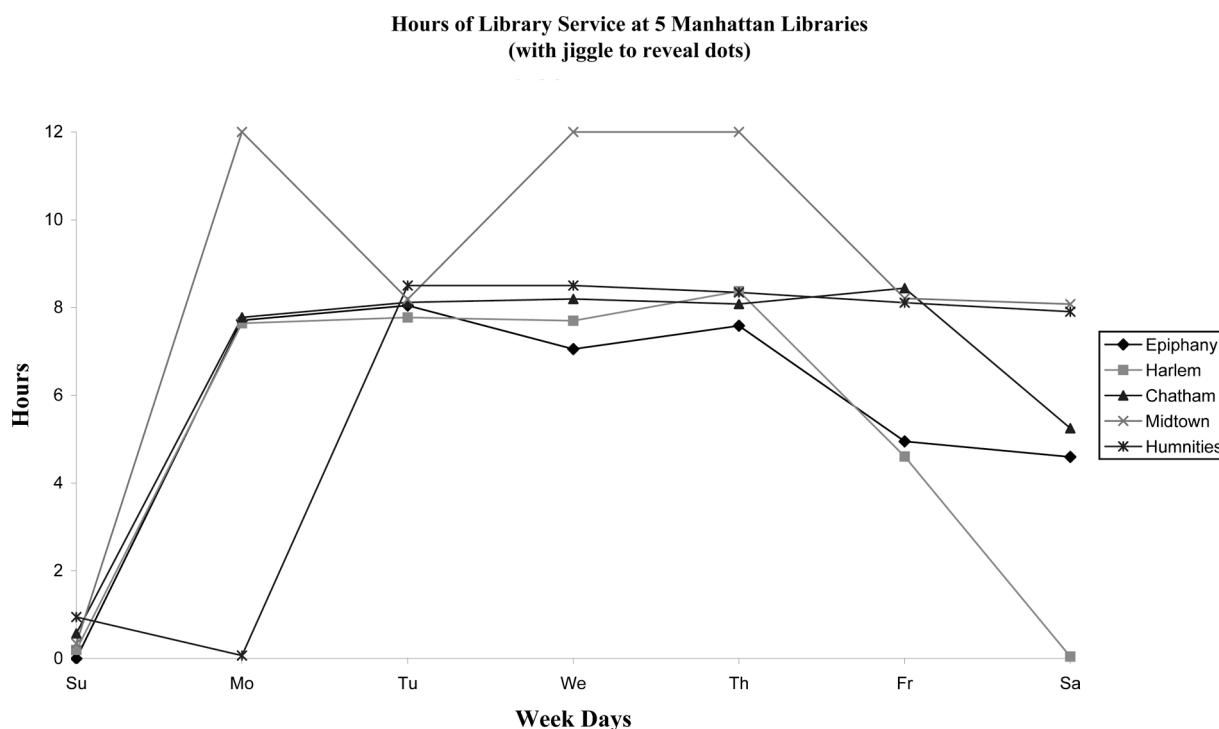


Fig. 3 Seasonality within weeks.

expect to get a reasonable interpretation of any seasonal effect. Data should be either evenly collected over the year, as with a survey of the population at large, or proportional to the data that are available, as with the number of service recipients coming to an organization's office.

Of course, it is often inconvenient to collect data over an extended period of time. The consequence of this is that, although seasonality remains a factor, its impact cannot be observed. Other remedial or preventive steps can be taken in some instances. For example, with longitudinal studies, it may not be economically feasible to collect data quarterly, which would mitigate against unobserved seasonal impact. However, one can avoid attributing seasonal change to other causes by avoiding changing seasons from one year to the next.

Once data have been collected with seasonality in mind, the next step is to look for seasonality in the data. Except with longitudinal data, such analysis will be risky. Data will vary from period to period and may exhibit an apparent pattern when no seasonality exists. The best way to discover seasonality is to compare multiple years of data. When looking at a single year of data, the presence of a pattern may imply seasonal effect, but more evidence is needed. Some sources of additional evidence are a qualitative explanation of the seasonality or the presence of a similar pattern, with similar data collected at another time. Even if seasonality cannot be firmly established or rejected, many

parameters estimated over the entire seasonal cycle are unbiased with respect to seasonality.

Unfortunately, the solutions that apply to the simple research design problem do not extend into forecasting or time-ordered regression models. These problems require addressing seasonality itself. There are two basic ways to do this: one is to take seasonality out of the data; the other is to put seasonality into the statistical model. Reasonable people disagree on this topic. This author is a proponent of taking seasonality out of the data. A reason for this will be offered later in this article. However, because of the equally popular opposing view, three methods of putting seasonality into the statistical model are also discussed.

Taking Seasonality out of the Data

The seasonal adjustment techniques discussed here rest on the assumption that data are systematically collected over at least two seasonal cycles, preferably three or more. This article introduces only the simplest techniques. Some directions are given for exploring more sophisticated techniques.

Sometimes, seasonality depends on the level (current mean) of the series at the time of the season. Other times seasonality is unrelated to the level. Judge this by asking whether the seasonal difference is additive

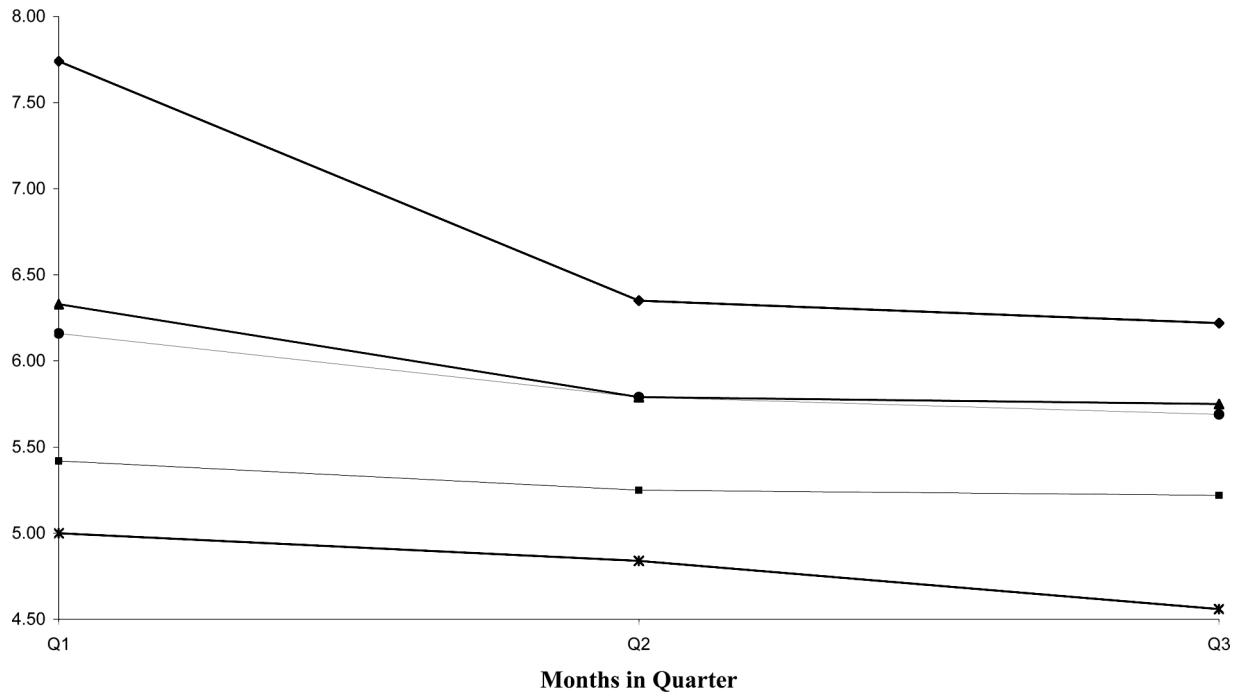


Fig. 4 Within quarter seasonality.

(more like “100 units more in August”) or multiplicative (more like “25% more in August”). Fig. 5 shows multiplicative seasonality around linear growth of 10 units a month. With larger values of the level, the

seasonal peaks and troughs get farther away from the line. Fig. 6 shows additive seasonality around the same 10 units per month growth; in this figure, the size of the peaks and troughs is unrelated to level.

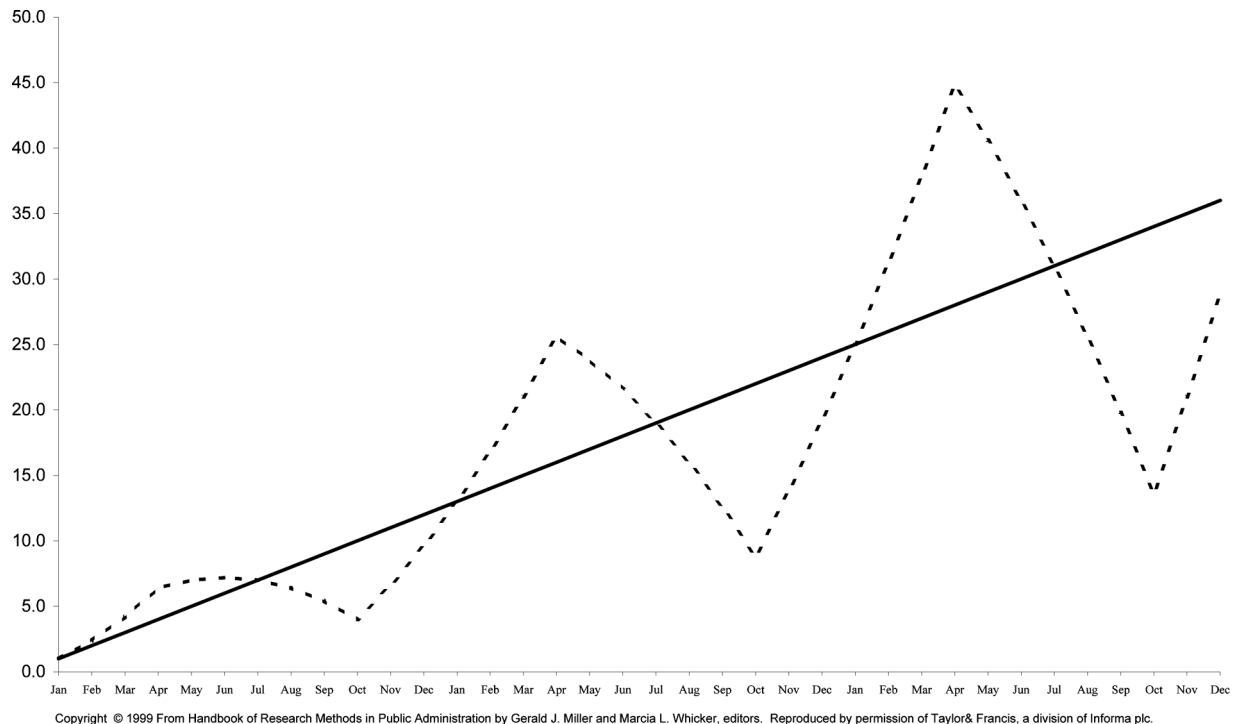


Fig. 5 Multiplicative seasonality.

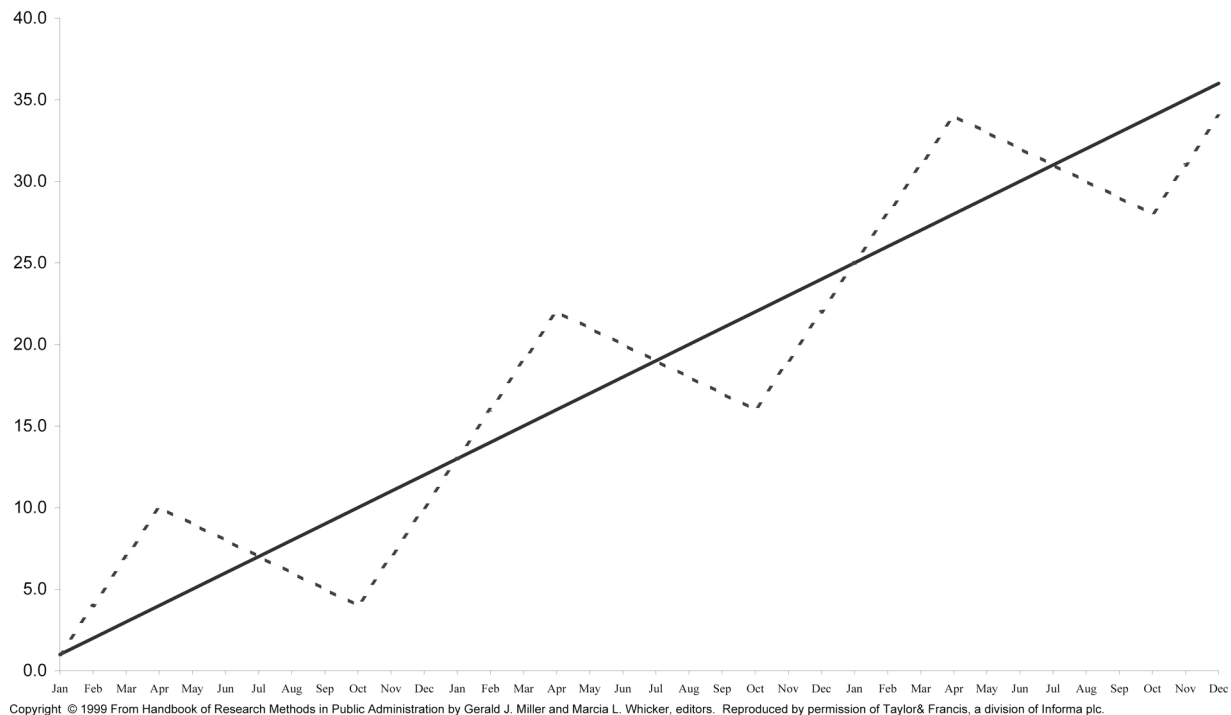


Fig. 6 Additive seasonality.

Classical Decomposition

Deseasonalizing means adjusting a series to remove seasonal impact. The following shows the calculation of a simple annual seasonal index for monthly data for both multiplicative and additive techniques. This technique requires a minimum of two seasonal cycles but works better with three or more seasonal cycles.

First, calculate a *double moving average* as follows: calculate a 12-period single moving average using:

$$MA_{(L+1)/2} = (X_t + X_M + \cdots + X_{t-(L-1)})/L$$

Where $L = 12$ (for monthly data). The single moving average (MA) is shown in the third column of (Table 1).

Then calculate the double moving average (shown as TC in fourth column of Table 1).

$$TC_{t+6} = (MA_{(L+1)/2} + MA_{(L+1)/2+1})/2$$

This new moving average is a 12×2 *double moving average*. For seasonal periods other than monthly, calculate an $L \times 2$ moving average, where L is the number of periods for one seasonal cycle.

The *center* of a moving average is found by the expression $(L + 1)/2$. For example, the center of the moving average for data over January through December (points 1–12) occurs at the end of June and the beginning of July. Since June is point number 6 and July is number 7, the midway-between point is 6.5. The next

moving average covers February through the next January, points 2–13, and, by the same reasoning as before, the midway point is 7.5. The first double moving average value represents the center points 6.5 and 7.5, so its center is at 7, which is July in this example. This is why we label the TC associated with time t with a subscript for time $t + 6$.

To prepare to calculate a seasonal index, first enter the result of the TC on the same row as the actual observation for its centered period (July for years starting in January). There are 12 fewer moving average values than original (raw) values (the six observations at the beginning and the six at the end do not have enough observations for a 12×2 period moving average to be centered beside them). You will see in Table 1 that both MA and TC begin on the row labeled July. MA should actually begin at *June and one-half*, but there is no row for June and one-half. TC begins on the correct row.

These calculations produce an estimate of the trend cycle in the data, as shown by the solid line in Fig. 7. The (original) seasonal data are shown with the scatter plot. The centered trend-cycle estimate extends from the seventh period through the $(n - 6)$ th period (the period that is seven periods before the end).

To calculate a *multiplicative seasonal index*, proceed as follows:

Calculate an *approximate index* (I' , fifth column of Table 1) by dividing each actual value by the

Table 1 Seasonality, classical decomposition with normalization

Month	Data	MA	TC	Multiplicative seasonality				Additive seasonality			
				Apx index I'	Average ind. I''	Index I	Multi. deseas.	Apx fact. F'	Avg fact. F''	Factor F	Additive deseas.
Jan	243.7					1.084	224.75219			23.8	219.8
Feb	240.2				12.015	1.143	210.1		11.070	41.7	198.4
Mar	226.2				12	0.990	228.4		12	−3.1	229.3
Apr	238.6			$P =$	0.998764	1.029	232.0	$P =$	0.922468	8.1	230.6
May	241.0					1.011	238.3			3.2	237.8
June	218.0					0.880	247.6			−37.6	255.5
July	201.3	241.7	243.7	0.826	0.822	0.821	245.2	−42.3	−47.607	−48.5	249.9
Aug	200.9	245.6	248.2	0.809	0.805	0.804	250.0	−47.3	−53.173	−54.1	255.0
Sept	231.7	250.9	252.5	0.917	0.918	0.917	252.7	−20.8	−22.717	−23.6	255.3
Oct	265.8	254.1	256.0	1.038	1.046	1.045	254.4	9.7	13.078	12.2	253.6
Nov	311.5	257.9	259.6	1.200	1.193	1.191	261.4	51.9	54.644	53.7	257.8
Dec	282.2	261.2	262.7	1.074	1.086	1.085	260.1	19.5	25.167	24.2	257.9
Jan	289.5	264.1	265.6	1.090	1.086	1.084	267.0	23.9	24.768	23.8	265.6
Feb	303.9	267.1	268.5	1.132	1.144	1.143	265.9	35.3	42.642	41.7	262.2
Mar	265.4	270.0	271.8	0.976	0.991	0.990	268.0	−6.4	−2.173	−3.1	268.5
Apr	283.9	273.6	275.9	1.029	1.030	1.029	276.0	8.0	8.987	8.1	275.8
May	281.1	278.2	280.5	1.002	1.013	1.011	277.9	0.5	4.099	3.2	277.9
June	252.1	282.8	285.4	0.884	0.881	0.880	286.4	−33.2	−36.643	−37.6	289.7
July	237.2	287.9	290.1	0.818		0.821	288.9	−52.9		−48.5	285.7
Aug	235.9	292.2	294.9	0.800		0.804	293.5	−59.0		−54.1	290.0
Sept	275.4	297.6	300.0	0.918		0.917	300.4	−24.6		−23.6	299.0
Oct	320.9	302.4	304.5	1.054		1.045	307.2	16.4		12.2	308.7
Nov	366.2	306.6	308.9	1.186		1.191	307.4	57.4		53.7	312.5
Dec	343.6	311.1	312.8	1.099		1.085	316.7	30.8		24.2	319.4
Jan	341.2	314.4	315.5	1.081		1.084	314.7	25.7		23.8	317.4
Feb	368.0	316.7	318.1	1.157		1.143	322.0	49.9		41.7	326.3
Mar	323.1	319.5	321.0	1.006		0.990	326.3	2.1		−3.1	326.2
Apr	334.2	322.6	324.2	1.031		1.029	324.9	10.0		8.1	326.2
May	335.7	325.9	328.1	1.023		1.011	331.9	7.7		3.2	332.5
June	291.8	330.3	331.8	0.879		0.880	331.4	−40.1		−37.6	329.3
July	263.8	333.3				0.821	321.4			−48.5	312.4
Aug	269.9					0.804	335.8			−54.1	324.0
Sept	312.3					0.917	340.7			−23.6	335.9
Oct	360.5					1.045	345.1			12.2	348.3
Nov	419.3					1.191	351.9			53.7	365.6
Dec	380.3					1.085	350.5			24.2	356.1

corresponding TC_t value:

$$I'_7 = X_7/TC_7, I'_8 = X_8/TC_8, \dots, I'_{n-6} = X_{n-6}/TC_{n-6}$$

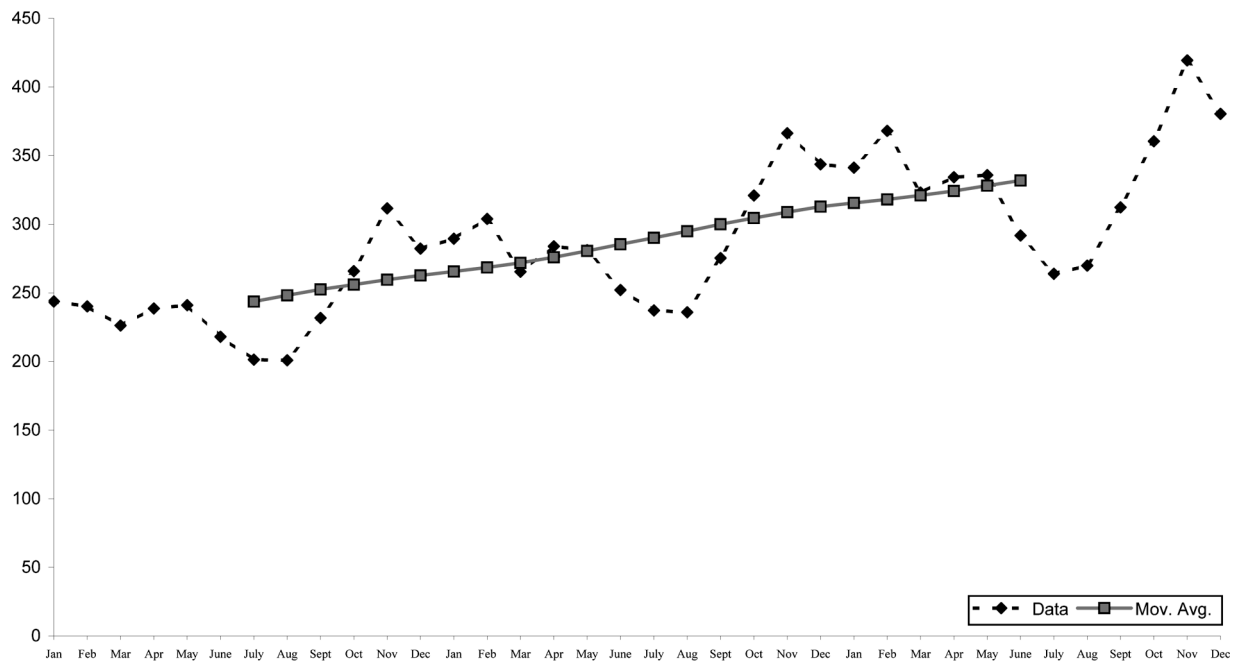
Average the index estimates for each month to get a smoother index (I'' , sixth column of Table 1).

This produces 12 values, one for each month:

$$I''_{JUL} = (I'_7 + I'_{19} + I'_{31} + \dots)/\text{count of Julys}$$

$$I''_{AUG} = (I'_8 + I'_{20} + I'_{32} + \dots)/\text{count of Augusts, etc.}$$

Normalize the index. Sum the 12 seasonal factors (shown above the average, also in the sixth column).



Copyright © 1999 From Handbook of Research Methods in Public Administration by Gerald J. Miller and Marcia L. Whicker, editors. Reproduced by permission of Taylor & Francis, a division of Informa plc.

Fig. 7 Trend cycle.

Use this sum as the denominator in a fraction whose numerator is 12 (or L when the seasonal data are not monthly). The resulting adjustment value, P , is multiplied by each value I''_t to produce an index I_t that sums (over all 12 months) to 12.

$$P = 12 / \sum_{i=1}^{12} I''$$

$$I_{\text{JUL}} = I''_{\text{JUL}} * P$$

$$I_{\text{AUG}} = I''_{\text{AUG}} * P, \text{ etc.}$$

This produces the $L = 12$ seasonal values. The results are shown in the boxed area in the seventh column of Table 1. The index is extended to the entire range of the data by repeating the index number for the month in every cycle.

To obtain deseasonalized data (DESEAS), divide the actual data by the index as shown in the next column of Table 1:

$$\text{DESEAS}_t = X_t / I_t$$

As shown by the circles in the deseasonalized data in Fig. 8, multiplicative classical decomposition substantially mitigates against the seasonal variation in the original data (marked with diamonds).

For *additive seasonality*, the steps mirror those for multiplicative seasonality, but substituting addition and subtraction for multiplication and division, as follows:

Calculate an approximate factor (F') by subtracting each TC_t from the actual observation, as shown

in the first column of the section labeled additive seasonality of Table 1:

$$F'_7 = X_7 - \text{TC}_7, F'_8 = X_8 - \text{TC}_8, \dots, F'_{n-6} = X_{n-6} - \text{TC}_{n-6}$$

If TC is greater than X , the approximate factor will be negative.

Average the factor estimates for each month to get a smoother factor (F''), as shown in the next column:

$$F''_{\text{JUL}} = (F'_7 + F'_{19} + F'_{31} + \dots) / \text{count of Julys}$$

$$F''_{\text{AUG}} = (F'_8 + F'_{20} + F'_{32} + \dots) / \text{count of Augusts, etc.}$$

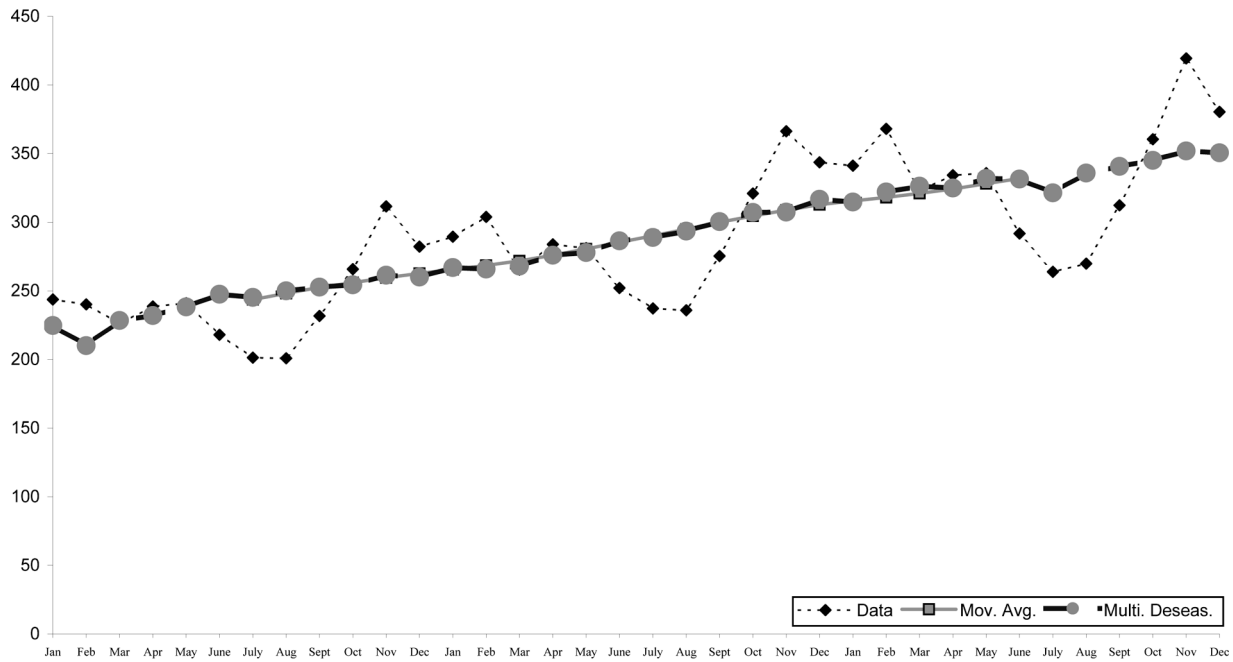
Normalize the index. Sum the 12 seasonal factors. Divide this sum by 12. The resulting value is an adjusting value, P , that is subtracted from each value F''_t to produce an index that sums to 0. When periods are not months, adjust the denominator to the number of periods in the index. This is shown at the top of the second column of the additive seasonality section of Table 1.

$$P = \sum_{i=1}^{12} F'' / 12$$

$$F_{\text{JUL}} = F''_{\text{JUL}} - P$$

$$F_{\text{AUG}} = F''_{\text{AUG}} - P, \text{ etc.}$$

The results of this calculation are shown in the boxed area of the next column of Table 1. They are



Copyright © 1999 From Handbook of Research Methods in Public Administration by Gerald J. Miller and Marcia L. Whicker, editors. Reproduced by permission of Taylor & Francis, a division of Informa plc.

Fig. 8 Multiplicative deseasonalization.

then extended to the full range of the data by repeating the monthly values for every seasonal cycle.

Subtract the factor from the actual data to obtain deseasonalized data (DESEAS):

$$\text{DESEAS}_t = X_t - F_t$$

These results are shown in the last column of [Table 1](#) and in the circles in [Fig. 9](#). In comparing [Fig. 9](#) with [Fig. 8](#), we do not see a substantial difference in the deseasonalization effect. Armstrong argues that when there is modest or no trend, it makes little difference whether one chooses to use multiplicative or additive deseasonalization methods.^[2] The effect is found when the trend is strong. The reason for this effect is that the particular consequence of multiplicative seasonality techniques is that they make the size of the seasonal factors conditional on the size of the short-term mean value of the series. If that value does not change much, then there is little or no effect on the seasonal factors; therefore, the multiplicative seasonal factors mimic additive seasonal factors. It is, therefore, more important to choose the correct method when there is a strong trend, in which case one should return to [Figs. 5](#) and [6](#). If the amplitude of the seasonal plot does not change as the series increases, the seasonality is additive. If the amplitude increases, the seasonality is multiplicative.

This discussion has focused on seasonality of monthly data, with the year beginning in January,

but the actual data may begin in any month, be quarterly data, or have seasonality over some period other than a year. If appropriate adjustments are made, this method can be used with data cumulated over any interval and with any seasonal cycle.

Differencing

An alternative to seasonal adjustment is to calculate the first differences of the seasonal period. That is, the difference between observations that occur at the same point in two sequential seasons. For example, if monthly data follow annual seasonality, as with the previous discussion, calculate differences across years by subtracting the observation in January of year 1 from the observation in January of year 2. For year 3, subtract year 2 data from year 3 data, continuing until the data run out. Observations are now differenced (subtracted one from another) at the same point in the season, so between them there is no seasonal effect. All the new observations are without seasonality.

The new observations are increments of change, not the underlying level of the data. There are many times when this is the most meaningful observation to analyze. However, it is likely that increments of change can be associated only with other increments of change. Thus, if these are data used as a variable in a regression model, and there are other time-ordered variables, they should be differenced as well.

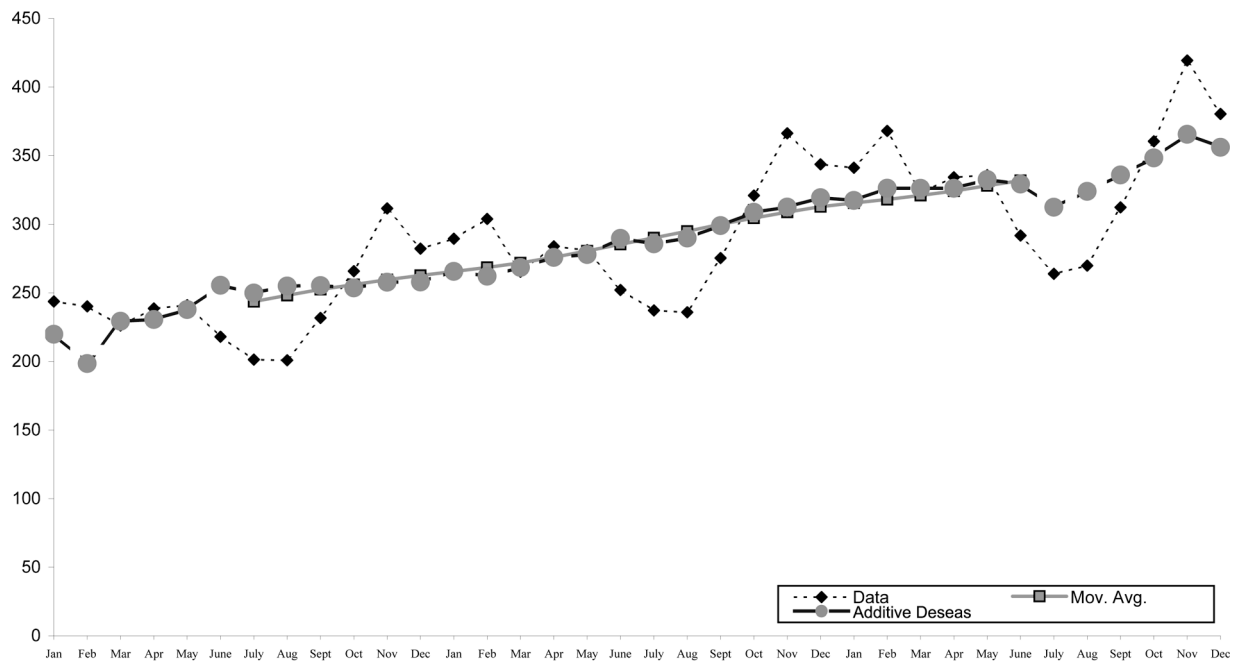


Fig. 9 Additive deseasonalization.

More Sophisticated Decomposition Methods

Classic decomposition has the limitation that its estimated seasonal factors are static, whereas actual seasonal factors may change over time. The reader may refer to the software and documentation available at the U.S. Census Bureau at <http://www.census.gov/srd/www/x12a/> to find more sophisticated methods. This website provides the basic equipment for using a procedure known as X-12, which is a much more sophisticated (and complex) version of the procedure discussed above, and is likely the most commonly used form of seasonal adjustment in the world. Use of X-12 is documented on the website and is also discussed in Ref.^[1] and by Ladiray and Quenneville.^[3] X-12 is a recently updated version of software long known as X-11, and much of the material still refers to the earlier version. The Ladiray and Quenneville text is for the advanced student.

Another advanced method is labeled TRAMO/SEATS. This is a mixed regression and ARIMA method that has been developed by Victor Gomez and Agustin Maravall of the Bank of Spain. Explanation is beyond the scope of this article. Both X-12 and TRAMO/SEATS have been implemented in open-source software such as GRETL (<http://gretl.sourceforge.net/>). X-12 is also available in the SAS statistical package and in a variety of forecasting packages.

Confounding Seasonality

Miller and Williams have shown that some typical seasonal methods confound other sources of variation with seasonal factors.^[4,5] While this demonstration is not made separately for each distinctive method for addressing seasonality, there is no special reason to think that any particular method is immune to the confounding effect. The chief focus of these articles is with the confounding of seasonality with random variation. However, other unexplained variation could also be subject to the same confounding effect. Miller and Williams demonstrate techniques for correcting this confounding effect. As with X-12, these techniques are sophisticated beyond the scope of this article. However, once the analyst has become sufficiently comfortable with the techniques demonstrated here, examination of these articles is strongly recommended. Software (spreadsheets) that implement these procedures can be found at <http://hops.wharton.upenn.edu/forecast/software.html#Freeware>.

Using Deseasonalized Data

Once seasonality has been taken out of the data, the data can be forecast, correlated, or otherwise analyzed, as it might otherwise be without concern for the seasonal impact. If it is forecast, it should be

reseasonalized—that is, the seasonal factors should be used in reverse—to complete the forecast.

PUTTING SEASONALITY INTO THE STATISTICAL MODEL

The other option to dealing with seasonality is to address it within the statistical model one builds. This method is not recommended. There are several reasons for this. First, for forecasting, Miller and Williams have demonstrated that typical methods of accounting for seasonality confound seasonal and random variation.^[4,5] Current methods for correction only exist for classical decomposition and X-12. While the confounding effect has not been demonstrated for each method of seasonal estimation, there are good reasons to believe that it exists. Therefore, it is best to use those methods that can be corrected. Second, one of the most common seasonal forecasting methods, Holt–Winters exponential smoothing, is subject to severe drift. This means that the seasonal factors tend to sum to values other than L within relatively few seasonal cycles. The correction for this is to renormalize the factors. But renormalizing reduces the effect of updating and is very complex. Compromises such as renormalizing once a quarter or once a year are arbitrary. While these may not be sufficiently satisfactory reasons for every forecaster, forecasting experience has led the current author to be reluctant to use seasonal forecasting models.

The reason for deseasonalizing when not forecasting, as with using time series data in regression models, is more apparent. Seasonal factors, which are explained in a later section, substantially affect regression diagnostics. When working with independent and dependent variables that are seasonal in the same period and with the same general seasonal model, one can anticipate strong correlation (and thus, good regression diagnostics), simply because of the seasonality. This effect, however, should be uninteresting to the analyst. The seasonal correlation may mask a complete lack of period-to-period covariation. This masking effect renders the use of seasonal data contrary to good practices.

Nevertheless, the informed reader should know about the principal methods within this class. The following is a brief review of three methods of including seasonality within statistical models. The first two specifically address forecasting models. This discussion is not intended to be exhaustive.

Holt–Winters Exponential Smoothing

A technique developed by Winters permits seasonality to be updated within the Holt exponential smoothing model, a well-known trend-forecasting method.^[6]

These methods have been modestly updated by Williams.^[7] When the seasonal model is additive, one can use the Winters' additive exponential smoothing model.^[8] These models use multiplicative indexes or additive factors to adjust for seasonality in a manner similar to classic decomposition. However, the index is included within an exponential smoothing (also known as exponentially weighted moving average) framework. With exponential smoothing, the current period forecast error is used to update the value of the forecast component; thus, in the Winters' model, the error is used to update the seasonal index for the same period in the next seasonal cycle. In the Holt model, the same error is also used to update two other elements of the forecast. This updating process allows the error to automatically correct the values of the various forecast components over time. There is considerable evidence that the Holt class of models are among the more effective time series forecast techniques.^[9] Thus, Holt–Winters models are a good candidate for integrating seasonality within forecast models. For more discussion of these models, see Chapter 4, on pages 135–184 of Ref.^[1]

ARIMA Modeling

ARIMA is a sophisticated forecasting method that is beyond the scope of this article. Nevertheless, it contains within it techniques for dealing with seasonality. The following symbols, which are specific to ARIMA modeling, show the sort of model that may address seasonality for monthly data, $ARIMA(p, d, q)(0, 1, 1)_{12}$. The “ p, d, q ” symbols within the first parenthesis are parameters that might take on alternate values for different types of data. The second parenthesis specifies a typical seasonal pattern, but as with the first parentheses, the parenthesis can contain (p, d, q) a great variety of parameter values. The subscript 12 refers to a 12-period (annual with monthly data) seasonality. For a more thorough explanation of ARIMA modeling, the reader may want to refer to Chapter 7, on pages 311–387 of Ref.^[1]

Regression with Indicators

In regression modeling, “dummy” or “indicator” variables can be created specifying 11 of the 12 months, three of the four quarters, etc. In basic multiple regression, you should have been taught not to create indicator variables the entire exhaustive set of options; one option must be left for inclusion in the base model. Thus, there are 11, not 12, indicators for months. An indicator is simply a variable that has a value of zero when a condition is not met and a value of one when it is met, so if it is the indicator for January, you would

code the entire column zero (0, not “ ” or “missing”) and then code just the Januaries one (1). Follow the same practice for Februaries through Novembers. If you expect it to be easier to interpret the data with Decembers coded, then omit some other month. The *t*-test for the indicator variable coefficient examines the degree to which the indicator month differs from *the month in the base model*. If none of the *t*-tests for the indicator variable coefficients are statistically significant, then it is arguably the case that there is no measurable seasonality. In the language of econometrics, this approach is seasonal “fixed effects.”

This process consumes 11 degrees of freedom from your regression, so you should be careful in using it. (It is arguably the case that other procedures discussed above sometimes consume degrees of freedom without revealing it.) It is sometimes suggested that this procedure should be tested with the “stepwise” procedure found in many regression packages. The current author discourages such testing.

CONCLUSION

Seasonality is a substantial concern for users of data in time. Procedures are provided that lead to correction of seasonal effect. Because of known modeling defects, it is recommended that statistical modelers use deseasonalized data; however, methods for inclusion of seasonality in statistical models are described. Procedures for classical decomposition are provided, and differencing is robustly described.

ACKNOWLEDGMENTS

I want to thank Don Miller for his assistance in reviewing this article and his helpful suggestions. Errors are my own.

REFERENCES

1. Makridakis, S.; Wheelwright, S.C.; Hyndman, R.J. *Forecasting: Methods and Applications*, 3rd Ed.; John Wiley & Sons: New York, 1998.
2. Armstrong, J.S. *Long-Range Forecasting, From Crystal Ball to Computer*, 2nd Ed.; John Wiley & Sons: New York, 1985, 163 pp.
3. Ladiray, D.; Quenneville, B. *Seasonal Adjustment with the X-11 Method*; Springer: New York, 2001.
4. Miller, D.M.; Williams, D. Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy. *Int. J. Forecast.* **2003**, *19* (4), 669.
5. Miller, D.M.; Williams, D. Damping seasonal factors: shrinkage estimators for the X-12-ARIM A program. *Int. J. Forecast.* **2004**, *20* (4), 529.
6. Winters, P.R. Forecasting sales by exponentially weighted moving averages. *Manage. Sci.* **1960**, *7*, 324–342.
7. Williams, T.M. Adaptive holt-winters forecasting. *J. Oper. Res. Soc.* **1987**, *38*, 553–560.
8. Pfeiffermann, D.; Allon, J. Multivariate exponential smoothing: method and practice. *Int. J. Forecast.* **1989**, *5*, 83–98.
9. Ord, K.; Hibon, M.; Makridakis, S. The M3-competition. *Int. J. Forecast.* **2000**, *16* (4), 433.

Senior Executive Service

Robert F. Durant

*School of Public Affairs, Department of Public Administration and Policy Analysis,
American University, Washington, District of Columbia, U.S.A.*

INTRODUCTION

Promoted and launched during the presidency of Jimmy Carter, the Senior Executive Service (SES) was comprised in 2006 of approximately 8000 career and non-career personnel who staff the top managerial and supervisory positions in the U.S. federal government. A central component of the Civil Service Reform Act (CSRA) of 1978, the SES was established “to provide more effective management of agencies and their functions, and the more expeditious administration of the public business.”^[1] Members of the SES were to serve just below top presidential appointees and afford critical links between appointees and the remainder of the career civil service. Today, they are involved with key government activities in approximately 75 federal agencies. Nearly three decades after its creation, however, the contemporary SES is the legatee of the conflicting purposes of its designers; congressional influence on the implementation of the SES; public attitudes toward government service; and the disparate emphases, strategies, and tactics of individual presidential administrations. Because of the interaction of these factors, the actual performance of the SES has not lived up to the promise envisioned by its creators.

ORIGINAL LOGIC

The culmination of nearly three decades of unsuccessful proposals to create an American higher civil service, the SES originally rested upon several core beliefs.^[2] First, public management skills and knowledge represent a coherent professional craft that is applicable across different types of organizations. Second, a cadre of top public managers could improve the efficiency and effectiveness of public agencies, as well as gain a better appreciation of the public interest, if they were able to escape the confines of traditional single-agency civil service careers. Third, reforms in the areas of pay and flexibility were needed to enhance the recruitment, retention, and development of senior government managers. Finally, the responsiveness of careerists to political executives could and should be maximized by giving the latter greater flexibility to assign, direct, and evaluate the work of the former.

The straightforwardness of these core beliefs belies the extent to which conflicting ideas existed among proponents about the ultimate aims of the SES.^[3] For some, the SES represented the embodiment of an elite European-style corps of generalist civilian executives carrying rank-in-person rather than rank-in-position, a concept rejected by earlier reformers. In contrast, others saw the SES as a vehicle for controlling an allegedly unresponsive federal bureaucracy. Still others viewed the SES as a way either to bring highly touted private sector management techniques into the federal government or to improve stealthily the pay and perquisites of career civil servants. The most talented among the latter had experienced pronounced pay compression early in their careers because maximum executive pay was limited to congressional pay levels.

In trying to accommodate these disparate purposes, the SES focused on four primary areas of personnel management: position allocation, staffing, performance incentives, and executive development.^[4] To combat the highly fragmented set of personnel authorities that existed prior to the CSRA, Congress created the SES as a unified executive personnel management system. It did so by converting formerly GS-16 through GS-18 and Executive Level IV and V positions into a cadre of SES positions filled by career and non-career managers. Persons in those positions were given the option of joining the SES. To facilitate central control of SES allocation while giving agencies flexibility and discretion in staffing SES positions, the CSRA provided that agencies submit requests for SES positions to the Office of Personnel Management (OPM). After reviewing these requests, the OPM was to allocate specific numbers of SES positions for a 2-year period. Agencies then could staff these positions by transferring career appointees to any SES vacancy after giving them 15 days notice of reassignment.

To enhance political responsiveness and provide for lateral or “outsider” appointments from the private sector, the CSRA also allowed a portion of all SES positions to be filled by non-careerist political appointees. The logic of this approach, questioned by many, is that the greater the number of political appointees in an agency, the more control a president has over

an agency's agenda.^[5] Yet to limit politicization and protect merit principles, the CSRA stipulated that no more than 10% of all SES positions could be allocated to non-careerists, and no more than 25% of all SES positions in any single agency (with some exceptions) could be staffed by political appointees.

To these ends, the CSRA designated two separate types of SES positions. The first type—career reserved positions—is for civil servants only because specialized technical or professional expertise is needed to perform the responsibilities of these jobs. The second type—general positions—is open to careerists and political appointees alike because specialized skills are not deemed necessary for performing job duties. Controls also existed against prohibited personnel practices related to merit principles, as did provisions ensuring that SES members cannot be involuntarily reassigned within 120 days of the appointment of a new agency head or non-career supervisor. Neither could they be involuntarily transferred between agencies, be RIFed (reduction in force) in the wake of agency cutbacks or downsizing absent competitive exams, nor be demoted lower than the GS-15 level.

To encourage the participation of careerists in the SES program, to improve the performance of SES members, and to link their performance to organizational goals, the CSRA attached grade and rank to persons rather than to positions. It also replaced largely trait- and tenure-based appraisal with pay-for-performance evaluations and bonus-award systems, while touting executive development programs to improve the skills of SES members. In collaboration with their supervisors, SES members were to set organizational and personal goals annually, and were to be rated by their supervisors in terms of how well they met these goals. Until fiscal year 2004, performance appraisal was premised on a pass-fail system that was scrapped in the fiscal year 2004 National Defense Authorization Act (Public Law 108-136) in favor of more rigorous and realistic ratings (see below).

The SES members also were eligible for rank awards and bonuses for superior job-related performance. Two types of rank awards were created, with SES members eligible to receive either type once every 5 years: a \$10,000 Meritorious Rank Award and a \$20,000 Distinguished Rank Award. In addition, members were eligible to receive a performance bonus of 5–20% of their base salary. To avoid politicization of the awards process and to ensure fairness, Performance Review Boards comprised of a majority of non-career SES members were to determine who qualified for bonuses each year.

By the same token, however, SES members also could be removed from the SES for “less than fully satisfactory” performance. Persons slated for removal could seek an informal appeals hearing with the Merit

Systems Protection Board (MSPB). If not recertified, SES members had the right to a formal MSPB appeals hearing. In no instance were SES members to give up any protections or appeals rights afforded to non-SES members of the federal service in cases of adverse actions. Nonetheless, removal from the SES for performance would not qualify as an adverse action.

THE POLITICAL APPOINTEE–CAREERIST NEXUS

How well have these conflicting SES core beliefs, values, and purposes fared in practice over the past quarter of a century as they relate to the political appointee–careerist nexus in the federal government? Studies of the SES that have relied primarily on staffing and perceptual data of SES members suggest that those who either feared or welcomed politicization of the higher civil service confront a decidedly more complex picture. During the early Reagan presidency (especially 1981–1983), fears or hopes of politicization were well founded. Under the CSRA, after all, political appointees were the ones “monitoring, evaluating, and rewarding or punishing” SES members.^[6] While the Carter administration used these tools, the Reagan administration had the first opportunity to engage in full-scale implementation of the SES provisions. Reagan appointees worked assiduously to use non-career SES positions (as well as even heavier use of Schedule C appointments at the GS 13–15 levels than had their predecessors) to advance the President's policy agenda.

Between 1980 and 1986, for example, the Reagan administration significantly increased the proportion of non-career SES members overall to maximum allowable limits (an increase from 8.4% to 9.8% of all SES positions). The administration also targeted specific agencies for installing disproportionate numbers and percentages of non-career SES members between 1980 and 1983. For example, the total number of non-career SES appointments at the Department of Defense ($N = 88$), the Department of Justice ($N = 50$), the Environmental Protection Agency (EPA; $N = 23$), and the Department of Housing and Urban Development ($N = 32$) reflected increases from the Carter years of 47%, 117%, 28%, and 33%, respectively.^[5,7] Likewise targeted for disproportionate increases in non-career SES appointees were central staff agencies such as the Office of Management and Budget (OMB), the OPM, and the General Services Administration (GSA). Anecdotal evidence also clearly suggests that the Reagan administration used unattractive transfers to coax careerists into accepting (or at least not opposing) Reagan's agenda.

Scholarly research also finds evidence that the Reagan administration “sorted out” career SES members in terms of their likelihood of supporting the President’s agenda. Those most likely to oppose Reagan’s policies were placed in less influential organization positions.^[8] As such, by 1986, the number of SES members indicating that they had “a great deal” of influence over policy dropped to 20% from nearly 50% 5 years earlier, and a significant exodus from the SES ranks occurred.^[6] This exodus was prompted also, however, by Congress subsequently reducing the percentage of SES members eligible for bonuses from up to 50% to 25% of total SES membership in an agency (later reduced to 20% by the OPM).

Yet, data from 1986 to 2000 indicated a decidedly different pattern that led researchers to dismiss politicization as an apt description of what had happened to the SES since the early Reagan years. One researcher summarizing a host of General Accounting Office (GAO) studies concluded in the late 1980s that “[e]vidence of the wholesale politicization of the [SES]...is slight.”^[9] Similarly, one of the nation’s leading experts on these trends concluded in the mid-1990s that efforts to politicize the senior civil service had reached their practical limits by the mid-1980s. Most likely this was because of the complexity of the appointment process generally and the high level of turnover in appointee positions.^[6] Thus, the percentage of non-career appointees fell from 9.3% in 1986 to 8.7% in 1992, whereas the total number of SES positions encumbered by the Clinton administration in 1998 declined by 16% from the last year of the first Bush administration.^[10]

These trends, however, do not mean that President George H.W. Bush and President Clinton failed to use SES appointments strategically to advance portions of their agendas. For example, the total number of SES appointees (i.e., positions filled) increased by nearly 22% (from 6702 to 8130) from 1986 to the end of the Bush administration in 1992. Likewise, although the total number of SES positions in 1998 fell to approximately 6804 positions under the Clinton administration’s “Reinventing Government” initiative, this drop merely returned the number of SES positions to historical levels. Patterns of disproportionate use of non-career SES members in different agencies also did not disappear. Reflecting, for example, the discretion that these heavily outsourced agencies have in awarding contracts, greater numbers and percentages of non-career SES members were placed in agencies such as the EPA, the Department of Energy, and the GSA. What did change, however, was that SES appointments during both the Bush and the Clinton administrations were decidedly less ideological in tone, more concerned about patronage (Bush) or diversity (Clinton)

than policy, and predicated more generally on promoting cooperative relationships with the career bureaucracy.^[11]

Under the George W. Bush administration, the total number of SES members was 7868 in 2004–2005. To advance the goals of the President’s Management Agenda to promote Strategic Management of Human Capital, however, the Bush administration launched a substantial reform of the SES that became part of the fiscal year 2004 Defense Authorization Act. Under the act and a subsequent executive order by Bush, agencies that were not certified by OMB as having rigorous performance standards that could make real distinctions among SES members were given different pay scales from those agencies that were certified. Non-certified agencies could have a minimum rate of basic pay equal to 120% of the rate for GS-15, step 1 (\$109,808 in 2006) and a maximum rate of basic pay equal to the rate for level III of the Executive Schedule (\$152,000 in 2006). Those with OMB-certified performance programs were given a higher maximum pay equal to level II of the Executive Schedule (\$165,200 in 2006). For agencies certified by OMB, the maximum rate of basic pay for senior level and science and technology SES members could match the Vice President’s salary (\$212,100 in 2006).^[12]

Also noteworthy is how much the SES has changed over the years in racial, gender, partisan, and ideological composition under both Democrat and Republican administrations. The representation of women in the SES increased from 10% to 22% between fiscal year 1988 and fiscal year 1999, whereas minority representation increased from 7% to 13%.^[13] By 2003, minority men and women comprised 14% of the SES, and white women comprised 19%. This left the SES still overwhelmingly dominated by white men (67%). The GAO also argues that enhanced efforts by agencies to improve diversity in the SES will be required because of projected retirements; analysts project that racial and ethnic minority representation will increase by only 0.7% if the same mix of appointees in federal agencies continues.^[14] Finally, although Democrats enjoyed a three-to-one margin among supergrade career civil servants in 1968, research indicates that the parties are in rough parity today in the SES.^[10] Leading students of this dynamic attribute current partisan parity to a combination of motivations and trends.^[10] These include turnover in the SES of Democrats disgruntled by the policies and attitudes toward careerists of the early Reagan administration (1981–1984);^[15] Republican presidents using SES appointments to promote and reward Republican careerists; the reflection in careerists of societal trends in party identification; and the partisan repositioning of some in the SES to advance their careers during Republican presidencies.

PROMISE VS. PERFORMANCE

Assessing the political appointee–careerist nexus in the SES tells little directly about how well it has attained the CSRA’s original management goals. Given the absence of pretest performance data, analyses of the differences between pre-SES and post-SES individual, program, or agency performance do not exist. Consequently, most scholarly research on the SES has examined performance only indirectly by assessing the extent to which SES provisions have been faithfully implemented. Most of these studies also rely on the perceptions of SES members rather than on objective implementation metrics. Nonetheless, both types of studies are revealing in noting continuities and discontinuities in perceptions over the decades.

A 1987 study using content analysis of SES performance standards between fiscal year 1980 and fiscal year 1984, for example, found no significant decrease in relying on personal trait evaluations, despite the CSRA’s emphasis on work performance evaluations.^[9] This finding arose amid other studies documenting the Reagan administration’s use of “clustering” career SES managers in smaller and less important programs, “shadow staffing” political appointees to bypass career SES managers, and leaving career SES slots vacant while filling all non-career positions available.

Other studies drew various negative management inferences from research done by the MSPB, the Federal Executive Institute Alumni Association (FEIAA), and the GAO.^[15] Not only had 40% of all participants who joined the SES in 1979 left the federal government by 1987, but also over two-thirds of respondents to FEIAA surveys in 1983 and 1985 argued that the Reagan administration had not created conditions favorable for good management.

As noted above, dissatisfaction also flourished among SES members because Congress failed to provide the originally promised levels of bonuses to them during these years. Reacting to these events, morale among SES members fell from a high of 83% satisfaction in 1980 to a low of 56% in 1983. It then rebounded to 74% over the remainder of the decade after a substantial pay hike by Congress, and returned gradually to initial 1980 levels by 1991.^[10] Moreover, by the late 1990s, a survey of SES members found that an astoundingly high 94% of SES members were either “very satisfied” (56%) or “mostly satisfied” with their jobs (38%).^[10]

Throughout the 1990s, researchers using a variety of more sophisticated statistical techniques also portrayed a somewhat less negative picture of the promise vs. the performance of the SES. In particular, support was found for several of the basic theories underlying the CSRA. For example, analyzing an SES subsample ($N = 4300$) of 21,620 federal respondents to a 1986

MSPB survey, researchers found a strong relationship between performance rewards, individual performance, and agency performance.^[4] They also found that as prohibited personnel practices decreased, both agency performance and public confidence in the agency increased.

Not all the news was good, however. Other researchers revisiting the exit rate question in the early 1990s still found that the second most powerful set of predictors of intent to leave the SES involved the political appointee–careerist nexus. One study, for example, found that the more concerned the SES members were about political control over their work, the more likely they were to be thinking about leaving the SES. In addition, this same study found that the single most powerful predictor of intent to leave the SES was how strongly respondents believed that they were unable to influence public policy. On a more positive note, however, research also suggested that widespread abuses of merit principles were no longer as great a perceived problem. Correlations in one study, for example, were insignificant between perceptions of arbitrary personnel practices and intent to leave the SES.^[16]

Nonetheless, things did not get better after the new pay-for-performance reforms instituted by Congress and the George W. Bush administration for fiscal year 2004 took effect.^[17] In 2006, the Senior Executives Association released the results of a survey of SES members that identified major problems in the program. Among other issues, respondents (with older and longer-serving SES members over-represented) indicated their belief that: 1) ratings quotas or forced distributions by the Bush administration were reducing high ratings; 2) a disconnect existed between rewards and performance; 3) the system was reducing morale and motivation; 4) eliminating SES ranks obfuscated important distinctions among job responsibilities; and 5) greater transparency of the ratings process was necessary.

CONCLUSION: FUTURE OPPORTUNITIES AND CHALLENGES

As one noted observer of the SES puts the dilemma, the concept of an SES “has not been a ringing success in any country (with the possible exception of Australia).”^[18] The SES, however, remains attractive because of its compatibility with “the flexibility in recruiting, assignment, and performance management necessary to achieve other reform objectives” that have swept the United States and abroad.^[18] Its emphasis on employee development is also compatible with what leading experts say is necessary to make the federal service a competitive employer in the

21st century. As the federal government tries to attract an increasingly “in and outer” cadre of federal workers who no longer see their careers in terms of a single organization, the SES emphasis on employee development, performance measurement, rank-in-person, and portability of benefits is well tailored for the multi-organizational and cross-sectoral workforce of the 21st century.

Prior research also suggests, however, that serious challenges will continue to confront the SES. First, to the extent to which the findings of the 2006 Senior Executives Association survey are generalizable and valid for all SES members (not just those with longer tenure), significant reform is needed in the George W. Bush administration’s implementation of pay-for-performance reforms. Second, diversity needs continuing improvement, salaries commensurate with responsibilities and comparable positions in the private sector are critical, and issues involving the appropriate balance between political and careerist influence over policy will remain salient. Third, advances in succession planning in the wake of an aging SES workforce are critical. Occupationally, the highest rates of eligibility for retirements will occur in health system administration and criminal investigation, but all agencies will be affected.^[19] For example, the EPA, the Nuclear Regulatory Commission, and the Veterans Administration will be especially hard hit. Not all those eligible to retire will do so, but unprecedented numbers will leave. These retirees, in turn, will take with them the experiential base, leadership background, and institutional memory and skills that historically have contributed to the allure of creating an SES in the first place.

REFERENCES

1. *Civil Service Reform Act of 1978*. Public Law 95-454, Section 3.
2. Rosenbloom, D.H.; Kravchuk, R.S. *Public Administration: Understanding Management, Politics, and Law in the Public Sector*, 6th Ed.; McGraw-Hill: Columbus, OH, 2004.
3. Huddleston, M.W. To the threshold of reform: The Senior Executive Service and America’s search for a higher civil service. In *The Promise and Paradox of Civil Service Reform*; Ingraham, P., Rosenbloom, D.H., Eds.; University of Pittsburgh Press: Pennsylvania, 1992; 165 pp.
4. Perry, J.L.; Miller, T.K. The Senior Executive Service: Is it improving managerial performance? *Public Admin. Rev.* **1991**, *51* (6), 554–563.
5. Nathan, R. *The Administrative Presidency*; Wiley: New York, 1983; Waterman, R.W. *Presidential Influence and the Administrative State*; University of Tennessee Press: Knoxville, 1989; Durant, R.F. *The Administrative Presidency Revisited: Public Lands, the BLM, and the Reagan Revolution*; State University of New York Press: Albany, 1992.
6. Ingraham, P.W. *The Foundation of Merit: Public Service in American Democracy*; Johns Hopkins University Press: Baltimore, MD, 1995; 99 pp.
7. Ingraham, P.W.; Thompson, J.R.; Eisenberg, E.F. Political management strategies and political/career relationships: Where are we now in the federal government? *Public Admin. Rev.* **1995**, *55* (3), 263–272, specifically 265.
8. Aberbach, J.D. The President and the Executive Branch. In *The Bush Presidency: First Appraisals*; Campbell, C., Rockman, B., Eds.; Chatham House: Chatham, NJ, 1991; 223–247.
9. Yeager, F.A. Assessing the Civil Service Reform Act’s impact on senior manager work priorities. *Public Admin. Rev.* **1987**, *47* (5), 417.
10. Aberbach, J.D.; Rockman, B.A. *In the Web of Politics: Three Decades of the U.S. Federal Executive*; Brookings Institution Press: Washington, DC, 2000.
11. Michaels, J.E. *The President’s Call: Executive Leadership from FDR to George Bush*; University of Pittsburgh Press: Pennsylvania, 1997; Schultz, D.A.; Maranto, R. *The Politics of Civil Service Reform*; Peter Lang: New York, 1998.
12. www.opm.gov/ses/compensation.asp (accessed October 22, 2006).
13. U.S. General Accounting Office. *Senior Executive Service: Diversity Increased in the Past Decade*; U.S. General Accounting Office, GAO-01-377; Washington, DC, 2001; 3 pp.
14. U.S. General Accounting Office. *Senior Executive Service: Enhanced Agency Efforts Needed to Improve Diversity as the Senior Corps Turns Over*; U.S. General Accounting Office, GAO-03-34; Washington, DC, 2003; 3 pp.
15. Ingraham, P.W. Building bridges or burning them? The president, the appointees, and the bureaucracy. *Public Admin. Rev.* **1987**, *47* (5), 425–435.
16. Wilson, P.A. Power, politics, and other reasons why senior executives leave the federal government. *Public Admin. Rev.* **1994**, *54* (1), 12–19.
17. American Society for Public Administration. Performance management system “lost in translation.” *PA Times* **2006**, *October*, 2–3.
18. Ingraham, P.W. Play it again Sam; It’s still not right: Searching for the right notes in administrative reform. *Public Admin. Rev.* **1997**, *57* (4), 325–331, specifically 330.
19. U.S. General Accounting Office. *Senior Executive Service: Enhanced Agency Efforts Needed to Improve Diversity as the Senior Corps Turns Over*; U.S. General Accounting Office, GAO-03-04; Washington, DC, 2003; 2 pp.

Seniority

Alan L. Saltzstein

Division of Political Science and Criminal Justice, California State University, Fullerton, California, U.S.A.

INTRODUCTION

Seniority is best defined as the granting of preference in certain personnel actions based on an employee's length of service. Seniority refers to salary increases, promotion, and the receipt of other benefits, based on the amount of time in service. Seniority is considered a valid principle for personnel actions because years within a position may provide one with superior organizational insight, experience, and loyalty.^[1]

ORIGINS OF SENIORITY IN PUBLIC ADMINISTRATION

Many consider seniority an element of an efficient bureaucratic organization. Weber, for example, lists "promotion according to seniority or to achievement" as an important component of an efficient bureaucracy.^[2] He argued that entrance into an office "is considered an acceptance of a specific obligation of faithful management in return for a secure existence."^[3] Employment security then contributes to the objective administration of duties "where legal guarantees against arbitrary dismissal or transfer are developed, they merely serve to guarantee strictly objective discharge of specific office duties free from all personal considerations."^[3] Thus, secure employment, of which seniority is a major component, provides the officer with protection from political interference and encourages loyalty to the goals of the organization. Generally, then, seniority encourages organizational loyalty and commitment to the objective application of laws.

The concept, however, is frequently viewed negatively in the American context because of perceived conflicts with three other principles of personnel action: merit, political direction, and equal employment. Merit is generally defined as the attempt to make employee competence the major criteria affecting employee decisions. Advances obtained by seniority are believed to favor those with lower qualifications for the position over those hired or promoted through comprehensive testing or analyses of skills and accomplishments. Political direction refers to the link between the aims of political leaders and bureaucratic

response. Seniority appointments are generally less likely to follow the lead of political supporters than those who are politically appointed. Equal employment means treating all groups equally within the personnel process. This includes those groups and people who have been historically discriminated against. Seniority is viewed as a policy likely to favor those people and groups who have historically held positions. Often those who hold existing positions advanced, while those who faced discrimination were held back. Thus, in organizations where discriminatory practices favored the hiring and retention of White males, seniority is generally a policy that encourages those discriminatory patterns to continue.

Seniority and Merit

Conflicts between seniority and merit are common when employees are unionized and when collective bargaining and negotiation influence personnel decisions. Union members generally want seniority to influence policies related to promotion and layoff of employees. Merit principles would encourage promotions and layoffs based on performance or objective testing. Unions often support application of the merit principle in the hiring of employees but seek to limit its application in promotion and layoffs. They argue that the hiring process should be the major application of merit in public service. They maintain that promotions and layoffs ought to be governed by seniority, to protect the interests of the worker and promote employee stability and loyalty.

Seniority and Political Appointment

Concerns for seniority also conflict with the interests of political leaders. When the spoils system predominated, hiring, pay, and promotion were determined by the party in power. Thus, unless an employee was politically active and the party was victorious, job security was unlikely. The advent of civil service, of course, limited the extent to which those in power could control appointments. The aim was a bureaucracy guided by those with extensive knowledge and experience in the managing of government.

Interest in greater political direction of the administration frequently leads to calls for political appointees. Those who support the values promoted by seniority, stability, expertise, and non-partisan competence often oppose such efforts. The Senior Executive Service provisions of the 1978 Civil Service Reform Act, for example, increased the numbers of politically appointed positions and initially encouraged lateral entry to the Senior Executive Service. Members of the civil service in particular opposed the lateral entry provisions. From the perspective of higher-level civil servants, lateral entry represented a challenge to the positive values promoted by a career civil service. The Senior Executive Service positions were viewed as jobs rightly obtained by those with long service in the bureaucracy.

Seniority and Equal Employment

Basic conflicts often arise between seniority and equal employment of individuals and groups that have experienced discrimination in employment. If groups have suffered discrimination historically, they are less likely to populate the upper regions of the bureaucracy. Seniority often prevented woman and minority members from advancing in the bureaucracy. Further, if seniority influences layoff procedures, discriminated groups will lose many of the gains they have won more recently. “Last hired, first fired” is a commonly heard mantra that often, in fact, reflects the reality of layoff policy in modern government.

Conflicts between layoff policies and equal employment bring forth fundamental disputes within the American value system. The equal protection clause of the Fourteenth Amendment and the 1964 Civil Rights Act are interpreted as supporting the rights of all people to be treated equally in personnel policies. They give standing in court to those charging government with discriminatory personnel policies.

However, the value of seniority in personnel practices was formally written into the 1964 Civil Rights Act through section 703(h), which states:

It shall not be an unlawful employment practice for an employer to apply different standards of compensation or different terms and conditions or privileges of employment pursuant to a bona fide seniority or merit system. Provided such differences are not the result of intention to discriminate because of race, color, and religion, sex or national origin.

This provision was placed in the act in response to supporters of seniority. The reference to intentional discrimination and the use of the term “bona fide” were added as a compromise between those who believed that the seniority provision would curtail the

ability of excluded groups to challenge employment policy and those who advocated seniority.

The presence of section 703(h) within legislation designed to enhance the employment rights of excluded groups has led to a series of challenges often reaching the Supreme Court.^[4]

Conflicts between the two principles remain unresolved, as the court tries to deal with cases of employment discrimination. Decisions have tended to side with seniority, rather than with equal employment. *Wygant v. Jackson Board of Education* focused on a collective bargaining agreement that contained a provision mandating that, when layoffs occurred, seniority would be maintained, except that at no time would there be a greater percentage of minority personnel laid off than the current percentage of employment.^[5] This resulted in the laying off of non-minority teachers, while certain minority teachers were retained. The court sided with the non-minority teacher and restored the teacher’s position. The decision was based in part on the seniority provision in the 1964 Civil Rights Act. Justice Powell, in his majority opinion, stated the case for seniority most clearly:

Many of our cases involve union seniority plans with employees who are typically depended on wages for their day-to-day living. Even temporary layoff may have adverse financial as well as psychological effects. A worker may invest many productive years in one job and one city with the expectation of earning the stability and security of minority. At that point, the rights and expectations surrounding seniority make up what is probably the most valuable capital asset that the worker ‘owns,’ worth even more than the equity in his current home . . . Layoffs disrupt these expectations in a way that general hiring goals do not.^[5]

More recently, the Supreme Court ruled in favor of seniority, even when a change in the seniority system in effect limited the seniority of recently hired female employees. The case involved workers in an AT&T Technologies plant. A prior collective bargaining agreement determined that workers accrued seniority on the basis of plant-wide service, and plant-wide seniority was transferable upon promotion to a more skilled position. A more recent agreement changed seniority, making it dependent on the amount of time in a new position. Women who lost the position had been advanced to receive demotions they would not have sustained under the old agreement. The court, however, sustained the new seniority system, arguing that it was not discriminatory on face value. Justice Scalia argued that a discriminatory impact is not sufficient to invalidate the system: “actual intent to discriminate must be proved. By allowing facially neutral systems to be challenged and entitlements under it altered many years after its adoption, would disrupt

those valid reliance interests that 703(h) were meant to protect.’’^[6]

A vigorous dissent by Justice Stevens argued that the majority interpretation here is contrary to the intent of provision 706(e). The provision, he argued, was never meant to immunize seniority systems from the requirements of Title VII.

CONCLUSION

Seniority remains a guiding principle of American personnel policy. Its strict application conflicts with concerns for merit, political direction, and equal employment. It stands for stability and continuity against those who want to change employment policy to reflect political and social change. The frequency of court challenges focusing on seniority suggests that

its place within the public administration process will remain uncertain.

REFERENCES

1. Berman, E.M.; Bowman, J.S.; West, J.P.; Van Wart, M. *Human Resource Management in Public Service*; Sage Publications: Thousand Oaks, CA, 2001; 367 pp.
2. Weber, M. *The Theory of Social and Economic Organization*; The Free Press: New York, 1947; 334 pp.
3. Gerth, H.; Mills, C.W. *From Max Weber: Essays in Sociology*; Oxford University Press: New York, 1958; 199, 202.
4. Riccucci, N.M. *Women, Minorities, and Unions in the Public Sector*; Greenwood Press: New York, 1990.
5. *Wygant v. Jackson Board of Education*, 476. United States 267, 1986.
6. *Lorance v. AT&T Technologies, Inc.* 490. United States 900, 1989.

Sexual Harassment

Sally Coleman Selden

School of Business and Economics, Lynchburg College, Lynchburg, Virginia, U.S.A.

INTRODUCTION

Twenty years ago, in its landmark case, *Meritor Savings Bank, FSB v. Vinson*, the U.S. Supreme Court ruled that sexual harassment, including a sexually hostile work environment, constitutes unlawful sex discrimination.^[1] Responding to these incidents, the Supreme Court offered multiple decisions in the late 1990s and early 2000s that modified sexual harassment laws and might have slowed the number of sexual harassment claims filed with the Equal Employment Opportunity Commission (EEOC). The recent allegations of sexual harassment against Congressman Foley by a teenage page highlight a growing concern in public and private workplaces, especially in retail and food service establishments, about sexual harassment of teen employees. The EEOC launched its Youth at Work campaign to educate teens and younger workers about discrimination in the workplace. In 2005, the EEOC filed two employment discrimination lawsuits on behalf of teen workers against franchises of McDonald's restaurant in Arizona and New Mexico.^[2] Sexual harassment in the workplace and in educational institutions is not a new issue, but its heightened visibility has forced human resource (HR) managers to review organizational policies and dissemination strategies.

The purpose of this article is to introduce the concept of sexual harassment and discuss the current policy environment.^a First, it provides statistics about incidents of sexual harassment in the workplace. Second, the article provides an overview of the policy and legal frameworks necessary to manage harassment, including pertinent legislation, relevant Supreme Court decisions, and administrative regulations governing sexual harassment. Finally, broad guidelines for designing an organizational sexual harassment policy are outlined.

CURRENT ENVIRONMENT: INCIDENTS AND REPORTING OF SEXUAL HARASSMENT DECLINING, BUT A GAP REMAINS BETWEEN THE TWO

Although sexual harassment remains an area of concern for all organizations, including educational

institutions and their employees, a recent survey shows that complaints of sexual harassment have steadily declined over a 3 yr period.^[3] In 2005, 48% of corporate attorneys and HR managers surveyed indicated that their companies had no complaints of sexual harassment, up from 44% in 2004 and 37% in 2003.^[3] Sexual harassment filings with the EEOC and the Fair Employment Practices agencies (FEPA) have also decreased during this same period (see Table 1). Between 1992 and 2001, the number of sexual harassment charges filed with the EEOC under Title VII grew 46.9% from 10,532 to 15,475. Between 2001 and 2005, the EEOC shows an 18% decline in formal sexual harassment charges filed. However, staffing levels at the EEOC during this 4 yr period fell by 19%. Without further study, it is hard to discern whether the decline in formal complaints with companies, the EEOC, and Fair Employment Practices reflects an actual decline in incidents of sexual harassment, a lack of willingness to report such incidents, understaffing, or better training. For example, in 2002, 21% of women and 7% of men surveyed reported that they had been sexually harassed at work.^[5] By 2005, according to a Louis Harris poll, the percent of women experiencing harassment grew to 31%, while the percentage of men remained steady at 7%.^[6] Moreover, the poll showed that 62% of persons experiencing harassment had never reported it to their organizations.^[6] Given that the percentage of complaints filed by men has steadily increased since 1992 (see Table 1), men may feel more empowered to report incidents of sexual harassment in the workplace.

Studies have shown that both women and men experience harassment and that some harassment occurs repeatedly and for long periods of time.^[7,8] However, many individuals who are sexually harassed elect not to report the incident or to use an employer's sexual harassment policy.^[9] Because of this tendency, the number of formal complaints filed may underestimate the pervasiveness of sexual harassment in an organization. While the evidence suggests that sexual harassment is an enduring and pervasive workplace concern, it is difficult to ascertain the exact magnitude of the problem because individuals define and experience sexual harassment differently and often elect not to share their experience with the organization formally.^[6]

^aThis article is an abbreviated and updated version of the author's previously published work.

Table 1 Sexual harassment charges filed with EEOC and FEPA under title VII FY 1992–FY 2005

Fiscal year	Number of charges filed	Percentage filed by men (%)
1992	10,532	9.1
1993	11,908	9.1
1994	14,420	9.9
1995	15,549	9.9
1996	15,342	10.0
1997	15,889	11.6
1998	15,618	12.9
1999	15,222	12.1
2000	15,836	13.6
2001	15,475	13.7
2002	14,396	14.9
2003	13,566	14.7
2004	13,136	15.1
2005	12,679	14.3

Abbreviations: EEOC = Equal Employment Opportunity Commission; FEPA = Fair Employment Practices Agencies.
(From Ref.^[4].)

THE PUBLIC POLICY ARENA OF SEXUAL HARASSMENT

Relevant Legislation

Sexual harassment is a complicated area of employment law, partly because employers must comply with both state and federal laws, which have been interpreted differently by the courts. Before turning to a discussion of the judicial decisions that provide broader policy direction to the interpretation of federal law, this section briefly reviews laws that govern sexual harassment cases and workplace policies in both the public and the private sector. Practitioners should note that when state sexual harassment laws offer greater employee protection, they take precedence over federal guidance. Illinois, for example, was the first state to ban sexual harassment of state employees. Considerable variation exists among states with respect to the legal doctrine governing sexual harassment. Some states, such as New Hampshire, New Jersey, North Dakota, Pennsylvania, Rhode Island, West Virginia, and Wisconsin, have antisexual harassment laws. These antisexual harassment laws employ EEOC's definition of sexual harassment, but they vary in terms of policy requirements and punishments. For example, in Rhode Island, employers with 50 or more employees are required to adopt and distribute a workplace policy and statement against sexual harassment. In 2004, California passed legislation requiring employers of 50 or more to provide 2 hr of sexual harassment

prevention training to supervisors every 2 yr. Wisconsin's law requires the board of regents to include sexual assault and sexual harassment in its orientation program for newly entering college students.

At the federal level, three laws are of particular significance to employers: Title VII of the Civil Rights Act (CRA) of 1964; Title IX of the Education Amendments of 1972; and the CRA of 1991. Title VII of the CRA of 1964 prohibits employment discrimination based on race, sex, national origin, or religion. It applies to public, non-profit, and private organizations, with one exception (see Ref.^[10] for a discussion of how the law pertaining to a hostile environment varies among employers). Military personnel are not covered by Title VII, although Department of Defense civilian employees are.^[11] Over time, Title VII prohibition against sex discrimination has been interpreted to include sexual harassment. In 1976, a federal district court found that firing an employee for refusing the sexual advances of a supervisor was illegal under Title VII.^[12] The next year, three federal courts of appeals reversed lower court decisions that had held that sexual harassment was not within the scope of Title VII.^[13–15] In 1980, the EEOC issued regulations that defined sexual harassment as a form of sex discrimination prohibited by the CRA of 1964. These EEOC regulations were validated by the 1986 Supreme Court Case of *Meritor Savings Bank, FSB v. Vinson*.^[1]

Title IX of the Education Amendments of 1972 prohibits sex discrimination in schools and school programs that receive federal funding. The law applies to elementary, secondary, and postsecondary schools and is enforced by the Office for Civil Rights of the Department of Education. When sex discrimination has occurred, schools, colleges, and universities are given an opportunity to correct the problem. If an appropriate action is not taken by the educational institution, federal funding may be revoked.

Lastly, the CRA of 1991 includes provisions that expand a victim's right to sue and collect compensatory (punitive) damages for sexual discrimination or harassment. Prior to this legislation, plaintiffs could only collect backpay and attorneys' fees. In an organization with more than 200 employees, for example, a victim of sexual harassment may recover a combined award of compensatory and punitive damages up to \$300,000. Moreover, and arguably, the most significant provision of the 1991 CRA grants sexual harassment plaintiffs the right to jury trials.

Supreme Court Decisions

Defining sexual harassment under Title VII of the CRA of 1964: The courts have identified two basic types of sexual harassment that are actionable under

Title VII—quid pro quo and the hostile work environment. Much of the case law defining the parameters of sexual harassment emerged first from lower courts. For example, the 1981 judicial decision *Bundy v. Jackson* was critical in forming today's legal definition of sexual harassment.^[16] The Court ruled that Bundy was the victim of a practice of sexual harassment and a discriminatory work environment, and her rights under Title VII were therefore violated. The sexually stereotyped insults and demeaning propositions to which Bundy was subjected caused her undue anxiety and created an "offensive or hostile" work environment. Thus, a two-dimensional approach to sexual harassment claims emerged. Sexual harassment claims could be classified as quid pro quo or hostile work environment.

Five years after the *Bundy* case, the Supreme Court issued its first landmark decision regarding sexual harassment—*Meritor Savings Bank, FSB v. Vinson*.^[17] The Court affirmed the distinction that emerged from *Bundy v. Jackson*, holding that two types of harassment are actionable under Title VII: Harassment that involves the conditioning of employment benefits on sexual favors (quid pro quo) and those where a hostile environment is created.^[16] Mechelle Vinson testified that Sidney Taylor, her supervisor and a vice president of Meritor Savings Bank, repeatedly demanded sexual favors from her, usually at the bank both during and after work hours. Finding for Vinson, the Court ruled that non-economic injury was a form of sex discrimination covered by Title VII, but the Court did not address employer liability with the same clarity. The Court strongly agreed with the EEOC that Congress wanted the courts to apply agency principles—defining a supervisor as an agent of the employer and imputing responsibility to the employer when a supervisor exercises authority delegated to the supervisor by the employer—to guide decisions on employer liability. Nevertheless, the Court remanded the case to lower courts, to determine the conditions under which employers are held responsible. The Supreme Court disagreed with the Court of Appeals decision "that employers are always automatically liable for sexual harassment by their supervisors... For the same reason, absence of notice to an employer does not necessarily insulate that employer from liability." While the *Meritor* decision offered clarity to the legal definition of sexual harassment, it left considerable uncertainty around what forms of behaviors constitute harassment, especially those that create a hostile environment, and the extent of employer liability or responsibility for such an environment.^[10]

By the early 1990s, lower court rulings varied tremendously, concerning the severity and pervasiveness of behaviors required to create a hostile work environment.^[10] In 1993, the Supreme Court addressed

this issue in its decision, *Harris v. Forklift Systems, Inc.*^[17] The Court ruled that victims need not show that they suffered serious psychological injury as a result of the harassment, in order to prevail in court. The Court affirmed the standard stated in *Meritor Savings Bank v. Vinson*—"Title VII is violated when the workplace is permeated with discriminatory behavior that is sufficiently severe or pervasive to create a discriminatorily hostile or abusive working environment" as determined or perceived by a "reasonable" person. While the Court did not elaborate on the definition of a "reasonable" person, the EEOC has stated that "the reasonable person standard should consider the victim's perspective and not stereotyped notions of acceptable behavior."^[18] The issue is whether a "reasonable" person in the "victim's circumstances would have found the alleged behavior to be hostile or abusive."^[18]

In this case, Charles Hardy, the president of the company, had referred to Harris, the plaintiff, as a "dumb-ass woman," made unwanted sexual innuendos to her, and asked her (and other female employees) to get coins from his front pants pocket. Harris complained to Hardy about his conduct, who promptly apologized and promised to cease. Within a month, Hardy resumed his behavior, leading Harris to quit and, subsequently, file suit against Forklift Systems. The company argued that because Teresa Harris had not demonstrated any obvious psychological harm as a result of Hardy's actions, Title VII was not violated. Finding for the plaintiff, the Court unanimously ruled that Title VII "comes into play before the harassing conduct leads to a nervous breakdown. A discriminatorily abusive work environment, even one that does not seriously affect employees' psychological well-being, can and often will detract from employees' job performance, discourage employees from remaining on the job, or keep them from advancing their careers."

In 2001, the Supreme Court stressed the legal differences between single offensive comments and a hostile environment. In its decision, *Clark County School District v. Shirley A. Breeden*, the Supreme Court ruled that no one reasonably could believe that a single comment with sexual overtones violates Title VII of the CRA of 1964.^[19] Sexual harassment is only actionable if it is "so severe or persuasive" as to "alter the conditions of the victim's employment and create an abusive working environment." Shirley Breeden met with her male supervisor and another male employee to review four job applicants' psychological evaluations. One evaluation contained an applicant's sexually charged comment. Breeden's male supervisor read the comment, "I hear making love to you is like making love to the Grand Canyon" aloud and remarked, "I don't know what this means." To which, the male

colleague stated “I’ll tell you later.” After the incident, Breeden complained about the exchange to the offending supervisor and to two other supervisors in the organization. Breeden charged that she had been harassed and retaliated against because she had been transferred after her complaint. The Court ruled, because the remark did not constitute unlawful sexual harassment under Title VII, that Breeden did not have a valid claim of retaliation.

Prior to 1998, the Supreme Court had not heard a harassment case involving persons of the same sex. The Court ruled in *Oncale v. Sundowner Offshore Services, Inc.* that “nothing in Title VII necessarily bars a claim of discrimination . . . merely because the plaintiff and the defendants . . . are of the same sex.”^[20] In this case, Joseph Oncale worked on an eight-man crew aboard an oil platform in the Gulf of Mexico. Oncale contended that his male supervisors and coworkers abused and sexually taunted him, pushing a bar of soap into his anus and threatening homosexual rape. Although he complained to his employer, the company did not respond, so Oncale quit. The Court judged that Title VII’s prohibition against discrimination “because of sex protects men as well as women.” This decision provides numerous employees, previously overlooked, protection against sex discrimination under Title VII. The Court, however, noted that Title VII “does not prohibit all verbal or physical harassment in the workplace” but makes actionable only conduct “so objectively offensive as to alter the ‘conditions’ of the victim’s employment.” Thus, “ordinary socializing in the workplace—such as male-on-male horseplay or intersexual flirtation” is not illegal. The Court noted that “common sense, and an appropriate sensitivity to social context, will enable courts and juries to distinguish between simple teasing or roughhousing among members of the same sex and conduct that a reasonable person in the plaintiff’s position would find severely hostile or abusive.” The Court, however, gave little guidance for determining when conduct crosses the line to sex-based discrimination, leaving lower courts to work out what proof is required to establish that same-gender harassment is because of sex.^[21] As a result, the Court’s broad, conduct-based analysis might be interpreted by lower courts in ways that allow standards to vary depending on the work environment.

Defining employer liability for sexual harassment under Title VII of the CRA of 1964: Courts have addressed the conditions under which an employer is held liable or responsible for sexual harassment. In 1998, the Supreme Court ruled on two such cases, one involving a public employer and one involving a private employer. Both cases establish identical rules of laws.^[22]

In *Faragher v. Boca Raton*, the Court determined that public employers can be held liable for harassment

under Title VII, even if the employee has not explicitly alerted the employer about the sexual harassment.^[23] However, the employer may raise an affirmative defense to liability or damages depending on the situation. During Beth Ann Faragher’s 5 yr employment as a lifeguard for the City of Boca Raton, Florida, her immediate supervisors repeatedly subjected Faragher and other women lifeguards to “uninvited and offensive touching,” lewd remarks, and denigration of women. The City of Boca Raton asserted that because the harassment was not reported to higher authorities and occurred in a remote location, it should not be held liable. While the city had a sexual harassment policy, it had not widely disseminated the policy. The Court held that this failure defeated any invocation of an affirmative defense to liability. Clarifying conditions of an employer’s vicarious liability, the Court adopted the following stance:

An employer is subject to vicarious liability to an . . . employee for an actionable hostile environment created by a supervisor . . . When no tangible employment action is taken, a defending employer may raise an affirmative defense to liability or damages . . . The defense comprises two necessary elements: 1) that the employer exercised reasonable care to prevent and correct promptly any sexually harassing behavior and 2) that the plaintiff employee unreasonably failed to take advantage of any preventive or corrective opportunities provided by the employer or to avoid harm otherwise.

The Court confirmed that no affirmative defense is available when an employee is terminated, demoted, or undesirably reassigned.

In a separate but related decision about employer liability, *Ellerth v. Burlington Industries*, Kimberly Ellerth’s second-level supervisor, Ted Slowik, repeatedly threatened her job unless she succumbed to his advances.^[24] She never registered a formal complaint nor was her job ever jeopardized, but after 15 mo as a salesperson for Burlington Industries, she resigned because of the constant harassment. Burlington Industries argued that it should not be held liable because Ellerth had suffered no job consequence and had even been promoted; moreover, she had failed to use the company’s sexual harassment complaint procedure.

The Court stated that Congress gave an explicit instruction to the courts to interpret Title VII based on agency principles, by defining the term “employer” to include its “agents.” Building on this rationale and consistent with its decision in *Meritor*, the Court confirmed in *Ellerth* that the employer is strictly liable under the quid pro quo theory for a supervisor’s harassing or discriminatory actions that lead to significant changes in a person’s employment status. Moreover, consistent with its ruling in *Faragher*, the Court judged

that an employer could be held liable in situations where a supervisor causes a hostile work environment, even when the employee suffers no tangible job consequence, and the employer was unaware of the offensive conduct. That is, the employer may be liable where a supervisor causes a hostile work environment by his or her severe or pervasive inappropriate conduct, subject to an affirmative defense. Based on this case and *Faragher*, employers, both public and private, are best protected by developing and disseminating an antiharassment policy. However, if an employee has reason to fear retaliation for coming forward and using the policy, the employer is not able to use policy as an affirmative defense. The burden of proof for any affirmative defense rests on the employer.

In *Pollard v. E.I. du Pont de Nemours & Co.*, the Supreme Court ruled that the amount of “front pay” a victim may receive is not subject to the limit of \$300,000, established in 1991 for compensatory and punitive damages. The Court ruled that “when a court found that an employer had intentionally engaged in an unlawful employment practice, the court was authorized to award such remedies as injunctions, reinstatement, backpay, and lost benefits.” Based on this ruling, backpay, known today as “front pay,” includes the period between the date of judgment and the date of reinstatement.

In *Pennsylvania State Police v. Suders*, the Court held in an 8-1 decision that an employer can only be held strictly liable for a supervisor’s harassment if there has been an “official act” that accompanied an employee’s constructive discharge.^[25] While the Court did not clearly define the “official act” standard, the description of a tangible employment action as set forth in *Ellerth* is implied in the new standard, as well as “a humiliating demotion, extreme cut in pay, or transfer to a new position in which [a person] would face unbearable working conditions.”^[25] At the same time, the *Suders* decision provides employers with the opportunity to assert an affirmative defense to a claim of constructive discharge.

Defining employer liability for sexual harassment under Title IX of the Education Amendments of 1972: In the late 1990s, the Supreme Court ruled on two cases involving liability for sexual harassment in educational institutions. The first dealt with student–teacher harassment and the second with student–student harassment.

The 1998 Supreme Court case, *Gebser v. Lago Vista Independent School District*, concerned harassment of a student by a teacher.^[26] The Court ruled in a 5-4 decision that school districts were only liable for student harassment by teachers, if school officials knew of the problem and refused to intervene. In this case, when the school district discovered a teacher, Frank Waldrop, having sex with a student, he was arrested

and promptly fired. Prior to this incident, a group of parents had complained to the school about sexual comments Waldrop made in class, but these concerns had not been reported to the superintendent. Also, the school district had not promulgated or distributed a formal antisexual harassment policy or procedure for lodging such a complaint. The Court, however, viewed Title IX as a contractual framework, distinguishing it from “Title VII, which is framed in terms not of a condition but of an outright prohibition.” The Court was unwilling to subject schools to liability in absence of a clear directive from Congress. The Court voiced concern that the award of damages might unfairly exceed the amount of federal funding actually received by the school. The Court held that a school is liable for teacher–student sexual harassment only when someone with authority to take corrective action has received actual notice of the harassment and has exhibited “deliberate indifference.” In a strong dissent, Justice Stevens warned that the rule crafted by the Court might create a disincentive for schools to take steps to discover harassment and avoid liability through an “ostrich” defense.

In 1999, the Supreme Court tackled the question of a school district’s liability for harassment of a student by another student in *Davis v. Monroe*.^[27] LaShonda Davis was in the fifth grade when she first complained to her teachers that a student sitting next to her was sexually harassing her. Over a 5 mo period, the harassment took various forms, including offensive language, fondling, and sexual abuse. The student and her mother reported the encounters on numerous occasions both to her teachers and to the school principal. The school did not provide any protections to Davis other than, after 3 mo, changing her seat. After the school refused to take action, the child’s mother reported the situation to the sheriff’s department. The harassing student was charged with and, subsequently, pleaded guilty to sexual battery. The Davis family filed against the school district, because it took no disciplinary action to protect the victim. The Court found that the child was the victim of repeated acts of harassment over a 5-mo period and that the misconduct was severe, pervasive, and objectively offensive. Applying the standard set out in *Gebser*, the Court ruled that public schools can be sued and forced to pay damages for failing to stop sexual harassment by students, but only where the school district is cognizant of the incident and is “deliberately indifferent.” In addition, the harassment must be deemed so severe that it deprives the victim of access to educational opportunities or benefits provided by the school district.

The Supreme Court has provided guidance to employers about the parameters of sexual harassment and the conditions under which employers can be held

responsible. The Court has made a clear distinction about employer liability, depending on whether a case falls under Title VII of the CRAs of 1964 or under Title IX of the Education Amendments of 1972. Lee and Greenlaw project that the recent “rulings will lead to greater litigation in the realm of sexual harassment and may emphasize the need for some action by Congress and/or the Equal Employment Opportunity Commission.”^[10]

The EEOC: Providing Guidance and Enforcing the Law

The Equal Employment Opportunity Commission is the agency responsible for publishing regulations governing sexual harassment in the workplace and providing policy and enforcement guidance to help employers interpret and apply EEOC regulations, federal law, and judicial decisions governing sexual harassment.^[28,29] In 1980, the EEOC issued formal guidelines defining sexual harassment for all employers. According to the EEOC, sexual harassment includes “unwelcome sexual advances, requests for sexual favors, and other verbal or physical conduct of a sexual nature ... when submission to or rejection of this conduct explicitly or implicitly affects an individual’s employment, unreasonably interferes with an individual’s work performance or creates an intimidating, hostile or offensive work environment.”^[30] In the workplace, two types of sexual harassment are legally recognized: 1) quid pro quo and 2) hostile work environment. Quid pro quo harassment is defined as a situation where a person demands or requests sexual favors from an employee in return for employment benefits, such as a promotion or salary increase, or as a condition of employment. When an employee refuses to succumb to sexual advances, the harasser denies the employee tangible economic or job benefits. Hostile working environment harassment occurs when behaviors create an offensive or hostile work climate, such as touching, sexual remarks, and overtures. For both types of harassment, the victim, as well as the harasser, may be of either gender, and the victim does not necessarily have to be of the opposite gender as the harasser.

Annually, the EEOC publishes its regulations about sexual harassment in Title 29 of the Code of Federal Regulations. The EEOC also produces policy guidelines to help employers interpret and understand judicial rulings. Recent policy guidance provided by the EEOC, for example, deals with the issue of employer liability in light of the *Burlington Industries* and *Faragher* Supreme Court decisions. In addition to formulating regulations and policy guidance, the EEOC can be involved in the investigation and prosecution of sexual harassment claims. Despite its

multiple roles, the EEOC has continually reiterated that all employers need a written antidiscrimination (including sexual harassment) policy.

FORMULATING A WORKPLACE POLICY

Employers are liable for sexual harassment in their workplace that they are aware of, unless they take prompt and effective actions to remedy the situation. Even if employers are unaware of the harassment, they are still potentially responsible. Employers may limit their liability in hostile work environment cases, by taking appropriate actions and providing avenues to employees to raise harassment complaints. According to the EEOC, “It is necessary for employers to establish, publicize, and enforce anti-harassment policies and complaint procedures.”^[29] Such policies must be in writing and disseminated widely to employees. At a minimum, a sexual harassment policy must address three objectives. First, the policy’s goal must be to prevent incidents of sexual harassment in the workplace. Second, the policy must encourage employees to report incidents as soon as possible so that the employer can correct the problem. Finally, the policy must be consistent with the applicable legislative and case law.

In light of these objectives and as suggested by the above-mentioned EEOC regulations and guide posts, an antidiscrimination policy should do the following:

1. Clearly define behaviors that constitute harassment.
2. Declare that behaviors defined as harassment will not be tolerated and are prohibited in the workplace.
3. Assure that employees who make complaints of harassment or provide information related to such complaints will be protected against retaliation.
4. Establish a formal complaint procedure within the organization that mandates a prompt, thorough, and impartial investigation of every complaint.
5. Provide for prompt corrective action, including appropriate disciplinary action if unlawful harassment has occurred.
6. Assure that the employer will protect the confidentiality of harassment complaints to the extent possible.
7. Include training and education to sensitize all employees to harassment issues and the organization’s policy.

An employer’s policy should prohibit harassment by all persons in and associated with the organization,

including supervisors, coworkers, and non-employees. Moreover, the policy should encourage employees to report harassment *before* it becomes severe or pervasive, so as to prevent its escalation to a level that violates federal law.^[29]

A successful prevention program requires that employers communicate their policies to employees in as many ways as possible, starting with a training and education program. The policy should also be included in the staff handbook, posted in the organization in places where staff gather and obtain news, and communicated orally during employee orientation. Moreover, the policy should be redistributed at least annually to employees and managers and whenever it is updated.

An important ingredient of a successful sexual harassment program, often not discussed in the literature, is the HR department. The HR department plays a critical leadership role in the development, interpretation, and communication of an organization's sexual harassment policy. Typically, HR professionals are responsible for translating legal doctrine, including recent judicial rulings, into workplace policy and training employees about sexual harassment. The HR staff schedule meetings with employees to discuss sexual harassment, convey management's disapproval of sexual harassment, and inform employees of their rights and the methods for lodging complaints of harassment. HR professionals serve as important resources for managers and supervisors who need to be able to respond to situations involving sexual harassment quickly and effectively. In addition, HR departments are involved in the receipt and resolution of sexual harassment complaints.

CONCLUSION

Sexual harassment is an issue that confronts managers in all organizations and is costly to all parties involved. Although employers cannot completely eradicate the risk of sexual harassment incidents, they can significantly limit their liability by properly designing and implementing a sexual harassment policy. The role of the HR professional is critical to the successful implementation of a sexual harassment policy and complaint procedure. The HR professional needs to understand the issue's significance, the hidden and direct costs of sexual harassment, the legal and regulatory policy environment governing sexual harassment claims, as well as the employer's preventive policy and investigative procedures.

REFERENCES

1. *Meritor Savings Bank, FSB v. Vinson*. 106 S.Ct. 2399, 1986.
2. EEOC Reaches Out to High Schoolers to Combat Workplace Harassment of Teens. Press release 12-15-2004. <http://www.eeoc.gov/press/12-15-04.html> (printed October 27, 2006).
3. Lewis, J. Lawsuits Decrease, Supervisory Training Reaches Nearly 90%, According to Findings from 2005 Workplace Survey March 13, 2006. <http://www.jacksonlewis.com/legalupdates/article.cfm?aid=911> (printed October 30, 2006).
4. U.S. Equal Employment Opportunity Commission. Sexual Harassment Charges EEOC and FEPAs Combined: FY 1992–FY 2005 (2006). <http://www.eeoc.gov/stats/harass.html> (printed 10/27/06).
5. Hr Focus One-Fifth of Women are Harassed Sexually 79 (2002) p. 2.
6. Michael, K. Sexual harassment prevention training is law on Jan. 1 October 23 2005. The San Diego Union-Tribune, retrieved from Factiva October 30, 2006.
7. U.S. General Accounting Office. *Equal Opportunity: DOD Studies on Discrimination in the Military*; GAO: Washington, DC, 1995; GAO/NSAID-95-103.
8. U.S. General Accounting Office. *Equal Opportunity: NIH's Handling of Alleged Sexual Harassment and Sex Discrimination Matters*; GAO: Washington, DC, 1995; GAO/GGD-95-195.
9. Reese, L.A.; Lindenberg, K.E. 'Victimhood' and the implementation of sexual harassment policy. *Rev. Public Pers. Adm.* 1997, Winter, 37–57.
10. Lee, R.D., Jr.; Greenlaw, P.S. The legal evolution of sexual harassment. *Public Adm. Rev.* 1995, 55, 357–364.
11. Lee, R.D., Jr.; Greenlaw, P.S. A legal perspective on sexual harassment. In *Public Personnel Management: Current Concerns, Future Challenges*; Ban, C., Riccucci, N.M., Eds.; Longman Press: New York, 1997; 133 pp.
12. *Williams v. Saxbe*. 413 F. Supp. 654 (D.D.c. 1976), 1976.
13. *Barnes v. Costle*. 561 F.2d 983 (D.C. Cir. 1977), 1977.
14. *Garber v. Saxon Business Products, Inc.* 552 F. 2d 1032 (4th Cir. 1977), 1977.
15. *Tomkins v. Public Service Electric and Gas Company*. 568 F. 2d 1044 (3rd Cir. 1977), 1977.
16. *Bundy v. Jackson*. 641 F.2d. 934 (D.C. Cir. 1981), 1981.
17. *Harris v. Forklift Systems, Inc.* 114 S.Ct. 367, 1993.
18. U.S. Equal Employment Opportunity Commission. Enforcement Guidance on *Harris v. Forklift Sys., Inc.* (1994). <http://www.eeoc.gov/docs/harris.html> (printed 9/21/00).
19. *Clark Country School District v. Shirley A. Breeden*. 532 S.Ct. 00-866, 1991.
20. *Oncale v. Sundowner Offshore Services, Inc.* 118 S.Ct. 998, 1998.

21. Carleton, F.; Carleton, J. *Oncale and Its Aftermath: The Damage Done?* Paper presented at the Midwest Political Science Association Annual Meeting, 20000427.
22. Bland, T.; Stalcup, S.S. Managing harassment. *Hum. Res. Manage.* **2001**, *40*, 51–61.
23. *Faragher v. City of Boca Raton*. 118 S. Ct. 2275, 1998.
24. *Burlington Industries v. Ellerth*. 118 S. Ct. 2257, 1998.
25. *Pennsylvania State Police v. Suders*. 124 S. Ct. 2355, 2004.
26. *Gebser v. Lago Vista Independent School District*. 118 S.Ct. 1989, 1998.
27. *Davis v. Monroe*. 119 S. Ct., 1999.
28. U.S. Equal Employment Opportunity Commission. Policy Guidance on Current Issues of Sexual Harassment (3-19-1990). <http://www.eeoc.gov/policy/docs/currentissues.html> (printed 10/30/06).
29. U.S. Equal Employment Opportunity Commission. *Federal Sector Report on EEO Complaint Processing and Appeals: Fiscal Year 1999*; GPO: Washington, DC, 1999; 9 pp.
30. U.S. Equal Employment Opportunity Commission. Facts about Sexual Harassment (2002). <http://www.eeoc.gov/facts/fs-sex.html> (printed 10/30/06).

Singapore

Jon S.T. Quah

Department of Political Science, National University of Singapore, Singapore

INTRODUCTION

Singapore attained self-government in June 1959, when the People's Action Party (PAP) government assumed office after winning the May 1959 general election and after nearly 140 years of British colonial rule. The PAP government has governed Singapore for 45 years as it was reelected for the 10th time in the November 2001 general election. This article contends that the nature of public administration in Singapore is shaped by its British colonial heritage, its policy context, and the policies introduced by the PAP government.

PUBLIC ADMINISTRATION UNDER THE BRITISH (1819–1959)

In 1947, the Trusted Commission, which was appointed to review the salaries and conditions of service of public services in the Malayan Union and Singapore, recommended that a Public Service Commission (PSC) be formed as the adapted version of the Civil Service Commission in Britain; and that the Singapore Civil Service (SCS), following the example of the British Civil Service, be reorganized and divided into four divisions according to the duties and salaries of its members.

There are four major features of public administration in Singapore during the colonial period. First, the SCS did not play an important role in national development during the colonial period as it was preoccupied with enforcing colonial rule in Singapore and with economic exploitation of natural resources from the region for the benefit of the home government in Britain. In other words, the SCS focused on performing the traditional “housekeeping” functions of maintaining law and order, building public works, and collecting taxes.

Second, the SCS neglected administrative reform as it was responsible for introducing only two reforms: the fourfold division of the SCS in 1947, and the creation of the PSC in 1951. As the PSC was formed to keep politics out of the SCS by rejecting the spoils system and to speed up the localization of civil servants, its establishment led to the introduction of meritocracy in the SCS.

Third, the SCS was afflicted by the problem of corruption during the colonial period. Corruption was

made illegal in 1871, when the Penal Code of the Straits Settlements was enacted. However, the first anticorruption law was only passed 66 years later in December 1937, when the Prevention of Corruption Ordinance came into force.

Fourth, the civil servants during the British colonial period were criticized for having “a colonial mentality and were insensitive to the needs of the population.” The PAP leaders had serious misgivings about the SCS as its localization had not contributed to a national bureaucracy. Accordingly, the PAP leaders were compelled to introduce comprehensive reforms to transform the SCS after they assumed office in June 1959.

THE POLICY CONTEXT

A brief description of the policy context in Singapore is necessary as it is an important factor influencing the nature of public administration and the performance of the SCS. More specifically, Singapore's smallness, multiracial population, and level of economic development, and the PAP government's predominance will be discussed.

In terms of size, Singapore is a city-state with a total land area of 682.7 km², or the same size as Lake Biwa in Japan. Although Singapore's smallness has enhanced policy implementation as the SCS does not encounter the logistical and communications problems found in larger countries, its small size also constitutes a constraint as there is no large rural hinterland for the cultivation of crops or the mining of minerals. Indeed, as Singapore has no natural resources, it has to rely instead on its strategic location, deep harbor, and people.

In June 2002, Singapore had a population of 4,163,700 (including foreigners who have resided for at least a year) and a population density of 6075 persons/km². The resident population of 3,378,300 persons is heterogeneous in three aspects. First, it is multiracial as the Chinese constitute 76.5%, Malays make up 13.8%, Indians constitute 8.1%, and other races make up 1.6%. Second, although the resident population speaks many languages and dialects, there are four official languages: English (the language of administration), Mandarin, Malay, and Tamil. Finally, in terms of religion, 54% of Singaporeans are

Buddhists and Taoists, 15% are Muslims, 13% are Christians, 4% are Hindus, and 14% have no religion.^[1]

The resident population's diversity in race, language, and religion imposes two obligations on those governing the city-state. The first obligation is the necessity for any incumbent government to formulate and implement policies that contribute to nation building and racial harmony. Second, the political leaders must minimize discrimination of any kind by ensuring that all public and private organizations are fair and impartial in the treatment of members of the public, regardless of their race, language, or religion. The rights of the minorities and religious groups are also protected by the Presidential Council for Minority Rights and the Presidential Council for Religious Harmony, respectively.

In 1960, Singapore was a poor country with a per capita gross domestic product (GDP) of \$443, a high unemployment rate, a serious housing shortage, and rampant corruption. However, the PAP government's success in promoting economic development, solving the housing problem, and curbing corruption has resulted in: a per capita GDP of \$20,850 in 2001;^[2] 85% of the population living in public housing today; and Singapore being the least corrupt Asian country and the fifth least corrupt country among the 133 countries included in *Transparency International's* 2003 Corruption Perceptions Index.

The PAP government's predominance in Singapore can be attributed to four factors. First, during 1965–1970, it succeeded in garnering the support of the people to deal with the challenges of the withdrawal of the British military forces and the building of Singapore's own armed forces with the introduction of compulsory military service in 1967. Second, the PAP government acquired legitimacy among the population because of its effective response to communist and communal threats. Third, the PAP government's predominance in Singapore politics can be explained in terms of its ability to deliver goods and services to the population during its 45 years in power. Finally, there is no creditable alternative to the PAP government as the opposition political parties are weak organizationally, poorly funded, and unable to recruit professionals as members.

PUBLIC ADMINISTRATION UNDER THE PAP GOVERNMENT (1959–2004)

Public administration in Singapore during the 45 years of the PAP government's rule is characterized by these six features: meritocracy, minimization of corruption, comprehensive administrative reforms, competitive salaries for political leaders and senior civil servants, effective policy implementation, and policy diffusion.

Meritocracy

The PAP government retained the meritocratic system introduced by the British colonial government in 1951 with the creation of the PSC. However, Prime Minister Lee Kuan Yew expressed his disdain for the former British reliance on seniority and favored instead an emphasis on efficiency as the basis for promotion. Consequently, competent civil servants were promoted to more responsible positions regardless of their seniority. This focus on efficiency rather than seniority is responsible for the relative youthfulness of many of the permanent secretaries today. Lee's emphasis on meritocracy and the need to attract the "best and the brightest" to join the SCS can be attributed to his experience as the legal adviser to several trade unions in Singapore in the 1950s before his entry into politics. He won all his legal cases against the British colonial government as its lawyers were incompetent and poorly paid.

Accordingly, Lee reinforced the PSC's role in maintaining meritocracy by controlling the quality of personnel entering the SCS. The "best and the brightest" Singaporeans are recruited by the PSC for the SCS by awarding scholarships to the best students of every cohort to obtain university education in Singapore or abroad. In his 1996 study of the Economic Development Board (EDB), entitled *Strategic Pragmatism*, Schein^[3] astutely observed that "having 'the best and brightest' in government is probably one of Singapore's major strengths in that they are potentially the most able to invent what the country needs to survive and grow and to overcome the kinds of biases and blind spots."

Minimization of Corruption

As corruption was a serious problem during the British colonial period, especially after the Japanese Occupation (1942–1945), the newly elected PAP government initiated its anticorruption strategy in 1960 with the enactment of the Prevention of Corruption Act (POCA), which increased the penalty for corruption and empowered the Director of the Corrupt Practices Investigation Bureau (CPIB) and the officers to investigate persons accused of corrupt offences. In 1989, the POCA was amended to enhance the penalty for corruption from 3 to 5 years imprisonment, and the fine was increased 10-fold from S\$10,000 (\$6000) to S\$100,000 (\$60,000).

The PAP government succeeded in minimizing corruption because it did not rely on the British colonial method of using the Anticorruption Branch (ACB) in the Singapore Police Force (SPF) to combat corruption. The ACB was ineffective in curbing corruption

as it was inadequately staffed and funded, and had difficulty in dealing with police corruption. The theft of S\$400,000 worth of opium by some detectives in the SPF in October 1951 forced the British colonial government to transfer the task of corruption control from the SPF to the CPIB in October 1952. However, the CPIB was ineffective during the next 8 years as its powers, personnel, and funds were limited.^[4]

The PAP leaders, especially Lee Kuan Yew, were committed to eliminating corruption in Singapore as they relied on the CPIB to impartially implement the POCA against corrupt persons in both the public and private sectors. Moreover, the POCA is constantly reviewed and amendments are introduced whenever necessary to remove any loopholes. In other words, the key to Singapore's success in combating corruption is the impartial implementation of the POCA by the CPIB and the public's perception that corruption in Singapore is no longer a "low-risk, high-reward" activity but a "high-risk, low-reward" activity.^[4]

Comprehensive Administrative Reforms

In contrast to the British colonial government, which neglected administrative reform, the PAP leaders initiated comprehensive administrative reforms after assuming power in June 1959 because of the favorable timing, their commitment, and the lower degree of risk in introducing reforms when compared with the maintenance of the status quo.

The timing was favorable for the PAP government to introduce major reforms as its assumption of power marked the end of nearly 140 years of British colonial rule. The PAP leaders emphasized both institutional and attitudinal reforms as they reorganized the SCS, established statutory boards, and changed the "colonial mentality" of the civil servants and their insensitivity to the population's needs. They were also committed to reform as manifested in their critical speeches in the Legislative Assembly on different aspects of the administration of the colony. The most important reason for the introduction of comprehensive administrative reforms was the PAP leaders' perception that the risk involved in not implementing the reforms was greater than the risk accompanying the implementation of the reforms.

The comprehensive reform of the public bureaucracy in 1959 involved these aspects: the formation of the Ministry of Culture and the Ministry of National Development; the creation of such statutory boards as the Housing and Development Board (HDB) and the EDB; the reduction of salaries of senior civil servants by discontinuing their variable allowances; and the establishment of the Political Study Center to change the "colonial mentality" of the civil servants.

In short, the PAP government initiated a comprehensive reform of the public bureaucracy in 1959 because it needed the support of the civil servants to implement its programs.

During its 45 years in power, the PAP government introduced these administrative reforms: the 1959 reforms; the POCA in 1960; the 1978 and 1989 budgetary reforms; the establishment of the Public Service Division as the second central personnel agency in 1983; the formation of the Education Service Commission and Police and Civil Defense Services Commissions in 1990; the establishment of the Service Improvement Unit in April 1991; the creation of a system of 31 personnel boards in 1995; the revision of salaries from 1972 to 1993; the benchmarking of the salaries of political leaders and senior civil servants to several private sector professions in October 1994; and the launching of Public Service for the 21st Century (PS21) in May 1995.

Competitive Salaries

As Singapore's economy grew in the 1970s, the higher salaries paid by the private sector led to a brain drain from the SCS, and civil service pay was increased to curb the loss of talent. In 1972, the National Wages Council recommended the payment of an additional month's salary to minimize the gap between salaries in the public and private sectors in Singapore. The salaries of the political leaders and civil servants were periodically revised in 1973, 1979, 1982, 1989, and 1993 to reduce the salary gap in the public and private sectors.^[5]

On October 21, 1994, a white paper on *Competitive Salaries for Competent and Honest Government* was presented to the Parliament to justify the pegging of the salaries of ministers and senior civil servants to the average salaries of the top four earners in six private sector professions. The white paper recommended the introduction of formal salary benchmarks for ministers and senior bureaucrats, additional grades for political appointments, and annual salary reviews for the SCS.

The adoption of the long-term formula proposed in the white paper has two advantages: it removed the need to justify "from scratch" every revision of the salaries of ministers and senior civil servants; and it ensured the building of an efficient SCS and a competent and honest political leadership, which are responsible for Singapore's prosperity and success. In short, the white paper institutionalized the PAP government's practice of "matching public pay to the private sector, dollar for dollar" as it enabled the government to revise automatically public sector salaries in response to increases or decreases in private sector salaries.^[6]

Thus, by periodically revising civil service salaries to keep pace with rising wages in the private sector from 1972 to 1993, and by benchmarking these salaries with some private sector professions from October 1994 onward, the PAP government has enabled the SCS to retain its talented personnel and also to maintain its quality service.

Effective Policy Implementation

The SCS and statutory boards in Singapore are effective in policy implementation because, once a policy is formulated, the government ensures its successful implementation by providing the required manpower, legislation, financial resources, and equipment to the implementing agencies. Apart from political will, the emphasis on meritocracy and competitive salaries ensures that the public bureaucracy is staffed with qualified and competent personnel. The CPIB's success in minimizing corruption has removed a serious obstacle to policy implementation as scarce resources are not wasted on bribes and delays are avoided. Finally, the small size of Singapore is conducive to effective policy implementation as the public bureaucracy does not face the same logistical and communications problems encountered in larger countries. A good example is the HDB's effective implementation of the public housing program, which resulted in increase of the proportion of Singaporeans living in public housing from 9% in 1960 to 85% today.

Policy Diffusion

Finally, the PAP government has relied on policy diffusion, or the borrowing of policy ideas and solutions from other countries and adapting these ideas and solutions to suit the context of Singapore. As it is unnecessary and expensive to "reinvent the wheel," the PAP leaders and senior civil servants consider what has been done elsewhere to identify appropriate solutions for resolving policy problems in Singapore. Those solutions selected will usually be adapted and modified to suit the local context. Schein^[3] has attributed the EDB's commitment to learning and innovation "to Lee Kuan Yew's and Dr. Goh's willingness to learn from other countries and from various non-Singaporean advisers, and is most clearly demonstrated in the continuous changing and refining of social policy."

Soon after the attainment of its independence in August 1965, Singapore looked at Israel and Switzerland as role models to provide inspiration for devising relevant public policies for defense and other areas.

Later, West Germany was added to the list for technical education, followed by The Netherlands (Schiphol Airport was the model for Changi International Airport) and Japan (for quality control circles and crime prevention). The critical lesson in these learning experiences is the adoption by Singapore of ideas that have worked elsewhere (with suitable modification to consider the local context, if necessary), as well as the rejection of failures in other countries.^[7]

In sum, policy diffusion in Singapore remains an asset if there is intelligent sifting of relevant policy ideas and solutions tested in other countries by the policy makers, who must avoid blind acceptance or wholesale transplantation of foreign innovations without modification to suit the local environment.^[8]

CONCLUSION

Public administration in Singapore is the product of its British colonial heritage, the constraints imposed by its policy context, and the various policies introduced by the PAP government during its 45 years in power. Indeed, the PAP government's success in transforming Singapore from a Third World state to a First World nation can be attributed mainly to its ability to attract the "best and the brightest" Singaporeans into the SCS. The high quality of Singapore's civil servants is manifested in the country's top ranking on the competence of public officials among 59 countries in the *Global Competitiveness Reports* of 1999 and 2000, and among 75 countries in the *Global Competitiveness Report 2001–2002*.^[9–11] In sum, Singapore's lack of natural resources has been overcome by the quality of its political leaders and civil servants, who are responsible for its rapid transformation and progress during the past 45 years.

REFERENCES

1. Foo, S.L., Ed.; *Singapore 2003*; Ministry of Information, Communications and the Arts: Singapore, 2003.
2. The economist. In *Pocket World in Figures 2004 Edition*; Profile Books: London, 2003.
3. Schein, E.H. *Strategic Pragmatism: the Culture of the Economic Development Board*; MIT Press: Cambridge, 1996; pp. 197, 221–222.
4. Quah, J.S.T. Singapore's anti-corruption strategy: is this form of governance transferable to other Asian countries? In *Corruption and Governance in Asia*; Kidd, J.B., Richter, F.-J., Eds.; Palgrave Macmillan: Basingstoke, 2003; 180–197; Chap. 10.

5. Quah, J.S.T. Paying for the 'best and brightest': rewards for high public office in Singapore. In *Reward for High Public Office: Asian and Pacific Rim States*; Hood, C., Peters, B.G., Lee, G.O.M., Eds.; Routledge: London, 2003; 145–162; Chap. 9.
6. Republic of Singapore. *Competitive Salaries for Competent and Honest Government: Benchmarks for Ministers and Senior Public Officers*; (White Paper presented to the Parliament on October 21, 1994, Command 13 of 1994).
7. Quah, J.S.T. Public Administration in Singapore: managing success in a multi-racial city-state. In *Public Administration in the NICs: Challenges and Accomplishments*; Huque, A.S., Lam, J.T.M., Lee, J.C.Y., Eds.; Macmillan Press: Basingstoke, 1996; 59–89; Chap. 3.
8. Quah, J.S.T. Singapore's model of development: Is it transferable? In *Behind East Asian Growth: The Political and Social Foundations of Prosperity*; Rowen, H.S., Ed.; Routledge: London, 1998; 105–125; Chap. 5.
9. *The Global Competitiveness Report 1999*; Oxford University Press: New York, 1999.
10. *The Global Competitiveness Report 2000*; Oxford University Press: New York, 2000.
11. *The Global Competitiveness Report 2001–2002*; Oxford University Press: New York, 2002.

Singapore: Economic Policy

Charles Conteh

Department of Political Science, McMaster University, Hamilton, Ontario, Canada

INTRODUCTION

This article seeks to throw light on the institutional properties and policy processes, through which Singapore has been able to employ corporatist models of state intervention in collaboration with business and labor in formulating and implementing policies that cater to the creation of a competitive and successful market system across the various phases of the country's economic development. The focus on institutional properties in this entry is a deliberate attempt to shift the discourse on Singapore's economic success away from the emphasis on the key role of the political leadership. While, no doubt, political leadership has certainly been important, the configurations of institutions and policy processes have not been given as much attention as they deserve in explaining Singapore's economic success.

Thus, the central focus is to understand the changing institutional and political mechanics that surround the rather complex partnership between the public and private sectors, within the framework of pragmatic economic management in Singapore, and the changing character of that partnership. Pragmatism refers to the mix of state economic planning and interventionism, characterized by a non-commitment to any particular ideological blueprint and more in tune with the creation, allocation, and management of resources, as necessitated by the expediencies of changing national development priorities and the imperatives of economic growth.

The article takes the following form: First, it presents a conceptual framework and brief theoretical review as a basis for the examination of Singapore's experience with economic policy formulation and implementation. Second, an analysis is made of key phases of economic development in Singapore over the past four decades, with the aim of putting in context the institutional and policy dynamics surrounding the government's active involvement or intervention in the country's economic development. Third, an attempt is made to throw light on the changing dynamics of state-business partnership, necessitated by the ever-changing imperatives of the domestic and international market, and, in turn, the implications that such changing partnership has for the state's present capacity to manage its interventionist model of economic development.

In light of the above, the following propositions about economic policy formulation and implementation in Singapore are examined throughout the rest of the article: First, the operational capacity of a country's administrative machinery, especially economic development agencies, in strategic partnership with private sector actors, significantly affects the long-term coherence, practical relevance, and successful implementation of economic development plans. Second, state intervention geared toward enhancing the capacity of the private sector necessitates a changing model of state-business partnership, as dictated by the imperatives of changing economic circumstances.

CONCEPTUAL FRAMEWORK AND THEORETICAL REVIEW

For the purpose of pragmatic economic management, state capacity comprises two dimensions: first, the existence of reasonably efficient administrative institutions responsive to political cues but at the same time *relatively free from—or able to resist*—the pressures and penetration of parochial and short-term special interests. Second is the *legitimacy* of the state and its administrative machinery in carrying out its interventionist activities within the market.

According to Max Weber, bureaucracy is necessary for state action. The existence of extensive, internally coherent bureaucratic machinery is an indispensable prerequisite for state intervention in pragmatic economic management.^[1] Moreover, in order for the state to engage in capitalist economic transformation, the workings of the administrative machinery must link up with the workings of the market. It is also important that key officials within the bureaucracy and public agencies share major assumptions and understanding of market variables and their interaction if the internal coherence of the state's intervention is to be maintained.

Furthermore, bureaucratic competence must coexist with some reasonable degree of policy autonomy, in order for the developmental state to translate its broad national goals into coherent and effective policy programs. In the absence of bureaucratic autonomy, public-private cooperation easily degenerates into situations in which state goals become directly reducible

to private interests.^[2] In this light, autonomy makes it possible for the state to coordinate the parochial and atomistic interests of private economic actors and alter the social and economic structure along the lines of its overall strategic development policies.

Although policy autonomy leading to internal organizational coherence and policy consistency is crucial for building and fully exploiting the technical capacity of bureaucratic agencies, nevertheless, this needs to be balanced by the imperatives of the state's external connectedness to the market environment and its main actors and interests if the intervention of the state is to remain relevant to the changing realities of the economy.^[3] This process could be referred to as market legitimacy, meaning the business community and organized labor's trust in the government's capacity-enhancing activities and programs, a recognition of the authority of agency officials and government bureaucrats to direct the trajectory of market transformation and a perception among these actors that the state's policies are consistent with their medium- to longer-term interests and overall economic development.

SINGAPORE'S ECONOMIC DEVELOPMENT IN PERSPECTIVE

Over the past four decades, Singapore has been transformed from having an economy based on staples export to one characterized by well-developed industrial and financial activities, with close integration into the global economy. The economy grew rapidly through two key stages between 1960 and mid-1990, and a third stage is presently unfolding.^[4] From staples export, the economy diversified into labor-intensive manufacturing, with the government's successful diversification of the economy toward financial and business services. As the economy matured and labor became scarce, the government further made a strategic switch to capital-intensive production around the turn of the 1980s. Singapore has enjoyed consistent exceptional growth rate, averaging over 7% GDP growth per annum, in the last 40 years, doubling its GDP about 11-fold.^[5] After the recession of the mid-1980s, the economy was set for yet another major restructuring that would direct Singapore toward its new vision of becoming a knowledge-based economy and an industrial hub in the Asia-Pacific region.

Unlike certain accounts that have merely stated that Singapore was well positioned at independence to benefit from global capitalism,^[6] it is necessary to explain how and why the state was able to redirect its society along this capitalist trajectory and take advantage of existing international economic opportunities. However, given that economic policy is a rather

broad subject, the analytical scope of this article is focused on industrial development policy, with a particular emphasis on the institutional and administrative implications for the most recent industrial restructuring over the past decade.

Seeking to understand the role of the state in the process of economic development requires an explanation of the nature of the state and its power, autonomy, and legitimacy vis-à-vis private societal actors, and how that affect the process of policy formulation and implementation. The significant question is not whether Singapore followed rational economic policy but why and how. Rather, it is about the processes that ensure that such "rational" policies are made and effectively implemented. The policies themselves are only reflections of institutionalized power relations among various actors.^[7] These institutions are, in turn, configured to deal with the imperatives of changing economic circumstances. Thus, the aim is geared toward an understanding of changing processes and power relations conditioned by wider socioeconomic variables and their impact on the selection and effectiveness of economic policies.

Since Singapore attained self-government on June 3, 1959, the government conceptualized national development in terms of its ability to respond to the material needs of its people and structurally transform their societies along the path of the newly adopted capitalist trajectory. Caught at the frontlines of communist-capitalist confrontation at the height of the cold war, Singapore's ruling People's Action Party (PAP) envisaged the country's socioeconomic destiny through a capitalist trajectory, thereby categorically rejecting communism in principle and perceiving economic development not merely as a goal but also as a question of national survival and identity. The domestic political and institutional processes, through which the Singapore government was able to delegitimize and exclude non-capitalist interests and directed the economy toward partnership with international capital (mostly multinational corporations or MNCs), will set the context for understanding the country's specific model of economic policy management. Moreover, it sets the context in explaining why the Singapore government chose to partner with MNCs as the key model for the country's economic development. It also put in perspective how the state was able to maintain the "acquiescence" of the local private sector (entrepreneur and labor alike), in their own marginalization, as significant participants in economic development.

The PAP was able to use strategic international concerns over the prevalent communist menace in the early years of self-government, to assert and consolidate its hold on power. Linking up with moderate labor unions within an alliance that seeks to strike a middle course (between the extremes of leftist

radicalism and rightwing conservatism), the PAP perceived that it had a sufficiently legitimate basis to embark on a systematic “liquidation” of any serious political opposition to its hold on power. Under the same “cloud” of the communist menace (both imagined and real), and armed with the overwhelming legitimacy provided by its alliance with moderate labor unions, the government was able to exclude and delegitimize more radical labor unions that were not inclined to follow the PAP’s path toward a pro-Western capitalist trajectory. It was the extension of this same logic of pro-Western capitalist trajectory that easily justified the PAP’s exclusion of local entrepreneurs who were showing sympathies for anti-Western sentiments.

Thus, the PAP created an artificial political stability and “social quiescence” as crucial foundations for building a socioeconomic infrastructure, characterized by instrumental rationality and elitist policy making in long-term strategic partnership with foreign investors.^[8] But the real “engineering” of the political context that shaped the atmosphere of economic policy formulation and implementation in Singapore was consolidated when certain institutions were created to give more concrete expression to the “middle road” of what could be referred to as strategic pragmatism of capitalist development. The Economic Development Board (EDB), established in 1961 to plan and promote economic development in the manufacturing sector, became the central agency responsible for market coordination and provision of incentive structures.^[9] Along with other agencies like the Jurong Town Corporation and the Urban Redevelopment Authority, the EDB maintained a long-term vision of market-capacity enhancement, by seeking to attract and build close partnerships with large MNCs.

A significant element of Singapore’s economic development strategy, therefore, is the strategic partnership between competent and autonomous agencies, led by the EDB, the multiple linkages between these agencies, and their close embeddedness within the market through consolidated networks with private economic actors.^[10] EDB’s partnership with MNCs, with its implicit exclusion of local enterprises, formed the overarching industrial development model that guided Singapore through the various phases of its economic development—at least until a decade ago, when serious economic and institutional reconfigurations started taking shape.

Another crucial institutional expression of Singapore’s economic pragmatism was the country’s corporatist “tripartite structure” of industrial relations management and market governance through the framework of the National Wage Council, consisting of state representatives, entrepreneurs, and labor unions.^[11] Over 90% of organized, unionized labor in Singapore has direct institutional affiliation with the

PAP-led National Trades Union Congress (NTUC). The head of the NTUC is a government minister. The most characteristic feature of the tripartite structure is that it provides a framework of labor management, through which dialogue between labor and capital is fostered, bargains are struck, and systemic long-term synergies are cultivated for economic development.

STATE CAPACITY IN INTERVENTIONIST ECONOMIC MANAGEMENT

One of the observations that have been made about Singapore is that the country inherited a strong administration and effective government from the British who governed Singapore until the country’s independence. This is not accurate because, in fact, the PAP government inherited an administration with significant structural weaknesses when it assumed power in June 1959. Fundamentally, the civil service did not have an orientation toward pursuing national development. Moreover, corruption was endemic in the Singapore public sector. The British generally neglected administrative reform. The PAP had to embark on a systematic and comprehensive reform and reorientation of Singapore’s Civil Service into a dynamic machinery of development administration.

However, a more notable feature of public administration in Singapore is the depoliticization of politics within the broader framework of policy formulation as a whole. Added to this was the strong belief of the ruling PAP in technocratic economic management, free from the uncertainties of interest politics.^[12] Through its subtle exclusion of local enterprise within the government’s EDB–MNC model of economic development, a potential political force was delegitimized from having any serious influence on the direction of economic development. Moreover, by effectively co-opting organized labor through the government’s NWC framework, a considerable constellation of labor interests as a political factor was “internalized” within the government’s calculus of opaque policy negotiation and “consensus” governance. This provides the government the dual advantage of “legitimizing” its active management of the economy and also reinforcing its operational autonomy from particularist interest articulation.

The operational autonomy of economic development agencies—and the civil service as a whole—extended not only to society at large but even to its own partners within the market. In its dealing with MNCs, for instance, the EDB maintains a close relational fusion with its MNCs partners but manage to keep a clear operational autonomy from their particularist interests. The success of this strategy was seen in

the fact that powerful economic interests could not deviate the state from its strategic thrusts along a certain direction. While the state remains sensitive to the needs and demands, the former was able to restructure the economy at will across critical junctures of its industrialization. It is no surprise then that Singapore has been rated repeatedly as having one of the most efficient and competitive public management systems in the world.

CHANGING DYNAMICS IN STATE-BUSINESS PARTNERSHIP

The focus of this section is to understand the policy and institutional and administrative implications involved in the recent strategic thrust over the last 10 yr toward a knowledge-based economy, with a focus on making Singapore an industrial hub in the Asia-Pacific region. A key part of this new thrust is an increasing realization that the old model of exclusive state partnership with MNCs will not sustain the country's long-term development within the region.^a As regional competition increases and international capital becomes more fluid, it is important that Singapore revisits its state-business partnership model.^[13] A result if this revisit has been the increased importance of more systematic inclusion of local enterprises into the state-business partnership.

This is not to say that local enterprise development is a new phenomenon, but that it is taking on more prominent place in the country's economic management strategy. Mechanisms such as market network linkages are being examined to cultivate closer cooperation between MNCs and small- and medium-size enterprises (SMEs). There are numerous schemes to nurture local enterprises into Singapore-made MNCs, with the capacity to engage more aggressively in the regional and even global economy.^[14] Some of the most popular SME schemes include the following: the Local Enterprise Finance Scheme; the Local Enterprise Technical Assistance Scheme; the Skills Development Fund; Corporate Advisers Programme; Micro Loan Programme; Technology for Enterprise Capability Upgrading; and Loan Insurance Schemes, among others. In general, Singapore has over 60 schemes to support SMEs.

There have also been increasing calls for the government to retreat from some of its productive activities in the market to make entrepreneurial room for more local businesses. In other words, the urgency of local enterprise is taking such a policy hold that local

entrepreneurs, policy think tanks (like the Institute for Policy Studies), and academics alike are becoming bolder in their call for the withdrawal of the once-revered government-linked companies from their "crowding out" activities in the market.

The implication of these recent developments is that institutional and policy coherence and continuity as understood in Singapore could be witnessing increasing strains. Even more significant is that within the arena of policy implementation, the institutional framework for economic management is undergoing radical reconfiguration, which is witnessing a fast redirection away from Singapore's signature strategy of one-stop service through the EDB for business development to multiple stops through not only a host of agencies under the Ministry of Trade and Industry but also other government ministries whose mandate may include something related to economic development.

Just to give a picture, in addition to the EDB and JTC, other agencies now responsible for assisting coordinating business development in Singapore include the National Computer Board (NCB), Trade Development Board (TDB), IDA (revamped), SPRING, EPC, A*STAR, STB, and STIF. Without dwelling on the meanings of all the acronyms of the above agencies and their functions, it is worth illustrating the complexity of the new institutional environment.

Agencies such as the NCB and the TDB, among others, were created (or older agencies have their mandates expanded) to cover the various sectors and sub-sectors, in managing the complex needs of a highly heterogeneous and specialized market. For instance, although the TDB had been formed as early as in 1983, it was a rather passive organization, since local enterprises were not a strategic part of Singapore's industrial policy. But after 1985, as the new economy takes shape and new policies emerge, giving greater attention to local enterprise development, the TDB was revamped into a proactive agency providing support to SMEs. Another example is the creation of SPRING, a relatively new agency that has been working with SMEs to develop intersectoral collaboration among local enterprises—what the agency calls "business fusion groups."

A significant implication of this strategic switch to multiagency approach is the challenge to institutional coherence and capacity that is being created by the multiplicity of agencies running various schemes in a complex and mature economy that is bent on going more regional. Part of this could be a function of the fact that as the jurisdiction and mandate of these various agencies develop and even duplicate each other, tensions and strains around cooperation are beginning to emerge. The clear leadership that the EDB was showing under the "old-order" economy is being eroded, as other agencies (such as SPRING) see their

^aFrom an interview with a senior official at the EDB, Singapore, at the EDB office on November 28, 2005.

primary mandate taking on more policy prominence within the new economy.

Local enterprises are already complaining that there is immense confusion over which agency does what. For instance, SPRING, as maintained above, has been working with SMEs to develop business fusion groups. One problem, however, is that while SPRING assumes greater and greater responsibility for the fostering of SMEs and even sees itself as a de facto "EDB" or central "nerve" for SME, its mission statement and the ranking of its priorities give the confusing impression that SME development is not the main function of SPRING. Instead, it has this vague statement claiming that SPRING seeks to "raise productivity," so as to enhance Singapore's competitiveness and economic growth for a better quality of life for Singaporeans. Just what this means in practical, operational terms remains unclear. Moreover, SPRING's relationship with the other agencies, including EDB, remains vague.

Another implication of this multiagency approach is the gradual erosion of state autonomy from societal penetration that state agencies once enjoy. Within their multiple and often duplicating functions, and the pressures of jurisdictional turf among agencies, the tendency toward increased porosity to the parochial demands of specific private sector interests cannot be overstated.^b It becomes less unlikely for these multiple agencies to lose sight of the overarching strategic goals and get focused on their specific mandates, with attendant sympathies for, and proximity to, their immediate clientele, relative to the broader and more abstract national goals.

Furthermore, as part of the institutional shift, Singapore's chambers of commerce and industry, as well as trade and industry associations (as representatives of local enterprises), have been incorporated into the policy arena of collaborative network partnership in governing the market. The implication here is that the character of market legitimacy of the state would require some reconfiguration under the new market arrangements. First, SMEs, unlike MNCs, have a much wider representation, very heterogeneous interests, thus requiring a more complex type of state-business partnership. This complexity is already being reflected in the move from the single agency to multiagency approach to business development.

Moreover, MNCs were essentially owned and run by foreigners in Singapore. Their approach to participation in policy formulation and implementation has the measured modesty and caution of "outsiders" within the political arena. Local entrepreneurs, on the other hand, do feel a greater sense of

ownership of the policy arena. The implication here is that the serenity and instrumental rationality of policy negotiation and implementation is beginning to witness a more convoluted process of widely heterogeneous demands and largely politically enfranchised interests, with higher expectations or demands on government agencies.

CONCLUSION

To put it in a nutshell, this entry has looked at the changing institutional properties and policy processes that characterize Singapore's corporatist model of state-market partnership. Using instruments of pragmatic economic management, the Singapore state has fostered strategic partnership with MNCs over the past 40 years with very successful economic outcomes. However, this model of exclusive partnership with MNCs is under strain, as the imperatives of industrial restructuring toward a knowledge-based economy necessitate a more systematic inclusion of local enterprises into the policy and administrative arena. Therefore, the institutional configuration toward increasing multiagency management of the economy, as dictated by the new direction of industrial development, is having some bearing on the state's capacity and legitimacy to effectively spearhead the pace and trajectory of economic development.

REFERENCES

1. Rueschemeyer, D.; Evans, P. The state and economic transformation. In *Bringing Back the State*; Evans, P.B., Rueschemeyer, D., Skockpol, T., Eds.; Cambridge University Press: New York, 1985; 50 pp.
2. Onis, Z. The logic of the developmental state. *Comp. Polit.* **1991**, *24* (1), 114.
3. Brett, E.A. Participation and accountability in development management. *J. Dev. Stud.* **2003**, *40* (2), 22.
4. Institute for Policy Studies (IPS). *The IFER Report: Restructuring Singapore Economy*; Times Academic Press: Singapore, 2002; 54–70.
5. Low, L. The Economic Development Board. In *Challenge and Response: Thirty Years of the Economic Development Board*; Low, L., Ed.; Times Academic Press: Singapore, 1993; 61–79.
6. Huff, W.G. *The Economic Growth of Singapore: Trade and Development in the Twentieth Century*; Cambridge University Press: Cambridge, 1994; 7–42.
7. Rodan, G. *The Political Economy of Singapore's Industrialization: National State and International Capital*; MacMillan: Hampshire, England, 1989; 1–30.
8. Huff, W. *The Economic Growth of Singapore: Trade and Development in the Twentieth Century*; Cambridge University Press: Cambridge, 1994; 299–360.

^bFrom an interview with a staff at the Singapore Civil Service College (CSC), at the CSC office on November 20, 2005.

9. Fletcher, R.G.; Moscové, B.J. Singapore's economic development: a financial perspective. In *Global Perspectives on Economic Development: Government and Business Finance*; Bingham, R., Hill, E., Eds.; Centre for Urban Policy Research: New Brunswick, 1997; 167–199.
10. Schein, E.C. *Strategic Pragmatism: The Culture of Singapore's Economic Development Board*; MIT Press: Cambridge, MA, 1996; 170–175.
11. Kuruvilla, S.; Erickson, C.; Hwang, A. An assessment of the Singapore skills development system: does it constitute a viable model for other developing countries? *World Dev.* **2002**, *30* (8), 1470–1471.
12. Seah, D.C. The administrative state L Quo Vadis? In *Singapore: Towards a Developed Status*; Low, L., Ed.; Oxford University Press: Oxford, 1999; 251–269.
13. Yuan, L.T. Singapore in economic transition. In *Singapore: Towards a Developed Status*; Low, L., Ed.; Oxford University Press: Oxford, 1999; 67–77.
14. Hew, D. SME policies and SME linkage development in Singapore. In *Entrepreneurship and SMEs in Southeast Asia*; Hew, D., Nee, L.W., Eds.; Institute for Southeast Asian Studies: Singapore, 2004; 175–198.

Social Networks

Brian E. Adams

*Department of Political Science, San Diego State University,
San Diego, California, U.S.A.*

INTRODUCTION

Social networks are the relationships between individuals or groups of individuals. These relationships could be formed through many channels, such as organizational membership, political activity, or recreational activity. A central characteristic of social networks is the presence of individuals known as “hubs” or “connectors” who have an exceptional number of links to various individuals and groups. These connectors reduce the social distance between disparate groups, forming bridges between groups that would not otherwise be connected.

Social networks are an important feature of community life that have a significant influence on the policy-making process. Social networks facilitate citizen involvement in politics, acting both as a mobilizing force and as a resource for citizens attempting to influence policy decisions. Further, they can potentially foster attitudes and dispositions, such as trust and norms of reciprocity, that build social capital and facilitate community cooperation in addressing social problems. However, networks do not always have the characteristics needed to promote these goals, and could act to limit the capacity of citizens to address public problems.

SOCIAL NETWORK STRUCTURE

A network is comprised of nodes—the things being connected together—and links, which connect the nodes. A *social* network is a network where the nodes are individuals (or groups of individuals) and the links are interpersonal relationships between them. Note that “social” here denotes that the network is comprised of people and their interpersonal relationships, as opposed to other networks where the nodes are inanimate object and the links are not personal relationships (such as the world wide web, where the nodes are web pages and the links electronic). Thus, social networks encompass more than just networks derived from “social settings.” It includes any set of individuals connected by personal relationships, regardless

of the context in which those relationships were formed.

Nodes in social networks vary in the number of links they have, i.e., some people have a greater number of relationships than others. Most social networks have a small number of nodes with many links, and many nodes with only a few links each. For example, in a group of 25 people, you may have one or two people who know almost everybody else, while most members of the group will only know a few others.

Nodes that have an exceptionally large number of links are referred to as hubs or connectors. For example, in our group of 25 people, the connectors would be the one or two people who know almost everybody else in the group. Buchanan refers to networks characterized by a few hubs with many links as “aristocratic” because they are dominated by a few nodes.^[1] “Egalitarian” networks, on the other hand, are ones where most nodes have about the same number of links. Buchanan uses this terminology to highlight the fact that when a few nodes have a disproportionate share of links, their importance, in terms of the functioning of the network, increases. Nodes also increase in importance when they have unique links to other nodes that make connections between otherwise unconnected groups. Thus, both the node’s number and the type of connections influences their importance and role within a network.

Hubs/connectors are important for three reasons. First, because they have so many links, they are the linchpins that hold networks together. If a network is dominated by a few major hubs, removing or incapacitating those hubs could undermine the whole network. For example, shutting down O’Hare airport, a hub of the air transportation system, would wreak havoc on that network, while shutting down a smaller airport with fewer flights would not. If you were to remove from a social network the handful of people that are the most well connected, the network would likely splinter into smaller networks, as parts of the network would lose connections to other parts. Because they have so many links, hubs significantly influence the boundaries and form of a network.

Hubs are also critical aspects of networks because they are central to their functioning. In the context of a social network, they facilitate the flow of resources and information through the network. This dynamic is aptly demonstrated by Gladwell in his book *The Tipping Point*.^[2] Gladwell argues that one way in which epidemics or fads take off is that a few “exceptional people” are responsible for spreading information or ideas. They are “exceptional” in the extent of their networks and their ability to use them. In other words, they are connectors. Information does not spread evenly or randomly through social networks; rather, it flows from connectors outward. Because they have so many links to a diverse array of individuals, they create the opportunity for information to flow from one group to another. The functioning of a social network as a conduit for information flows depends largely on the existence of connectors; without them, this task would be far more difficult.

Finally, connectors are important because they make the social distance between any two people within a network shorter. The number of nodes it takes to go from one node to another within a network is referred to as the “path distance” or “degree of separation.” In most networks, the path distance between randomly selected nodes is surprisingly small. For example, a study done by Stanley Milgram in the 1960s found that two randomly selected people in the United States can be linked through their social networks by an average of six people (although subsequent research has reached slightly different conclusions). One reason path distance is so short is because connectors provide links between individuals who otherwise would have no connection to each other. Because connectors have so many links, they are able to bring together nodes that are from different social worlds and otherwise would have no connection. This creates a “small world” phenomenon. By linking together so many different people, connectors significantly reduce the social distance between any two random people, thus allowing for the flow of resources and information across disparate groups that are otherwise not connected.

Hubs can compensate for a lack of network density, which refers to number of links between nodes; the more links, the denser the network. Thus, in a group of 25 people, if each person only knows one other person, it is not a dense network, while if everybody knows almost everybody else, it is very dense. When we think of the connectivity of a network, we typically do so in terms of density: the more links between people, the more highly connected the network. But density is not the only factor contributing to connectivity. A dense network of 25 people, if all the links are to each other, will lack connections to others outside of the immediate group. A less dense network could be

more connected (in the sense of having greater capacity to communicate with others) if it has a few links to hubs that in turn connect it to other groups. It is incorrect to assume that the networks that are most densely connected are also the most capable of performing desired functions, such as the flow of information or the sharing of resources. The *pattern* of links within groups and between groups is as important, if not more important, than the raw *number* of links between individuals.

TYPES OF RELATIONSHIPS IN SOCIAL NETWORKS

The nature of the links between nodes is also an important feature of social networks. The central distinction is between “strong ties,” which are links between personal friends, family, and others who know each other well, and “weak ties,” which are links between acquaintances or people who only know each other through a third person (e.g., a “friend of a friend”).^[3] In many cases weak ties are a more valuable resource than strong ties. For example, a job seeker may find weak ties to be more useful in generating leads than strong ties, since a person’s close friends and family are less likely to have information about jobs that the person do not already know about. Weak ties are valuable because they connect individuals with others who are from different social circles and thus have access to different resources and information. Strong ties, of course, are also important in that they represent strong interpersonal relationships that can serve as a foundation for social cohesion and cooperation. But in terms of individuals’ ability to gain access to information and resources through networks, weak ties are often more valuable.

Another distinction between different types of relationships in social networks is the difference between a “bridging” and a “bonding” network. Bridging social networks link together groups and individuals that are dissimilar in important respects, which form bridges between groups that otherwise would not have any connections. Bonding social networks tie together those who share commonalities and perhaps have connections in other contexts. Both types of networks are valuable: bonding networks can create social cohesion and group solidarity, while bridging networks reduce social distance and can enhance understanding and cooperation between groups (although critics argue that bonding social networks can also promote intolerant and insular communities). Connectors are the critical facilitator for bridging networks: they form the links that can tie together disparate groups. Bonding social networks, on the other hand, are more a function

of network density. Thus, the social function that network ties perform (i.e., bridging or bonding) is derivative of the structure of the network itself, in terms of density and the role of connectors.

The basic building blocks for social networks are, generally, clusters of individuals who form circles of friends, workplace colleagues, families, or political groups. How these small clusters are tied together into larger social networks significantly influences network dynamics. This is why connectors and weak ties are so important: because they are the links that assemble smaller clusters into larger networks, they have a tremendous impact on the manner in which networks operate. With fewer connectors and a dearth of weak ties, clusters are not likely to be as connected, thwarting the flow of information and resources through the network. With more hubs, and bridges between various clusters, networks become “small worlds” where the ability to communicate becomes much easier, even between socially distant clusters.

SOCIAL NETWORKS AND THE POLICY-MAKING PROCESS

Social networks are relevant to the policy-making process in three ways: they can mobilize citizens to participate, they can be used as a resource by citizens who are attempting to influence policy decisions, and they can create social capital, which in turn fosters attitudes and dispositions that can enhance a community's capacity to address social problems.

Mobilization

Social networks are a means through which citizens can be mobilized to participate in the policy-making process. Scholars examining political mobilization have made the distinction between *intentional* mobilization, which involves people directly asking others to get involved, vs. *unintentional* mobilization, which is when people hear about political activities through social networks, and as a consequence get involved even though there was no intent by anyone to mobilize them. With intentional mobilization, there are explicit efforts to draw on networks to engage other people. In unintentional mobilization, social networks facilitate the mobilization process by acting as a conduit for information, even though there is no explicit attempt to use them in this manner.

Whether it be intentional or unintentional, individuals with extensive social networks are more likely to be mobilized and hence participate in the policy-making process. The decision to participate is significantly influenced by whether an individual is asked

to participate by someone in their social network. As Rosenstone and Hansen put it, “citizens participate in elections and government both because they go to politics and because politics comes to them.”^[4] The extent of participation is not just a function of the attitudes and dispositions of citizens, but also the density and character of social networks in a community that can be used to prompt citizens to participate. Expansive networks that link together disparate groups can facilitate citizen participation, while networks that do not could constrain the ability of political actors to mobilize others.

Social Networks as a Political Resource

Social networks are key resources for citizens who desire to influence public policy. First, as just discussed, they can be used as a recruiting tool to mobilize other citizens to support one's effort. Social networks are also an important source of information about politics for citizens, who can use their networks to learn about the activities of government officials and other citizens, as well as acquire the “inside scoop” on political events. Conversely, they also are used to disseminate information that activists wish to make known throughout the community. Perhaps most importantly, social networks are a means through which citizens with similar political goals can find each other and pool their resources to pursue common objectives. In general, social networks are a resource akin to time, money, and civic skills that assist citizens in their efforts to influence public policy.^[5]

The Creation of Social Capital

Social capital is defined by Robert Putnam as the “connections among individuals—social networks and the norms of reciprocity and trustworthiness that arise from them.”^[6] In Putnam's usage, social capital includes both the networks themselves and the consequences of being embedded in networks, such as trust of others and adherence to democratic norms. Putnam focuses on social networks built through participation in community organizations (both political and non-political) because it is through these organizations where the benefits of social capital—the development of norms of trust and reciprocity—are most likely to develop. Putnam's central thesis is that participation in community organizations is declining, leading to a decline in the stock of social capital. Critics have questioned both Putnam's argument that civic participation is declining and the contention that a decline in civic participation will lead to a decline in social capital.^[7]

Scholars studying social capital hypothesize that being part of social networks formed through civic

engagement will lead individuals to be more trusting of other citizens and social institutions generally; interaction with others in the community will lead to a more positive assessment of society and fellow citizens. Involvement in community organizations is also hypothesized to instill democratic norms in individuals. Citizens who are active in community organizations will develop norms of reciprocity, become more tolerant of others' views, and be more public spirited. Survey research has found that there is a correlation between civic engagement and these hypothesized benefits. But the causation can work both ways: civic engagement could instill norms of trust and reciprocity, or individuals who already have these dispositions could be more willing to engage in civic activity.^[8] Thus, whether civic engagement really has the salutary benefits described by Putnam and others is still an open question.

CONCLUSION

Social networks are an essential part of political activity, which is fundamentally a group exercise. The nature and density of social networks that exist in a community can potentially have a profound impact on the policy-making process by influencing who participates, the effectiveness of citizen participation, and the dispositions of citizens toward the political process. While research has not conclusively demonstrated that civic engagement has all the positive benefits hypothesized, there is strong evidence that a community's social networks have important consequences for policy making and governance.

REFERENCES

1. Buchanan, M. *Nexus: Small Worlds and the Ground-breaking Science of Networks*; W.W. Norton: New York, NY, 2002; 18–20.
2. Gladwell, M. *The Tipping Point*; Little, Brown and Company: Boston, MA, 2000; 19–22.
3. Granovetter, M.S. The strength of weak ties. *Am. J. Sociology* **1973**, *78* (6), 1360–1380.
4. Rosenstone, S.J.; Hansen, J.M. *Mobilization, Participation, and Democracy in America*; MacMillan: New York, NY, 1993; 6 pp.
5. Adams, B.E. *Citizen Lobbyists: Local Efforts to Influence Public Policy*; Temple University Press: Philadelphia, PA, 2007.
6. Putnam, R.D. *Bowling Alone: The Collapse and Revival of American Community*; Simon and Shuster: New York, NY, 2000; 19 pp.
7. Theiss-Morse, E.; Hibbing, J.R. Citizenship and civic engagement. *Ann. Rev. Polit. Sci.* **2005**, *8*, 227–249.
8. Brehm, J.; Rahn, W. Individual-level evidence for the causes and consequences of social capital. *Am. J. Polit. Sci.* **1997**, *41* (3), 999–1023.

BIBLIOGRAPHY

1. Barabasi, A. *Linked*; Plume: New York, NY, 2002.
2. Wuthnow, R. *Loose Connections: Joining Together in America's Fragmented Communities*; Harvard University Press: Cambridge, MA, 1998.
3. Lake, Ronald La Due; Robert Huckfeldt. Social capital, social networks, and political participation. *Political Psychology* **1998**, *19* (3), 567–584.
4. Watts, Duncan J. *Six Degrees: The Science of a Connected Age*; W.W. Norton: New York, 2003.

Special District Governments

G. Ross Stephens

Department of Political Science, University of Missouri-Kansas City,
Kansas City, Missouri, U.S.A.

INTRODUCTION

Special districts governments are the most numerous and the fastest growing type of local government in the United States. In 1942, there were 8299 districts, having declined from 14,572 in 1932 over the course of the great depression. Since 1942, they have increased steadily to 35,052 in 2002 (GC02-1-1), and they are expected to be about 35,500 by the year 2007 if recent trends continue (see Fig. 1).^a For the present, 1997–2002, they constitute 40% of all local governments. Moreover, not all public agencies called “districts” or “authorities” are special district governments. Many are semi-independent subordinate agencies of other local or state governments.^b

DEFINITION

The Census Bureau defines special districts as, independent, special-purpose governmental units (other than school district governments)^c that exist as separate entities with substantial administrative and fiscal independence from general-purpose governments. Special district governments usually provide specific services that are not being provided by existing general-purpose governments. Most perform a single function, but, in some instances, their enabling legislation allows them to provide two or more, usually related, types of services [GC97(4)-2 and GC02(4)-2].

To be classified as a special district, a local unit must have three characteristics: existence as an organized entity, governmental character, and substantial autonomy. These are not easy standards to apply, because they are subject to varying practices in different states:

1. *Existence as an organized entity*: This means the presence of some form of organization, corporate powers such as perpetual succession, and the right to sue and be sued, to make contracts, to acquire and dispose of property, and so on. Districts are designated in law as “public corporations” or “bodies corporate and politic.” To be counted as a separate government, they must receive revenue, conduct activities, and have corporate officers.
2. *Governmental character*: This characteristic is indicated where the officers are elected or appointed by other state or local government officials or where other government officials serve as district officials in an ex officio capacity, and, also, where there is a degree of public responsibility indicated by the requirement for public reporting or election of officials. Governmental character is attributed where the district has the power to levy taxes or service charges,

^aMost of the data used are taken from the 1957 through the 2002 *Census of Governments* or from data provided to the author from the Governments Division of the U.S. Bureau of the Census, Washington, DC. These citations, for example, GC02(1)-1, stand for the 2002 *Census of Governments*, Volume 1, “Government Organization,” Number 1, Government Organization.

^bThere are literally tens of thousands of subordinate, semi-independent state, and local agencies that are sometimes called districts or authorities or a variety of other names that appear to be similar to special districts but do not meet all of the Census criteria for listing them as separate governments. As of 2002, the number of types of subordinate agencies ranges from 20 in Hawaii to 266 in Louisiana. One-third is under the aegis of the state, with the remainder being local. Under the state government, there are from eight in Kansas and North Dakota to 86 in Louisiana. Most of those under the aegis of state government are local or regional in character. The number of types of local subordinate agencies range from four in Hawaii to 180 in Louisiana. They are often used to take a service and its finances off-budget. Of the types listed, some have only a single unit, whereas others may have dozens or hundreds of units. A total of approximately 3458 types are listed by the Census [GC02(1)-1] of which 1149 are state, and the other 2321 are local, i.e., county, municipal, township, joint, or tribal. The average state has 23 types at the state level and 46 local. The District of Columbia has 12. This tabulation excludes similar local entities for elections and the administration of justice. The total number of these entities probably exceeds the total number of special district governments in the United States. Since 1962, the number of types of these subordinate agencies has been growing more than twice as fast as those classified as independent special districts. The existence of these agencies makes the structure of local government even more atomistic and amorphous.

^cSpecial districts and authorities must be distinguished from school districts, the other usually single-function type of local government. School districts provide K through 12 and sometimes community college public education, though at times they are also responsible for library services.

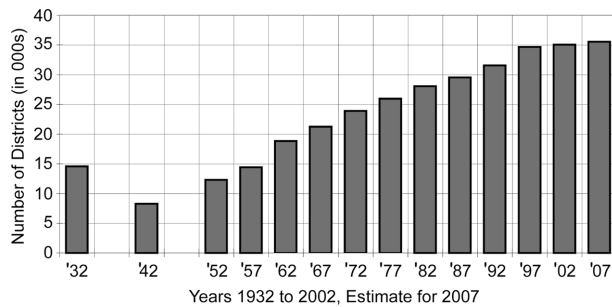


Fig. 1 Growth in the number of special district governments, 1932–2002, with estimate for 2007. (Data for 1932 through 1952 taken from Ref.^[1]; for 1957 through 1997 and 2002, GC02-(1)-1, from Ref.^[2].)

issue debt paying interest that is free from federal taxation, and/or perform a function that is generally considered a governmental activity.

3. *Substantial autonomy*: Autonomy means that the unit is substantially free from the supervision of other local units and the state, is free to set its own budget without strict review by other local governments and the state, and can largely determine the level taxes and service charges it levies. Additionally, administrative independence is considered if the governing body is elected, or represents two or more other local governments, or, if the governing body is appointed, it performs functions that are different from those of the appointing governments and are not subject to specifications by the creating governments. It should be noted that some states make a technical distinction between special districts and authorities, with special districts being capable of levying taxes and authorities operating on service charges.

Use

Special districts exist in all 50 states and the District of Columbia, with the 2002 numbers ranging from one in the District of Columbia, 14 in Alaska, and 15 in Hawaii to 2245 in Texas, 2830 in California, and 3145 in Illinois. The average state has about 700 special districts and authorities. Most, 91%, provide only a single function, while the remainder may provide two or more services. They perform virtually every service associated with government and some that are more commonly considered the role of private enterprise.

Though the numbers can vary from 1–3 to over 20, on average, they have a five-member governing board that may be elected, appointed, ex officio, or some combination thereof. Fifty-three percent of the governing

board membership is elected, with most of the remainder appointed, usually by other local governments or the state, though some are officials of other local governments or the state and serve ex officio. As of 2002, an estimated 180,000 individuals serve on the boards of special district governments.^d At the same time, these are governments that are largely *hidden* from public view. In other words, they largely operate below the level of public, media, and scholarly attention. Axelrod called them “shadow governments.”^[3] District elections rarely draw much voter attention, and the turnout is usually minuscule even by the low levels of voter participation that are normally associated with local elections.

Why, then, are there so many districts, and why do they provide such a variety of activities? First, they are usually created to provide a service or tap a revenue source not available to other local governments. For the 1987 and 1992 *Census of Governments* years, 93% of the variation in the use of special district governments is explained by the degree of state centralization and the level of activity on the part of other local governments. Stated differently, they are more likely to be used extensively where the state-local system is more decentralized and/or there is a lower level of services provided by other local governments. In other words, they are used to fill in the service gaps left by other local units and the state government.^[4] In Wisconsin, for example, townships can create coterminous township sanitary districts in order to provide this service, issue bonds, levy service charges, assessments, and a property tax [GC97(1)-1, A-273]. Foster blames private developers for the creation of many special districts, in order to obtain and control services for their projects.^[5]

Writing in 1957, Bollens noted that many special districts have been created, because existing governmental units have been found to be unsuited to present-day needs because of small area, lack of authorization for specific services, weak financial position, faulty administrative structure, antagonism, unresponsiveness, or resistance on the part of the officials of existing governments.^[6] Federal promotion, expediency, and clientele promotion (the desire to “free” a service from politics) are also factors (247–252).

Another aspect of their use is that they are often flexible in terms of the territory they serve—they can be created to serve larger or smaller geographic areas than those of existing local governments. Some districts provide one or more services to a metropolitan

^dSome of this information is based on the 1992 Census report on “popularly elected officials” [GC92(1)-2] and extrapolated to 1997 and 2002, given the increase in the number of special districts. To date, the Governments Division has not issued this report for 2002.

area. Further, they can often provide a service or services that are not authorized for existing local governments such as townships or counties.

A further reason for the creation of so many special districts is the policies of federal and state governments and, in some cases, the availability of grants-in-aid or shared revenue to fund their activities. Road districts have been created in some areas to receive a share of state revenue from taxes on motor vehicles and motor fuel. Fire protection districts in some states receive a share of state taxes on fire insurance. The U.S. Department of Agriculture stimulated the establishment of nearly 7000 special districts, providing drainage and flood control, soil and water conservation, and other natural resource activities. Federal policies and grants supporting planning, housing, and community development encouraged the establishment of nearly 3400 housing and community development districts. Other federal and some state policies have directly or indirectly encouraged the creation of other districts providing sewerage, water supply, solid waste management, as well as services related to transportation and transit.

Finances

Special districts are often used to fund new public facilities. Some districts, such as those for the construction of public school buildings in Indiana and Pennsylvania, are established specifically for the purpose of funding new facilities. In FY 2002, capital outlay accounted for 21% of district expenditures, which is almost twice the proportion of that for other local governments that average 11%. Even though special districts have less than 11% of local revenues and expenditures, they account for 21% of local government indebtedness, with over 89% of the 2002 debt of \$218 billion being the responsibility of the 1757 largest districts [GC02(4)-2].

Of the \$122.1 billion spent in FY 2002, no single function stands out as dominant, given the wide range of activities performed by special district governments. Leaving out utilities, the largest single function performed is hospitals at \$17.8 billion. Next is housing and community development, at \$12.4 billion, over four-fifths of which is paid for by federal grants. With expenditures of \$48.2 billion, utilities comprise two-fifths of all district expenditures. Nearly \$19.4 billion is for transit, \$17.6 for electric power, \$11.1 for water supply, and almost \$1.1 billion for gas utilities (see Fig. 2).

In FY 2002, total revenue amounted to \$123.7 billion. Well over half of all revenue, \$67.5 billion or 54.6%, is in the form of charges for services rendered such as hospitals, sewerage, and utilities. Taxes

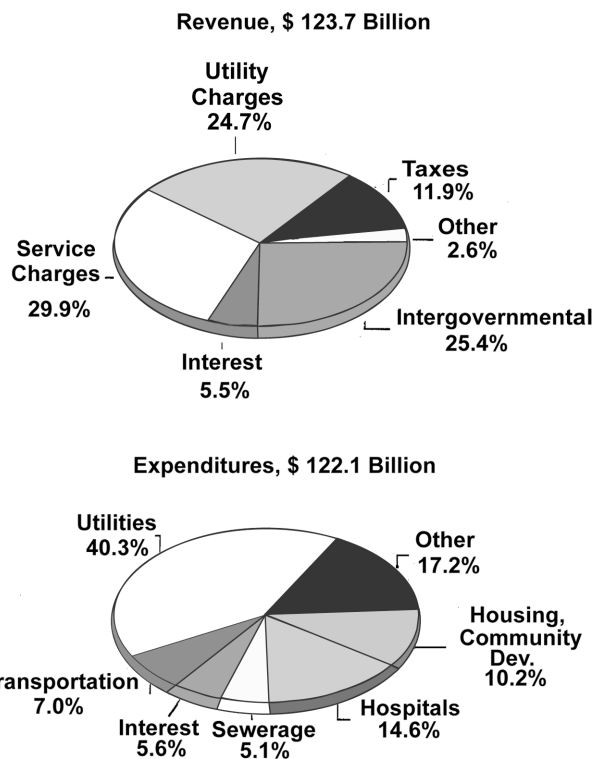


Fig. 2 Special district revenue by source and expenditures by function, 2002. (From Ref.^[8].)

brought in only 11.9% of total revenue, mostly from property and sales taxes. Intergovernmental revenue comprised one-fourth of district income at \$31.4 billion. This is 13.1% from the federal government, 6.9% from the state, and 5.4% from other local governments. At 80.0%, four-fifths of all district revenue comes from charges for services and utilities and from other governments [CG02(4)-2]. Foster distinguishes between districts that have the taxing power and those that operate on service charges, though quite a few have the ability to levy both.^[5]

For the 40-year period between 1957 and 1997, special district expenditures increased 723% in 1997 constant dollars compared to an increase of 301% for other local governments. In these terms, special district revenues rose 796%, as contrasted to an increase of 303% for other local governmental units. This was a period of time during which there was a large increase in the population of metropolitan areas and, in particular, major increases in suburban and fringe area residents in need of services that were not always being provided by general-purpose local governments.^[7]

In 2002, nearly \$28 billion, 21%, of all district revenue were spent by special districts in California. Texas, Illinois, and Florida districts each raised and spent about \$8 billion, with those in Pennsylvania close behind at nearly \$7 billion. Districts in Hawaii spent

only \$21 million. Five percent of these entities, 1757 large districts (defined as those that either raise or spend \$10 million or have a debt exceeding \$25 million), accounted for 82% of the revenue and were responsible for 83% of all district expenditures. The average district the Census classifies as smaller raised \$663,000 and spent \$612,000, but some of the smallest spent only a few hundred dollars. Transit and transportation districts in Los Angeles, Chicago, New York, Boston, Philadelphia, Washington, DC, and Los Angeles county raised or spent from \$1.2 to \$3.0 billion each, as was the case with Arizona's Salt River Project and the Southern California Metropolitan Water District.^[8]

Functions

While some special districts provide multiple functions, the overwhelming majority provide a single function (90.9%). Almost 7000 are natural resource districts, classified as drainage and flood control, soil and water conservation, and other natural resources. Many single-function entities, such as those for housing and community development (3399), water supply (3405), along with sewerage districts (2004), the multifunction sewerage and water districts (1446), as well as the natural resource districts, have been promoted by federal policies and federal grants. The most numerous single-function types are the 5725 fire protection districts (see Table 1). A study

Table 1 Function activities of special district governments, 2002

Function	Total number of districts	Percent of total
Single-function districts		
Natural resource districts		
Drainage and flood control	3,247	9.3
Soil and water conversation	2,506	7.2
Other natural resources	1,226	3.5
Other single-function districts		
Fire protection	5,725	16.3
Water supply	3,405	9.7
Housing and community development	3,399	9.7
Sewerage	2,004	5.7
Cemeteries	1,666	4.8
Libraries	1,580	4.5
Parks and recreation	1,287	3.7
Health	753	2.1
Highways	743	2.1
Hospitals	711	2.0
Education	518	1.5
Air transportation	510	1.5
Gas and electric utilities	485	1.4
Solid waste management	455	1.3
Industrial development and mortgage credit	234	0.7
Other transportation	205	0.6
Welfare	57	0.2
Other and unallocatable	1,161	3.3
Subtotal, single-function districts	31,877	90.9
Multiple-function districts		
Sewerage and water supply	1,446	4.1
Natural resources and water supply	102	0.3
Other multifunctional	1,627	4.6
Subtotal, multifunction districts	3,175	9.1
Total, all districts ^a	35,051	100.0

^aThe one district in the District of Columbia, the Washington Metropolitan Area Transit Authority, is not included. (From Ref.^[12].)

by the U.S. Advisory Commission on Intergovernmental Relations found that fire departments were able to obtain funds from about 50 different federal grants.^[9]

Some 19,252 or 55% of all special districts operate without paid personnel. In other words, they have no employee who represents a 0.5 fulltime equivalent (FTE), which the Census Bureau counts as one paid employee. This is, at least partly, due to the fact that districts are often used simply as a financing mechanism for some new facility such as street lighting or school buildings. Most of these zero FTE units contract out for the services they provide, while the 58% of zero FTE fire protection districts use only volunteer personnel. Libraries and gas and electric utilities, at about 9% zero FTE, are the least likely to contract out or use volunteers in the performance of the services they provide. As of 2002, zero FTE special districts existed in every state, with a range from one in Alaska and 11 in Hawaii to 1300 in California and 1878 in Illinois. In fact, another four states (Kansas, Nebraska, Pennsylvania, and Texas) have from 900 to 1250 of these zero FTE units. Research by the author for the 50 states indicates that there is a high correlation between FTE personnel per unit of population and activity on the part of special district governments.^[2,7,10,11]

Evaluation

Special districts have long been criticized by scholars for adding to the proliferation of local governments and the continuing confusion as to which local government is responsible for a particular area of public policy and the provision of whatever local public service is of concern.^[3,13–15] There is seldom any service or financial coordination among districts and between districts and other local governments. While most of the large numbers of districts concerned with natural resources are located in rural areas, many other types of districts are within the confines of metropolitan statistical areas (MSAs). At least 43% are located in MSAs, while the overwhelming majority of the 1757 special districts the Census classifies as large are located in MSAs [GC02(4)-2, 50–70].^[5] No one knows how many are located in the 575 micropolitan statistical areas.^[16]

As of 1997, 19 county areas in the United States have 100 or more special district governments, all located in or adjacent to MSAs. They ranged from 100 in Fairfield County, Connecticut, to 445 in Harris County (Houston), Texas.^c Moreover, Harris County, at 280, has more zero FTE districts than any other

county area in the country, most supposedly involved in the provision of sewerage and water services. The eight-county Houston consolidated metropolitan statistical area (CMSA) has a total of 741 special district governments; 60% of these districts operated as zero FTE units. Eight county areas with over 100 special districts were located in California, three in Illinois, two in Texas, and one each for Colorado, Connecticut, Nebraska, New Mexico, New York, and Pennsylvania. The total number of local governments in these county areas ranged from 125 in Fairfield County, Connecticut, to 547 in Cook County (Chicago), Illinois.

In terms of services provided, taxes, and service charges, it is highly problematic whether even a minority of residents of metropolitan areas are aware of which government is providing or funding a particular public service or public facility. In fact, counting all types of local governments, as of 1997, there were 79 county areas in the United States, with more than 100 local governments. It is fairly clear that special districts contribute significantly to the balkanization of local government. Over time, the massive increase in the number of districts makes it even more difficult for the public to hold government accountable for the delivery, overall quality, and the financing of public services. Even without considering the thousands of semi-independent districts, the situation with special district governments magnifies the incomprehensible maze of local governments in the nation's metropolitan areas. Sometimes, such confusion exists in the nation's 575 micropolitan statistical areas (microMSAs).^c

Notwithstanding the growth in the use of special districts, this device for the provision of public services raises a number of questions and long-standing concerns. Some of these concerns relate to an inherently undemocratic structure. Only a little over half of the members of governing boards are elected, usually with a minuscule number of voters participating. Other governing board members serve *ex officio* or are appointed by other local governments, the Governor, or the state legislature. In 1965, nearly four decades ago, Martin, in *The Cities in the Federal System*, criticized the use of special districts:

It serves to separate the programs in question . . . from the mainstream of city politics, thereby denying it the juices of democracy . . . [I]t may serve only to substitute for the general politics of the city the particular politics of a special clientele. This exchange is of dubious advantage, the more because special district politics are less visible, less public, than general government politics.^[14]

Martin was right in the sense that special districts are *hidden governments* that operate largely below

^cThese data are not available for 2003 from the Census Bureau. MicroMSAs range in size from about 12,000 to 171,000 residents.^[16]

the level of public, media, and even scholarly attention. More recently, Burns pointed out that:

Simply living in one of these districts does not qualify an American citizen to participate in the district's politics. Unlike other kinds of governments, special districts can have property qualifications for voters. In 1973, the Supreme Court ruled that special districts are exempt from the one-person, one-vote requirement placed on other local governments.^[15]

Burns might have added, where the governing board serves ex officio or is appointed by the Governor or other local governments, that they are at least one more step removed from the political process. Not only may a special district have property qualifications, but, in a few instances, the vote is weighted by the amount of land (area) owned by the voter. Axelrod is highly critical of special districts, which he calls *shadow governments* and what he describes as their nefarious activities.^[3] He contends that not only are they a diminution of democracy, but also they elude the scrutiny of elected officials, are often mismanaged and corrupt, and may be a convenient pork barrel for wealthy private interests. Further, he notes that they are the product of a contradictory political culture that demands public services of local governments but refuses to pay increased taxes, so they are provided on a fee for service basis.

Districts are inordinately involved in the provision of system-maintenance services such as fire protection, sewerage and water services, drainage and flood control, mass transit, roads and airports, and mass transportation. This device has not proved viable in responding to social problems such as crime, drugs, and poverty associated with our core cities and inner suburbs. On the other hand, special districts deliver or provide the financing for capital intensive system-maintenance services. They are able to surmount geographic, political, and financial barriers associated with many, if not most, general-purpose local governments. In metropolitan areas, they often can deliver a service or a combination of services on a regional or semiregional basis, thereby overcoming the limitations of municipal, township, or county boundaries. In some cases, they cross state lines. They may, thereby, even be capable of achieving economies of scale in the provision of selected services.

Special districts are often used as a mechanism to overcome the financial constraints, tax, and debt limitations, placed on other local governments by state laws and constitutions. Given these constraints, special districts have proven to be a viable financial mechanism through which it is possible to issue bonds, levy user fees, property taxes, and special sales taxes for the provision of capital intensive services. Special districts

make extensive use of service and utility charges as a source of revenue, and the proportion spent on capital outlay is nearly twice that of other local governments. Even though district revenues and expenditures are less than 9% of the total for local governments, they account for nearly one-fourth of all local government debt. At the same time, they are responsible for almost 33% of locally owned public utility revenue and over two-fifths of locally owned public utility expenditures.^[8]

CONCLUSION

The special district mechanism has proven to be attractive, as it offers a politically feasible alternative for the provision of specific services to two or more municipalities or townships, or even on occasion, for a metropolitan area, without changing the political structure for existing local governments. They can also be downscaled to provide a service to a small area within a municipality or county, where the service in question is not needed in other parts of the existing local entity. While these hidden governments clutter up the political landscape, operate largely without public scrutiny, and almost universally lack effective public representation, special districts are playing an increasing role in the American polity.

REFERENCES

1. Graves, W.B. *American Intergovernmental Relations*; Charles Scribner's Sons: New York, 1964.
2. U.S. Bureau of the Census, Census of Governments. In "Government Organization," GC57(1)-1 to GC02(1)-1; U.S. Government Printing Office: Washington, DC.
3. Axelrod, D. *Shadow Government: The Hidden World of Public Authorities-and How They Control Over \$1 Trillion of Your Money*; John Wiley and Sons: New York, NY, 1992.
4. Stephens, G.R.; Wikstrom, N. Trends in special districts. *State Local Gov. Rev.* **1998**, 30 (2), 133.
5. Foster, K.A. *The Political Economy of Special-Purpose Government*; Georgetown University Press: Washington, DC, 1997; 1-13.
6. Bollens, J.C. *Special District Governments in the United States*; University of California Press: Berkeley, CA, 1957.
7. Stephens, G.R.; Wikstrom, N. *Metropolitan Government and Governance: Theoretical Perspectives, Empirical Analysis, and the Future*; Oxford University Press: New York, NY, 2000; 16-17, 129; see also, Stephens, G.R.; Wikstrom, N. *American Intergovernmental Relations: A Fragmented Federal Polity*; Oxford University Press: New York, NY, 2006; 217-245.

8. Census of Governments. In *Finances of Special District Governments*; 2002; Vol. 4, No. 2; 6–7, 20–49, 50–70.
9. Reeves, M.M. The federal role in local fire protection. In *The Federal Role in the Federal System: Dynamics of Growth*; U.S. Advisory Commission on Intergovernmental Relations: Washington, DC, 1980; A-85, 178 pp.
10. Governments Division Census of Governments (Government Organization) (GC02(1)-1), 2002.
11. Stephens, G.R.; Wikstrom, N. *Midwestern Townships, Special Districts, and Other Toy Government Manuscript*; University of Missouri-Kansas City and Virginia Commonwealth University, 2001; 25 pp.
12. Koerben, W.K. *U.S. Bureau of the Census*; 2001.
13. Martin, R.C. *Grass Roots*; University of Alabama Press: Birmingham, AL, 1957; 26–28.
14. Martin, R.C. *The Cities and the American Federal System*; Atherton Press: New York, NY, 1965; 178 pp.
15. Burns, N. *The Formation of American Local Government: Private Values in Public Institutions*; Oxford University Press: New York, NY, 1994; 12 pp.
16. Bureau of the Census. In *Statistical Abstract of the United States*; Government Printing Office: Washington, DC, 2006; 901–922.

Stage Heuristics in Policy Sciences

Christine R. Martell
Peter deLeon

*Graduate School of Public Affairs, University of Colorado at Denver and
Health Sciences Center, Denver, Colorado, U.S.A.*

INTRODUCTION

This entry develops the relationship between the concept of policy sciences^[1–3] and their relevant stages—a common heuristic of the policy sciences. In concert, they offer a unique approach to problem solving. The first part of this entry develops the evolution of the underlying assumptions of the policy sciences while the second sets out an analytic framework of stages derived from the policy sciences as a means of viewing policy decisions. Finally, the third part offers observations about the challenges the policy sciences and the stages face in policy analysis.

THE EVOLUTION OF THE POLICY SCIENCES

The development of the policy sciences approach has generally been ascribed to the works of Harold D. Lasswell^[1–3] and his colleagues soon after the end of the Second World War. The stated purpose of the policy sciences was to improve the quality of information offered to government and, by so doing, to improve the quality of governance itself. But the ontogeny of advice to the ruler goes back to virtually the beginning of recorded history, when people wondered as to the application of knowledge to the public sector to much the same end.^[4] Certainly the priests of antiquity were actively involved in political decisions, as kings would ask them to divine (and possibly sway) the “mood” of the gods prior to battle or even before a royal nuptial.

By the eighteenth century, the Age of the Enlightenment infused Western Civilization with a belief in absolute certainty that—through study and reflection—a rational or absolute “truth” covering a wide range of subjects, including politics, could be discerned and acted upon. The Progressive Era in the United States at least partially mirrors the purported marriage of good science and good government,^[5] even if politics and their administrations were somehow considered by some to be mutually exclusive.

The socio-political history of the twentieth century contributed in a sustained way to the movement of higher education (i.e., universities) into the realm of social application. The First World War saw the

emergence of psychology as a means to help winnow recruits for the American Expeditionary Force. The worldwide and sustained traumas of the Great Depression brought economics into virtually every government office. The horrors of World War II fascism made it apparent that governments could easily be used for ill purposes as well as good. This realization was especially clear for Lasswell, who spent the wartime years studying propaganda in the Library of Congress.

THE POLICY SCIENCES FRAMEWORK

Beginning in the late 1940s,^[1,2] Lasswell began to sketch out the thrust of the policy sciences with three clear and abiding components that differentiated them almost radically from the traditional research canons as practiced in higher education. Specifically, the policy sciences were meant to be:

- *Problem-oriented:* That is, they were intended to address problems of significant social import rather than those of strictly academic concern. The policy sciences would not address the number of angels dancing on the head of any given pin, but rather the causes and alleviations of social unrest and inequalities. In Lasswell’s words,^[2] “The Policy [Sciences] approach does not imply that energy is to be dissipated on a miscellany of merely topical issues, but rather that fundamental and often neglected problems, which arise in the adjustment of man in society are to be dealt with” (p. 14).
- *Multi-disciplinary oriented:* Lasswell^[2,3] saw the problems of the “real world” as having multiple “explanations,” such that in almost all cases, the application of a single discipline—be it law or psychology or sociology or economics—would be an unnatural constraint and excessively narrow in scope, and therefore, in all likelihood, it would fail to reveal the complete scope of the problem or its resolution. This tenet, of course, violated the entire departmental structure of the contemporary university, which by this time was disciplinarily focused on increasingly narrow research topics.

- *Explicitly normative*: The effects of the Enlightenment were most apparent in the exposition of and general adherence to logical positivism, a “scientism” that neglected the role of norms and values and ethics in social decisions (Social Darwinism and logical positivism are examples). This was the situation in social sciences such as economics, psychology, and behavioralism in political science. Yet Lasswell was certain that norms had an explicit place in policymaking and needed to be emphasized in the policy sciences.

The third of these hallmarks was explicitly spelled out by Lasswell as he and Abraham Kaplan^[6] articulated that the policy sciences were designed to provide “intelligence pertinent to the integration of values realized by and embodied by interpersonal relations [such as] human dignity and the realization of human capacities” (p. 15). More specifically, Lasswell wrote that “the policy sciences of democracy [were] directed towards knowledge needed to improve the practice of democracy,”^[2] thus enshrining the normative concept of the “policy sciences of democracy” (p. 15).

Two additional features of the policy sciences were their emphases on contextuality and the policy process. The first referred to the concept that no policy research could be conducted without a thorough understanding of the inherent contextuality of the problem—in other words, what were the conditions leading up to the recognition of the problem and in what environment did the problem exist? While this concept is certainly important in a world in which the policy sciences emphasize social problems, it does lend itself to a fundamental, endemic concern of the policy sciences: if all knowledge is dependent upon the overall environment in which a problem exists, that naturally relegates a general *theory* of the policy sciences themselves to a secondary position.

Lasswell also indicated a framework for understanding the events of (what he termed) the “policy process.” Formally posed in 1956, it has become the touchstone of most research in the policy process; many authors^[7–9] have formulated their texts around the process and numerous others have written on the various stages. As posed by Brewer and deLeon,^[7] the stages are:

- *Initiation*: The recognition of a problem, an understanding of what goals (e.g., political, economic, and normative) are to be sought, and the preparation of possible options to attain these goals within their constraints.
- *Estimation*: The balanced assessment of the “costs” and “benefits” (broadly defined) characterizing the various policy options to be delivered to government policymakers.

- *Selection*: The stage in which decisions addressing the problem are made by the policymakers (generally, but not always and surely not exclusively) on the prior analytic work.
- *Implementation*: The stage in which the chosen option is put into practice (usually by administrators, but more recently through the nonprofit sector) such that the programs have an effect on their intended recipients.
- *Evaluation*: The stage in which the program is assessed from a retrospective vantage point, its various aspects are evaluated, and when and where necessary, changes are made in the program and its implementation.
- *Termination*: In some cases, the final stage in the policy process, in which a program (for a number of reasons) is seen to have lost its effectiveness and is therefore terminated.

It is important to recognize that this framework (some have called it a “heuristic,”^[10]) is not presented as a formal “model” nor does it assume any predictive capabilities. Instead, it is a framework for understanding and structuring a particular problem for analysis and, as such, it serves a significant purpose.^[11] Moreover, its stages are clearly interactive with multiple feedback loops between them, rather than necessitating a linear, lock-step progression.^[7]

THE STATE OF THE STAGES

While the stages continue to be used as a framework for understanding the policy process, as witnessed by recent publications, there is no shortage of critique of their use.^[4,11–13]

Most prominent among the critiques of the stages as they apply to policy analysis are that they lack causal power and an hypothesis testing basis, they neglect the role of ideas, and they fail to account for policy-oriented learning.^[14] DeLeon and Martell^[13] concur that the policy sciences poorly incorporates active policy learning: “In many ways, policy analysis failed to marshal the evidence that policymakers could have gleaned from past policy failures, suggesting a bigger human challenge to the policy sciences” (p. 27). The policy sciences assume an environment with a pluralistic democracy and may have less application in other forms of democracies and political cultures.^[15]

Others frame their critique of the use of stages and policy sciences in reference to a learning curve. Radin^[16] considers the historical use of the policy sciences to be a growth industry, whereby the policy sciences are increasingly used as the framework diffuses over time. DeLeon^[4] views the historical evolution of the policy sciences in terms of lessons learned,

largely due to overwhelming policy failure. The lessons are that policy initiatives must be consonant with policymakers' requirements, that the policy sciences must respond to complexity and implementation design, that the policymakers and analysts need to frame the right questions, that good analysis in the policy sciences tradition cannot be separated from values (core values, morals, and ethics), and that analysts need to reflect the required political will.^[4,13]

It is in the context of previous failures and critiques of the policy sciences and stages that this discussion proceeds. First, we summarize prescriptions that have been offered to improve the policy sciences and stages in policy analysis. Second, in light of these prescriptions, we consider what improvements can be made at each stage of the policy process. Finally, we conclude by addressing some issues that to date are unresolved.

Summarize Prescriptions

If a central problem is how to reconcile the policy sciences and stages with Lasswell's normative, value orientation, deLeon and Martell^[13]—based on the work of Fischer,^[17] Forester,^[18] Hajer,^[19] Dryzek^[20,21] and Roe^[22]—recommend integrating a postpositivist perspective into the policy sciences to deal with diverse, interconnected, value-laden policy issues. The main advantages of this strategy would be in providing *understanding* of a relationship, providing a wider context to a policy issue, and understanding of the policy process.

This view is analogous to Lasswell's prescription of a "decision seminar," but often it is not followed in practice to carry out a multi-dimensional analysis of a policy issue. Even if perspectives offer divergent views of an issue, examining a policy problem from different lenses "illustrates how a policy problem is multi-faceted, with each perspective introducing a host of issues. The more that the perspectives introduce issues, the more policy goals become multi-dimensional and the acceptable solutions necessarily become multi-disciplinary" (p. 4).

Improve Stages

Given these prescriptions and the various critiques of the policy sciences, each of the stages could be shored up to honor the potential value of the policy sciences.

- *Initiation*: Needs to be coordinated around the right issue at the right time to the right person/agency. In this sense, policy analysts need to consciously widen the parameters of question definition, explore multifaceted understandings of a problem, and accurately frame issues.

- *Estimation*: Needs to incorporate negotiation, mediation, and resolution as a means for settling or at least framing value disputes.
- *Selection*: Needs to apply learning from past experience and broaden the parameters so that contradictory opinions are not excluded.^[23]
- *Implementation*: Needs to recognize that policy design is crucial to the success or failure of most policies.^[24,25]
- *Evaluation*: Needs to account for intended and unintended consequences, including distributional effects. Policy analysts should evaluate how a chosen issue frame affects outcomes and be conscious of whose values are attended and ignored.
- *Termination*: Needs to recognize that termination may occur because of shifts in values, norms, ethics, or political will. Termination becomes a policy choice even if a specific program is functioning as intended.^[26]

Irresolvables

In spite of the added value the policy sciences have offered the policy analysis world, any discussion would be remiss to neglect the following troubling questions related to the problems of integrating the policy sciences with a value orientation. The problem arises in how to make use of two paths—the predictive product of a positivist approach and the understanding of a postpositivist approach, neither of which is epistemologically flawless.

Would a postpositivist approach to initiation and estimation better align or wedge apart analysts' supply with policymakers' demand? If postpositivist methodologies bring more citizen input to the analyst and open the analysis beyond the standard option set (Initiation and Estimation), then would it also further divorce the usefulness of analysis and the role of the analyst, from the policy maker who is driven by political will? For the policy sciences to be an effective framework, do values and political will have to be the same thing? Alternatively, what shape and what types of relationships should policymakers, analysts, and citizens have with each other? Is it possible for the policymaker to acknowledge, respect, and respond to the values put forth by citizens? If so, would that happen via a policy analyst intermediary or best through direct democracy?

Another unresolved issue about the policy sciences framework is "whose" values. Lasswell, deLeon, and others speak of the importance of norms and values. What the policy sciences fail to do is present a reliable framework for whose values, thus neglecting the power dimension that underlies the importance of postpositivist thought. The prescriptions only guide that values

should be identified, and they provide no guidance on *how* to evaluate values of competing perspectives or persons.

The policy sciences, manifested in the stages, continue to offer a framework with which to understand policy. It is argued that the usefulness of the policy sciences for understanding multifaceted, complex, interconnected, value-laden problems is best enhanced with additional elements for a postpositivist orientation. It is also acknowledged that greater discourse and participation does not guarantee that analysis done according to the policy sciences framework will identify the “right” values nor, of course, does it promise the political acceptance of such arguments. It may, however, allow for the appropriate discussion of the tradeoffs of competing values. While the positivist perspective is necessary for the more tractable problems, the postpositivist perspective aids in understanding. Both resources serve policy analysis and should be used as the research problem dictates.^[27]

ARTICLE OF FURTHER INTEREST

Policy Sciences Approach, p. 1495.

REFERENCES

1. Lasswell, H.D. *Power and Personality*; Norton: New York, 1949.
2. Lasswell, H.D. The policy orientation. In *The Policy Sciences*; Lerner, D., Lasswell, H.D., Eds.; Stanford University Press: Palo Alto, CA, 1951; 3–15.
3. Lasswell, H.D. *A Pre-View of Policy Sciences*; American Elsevier: New York, 1971.
4. deLeon, P. *Advice and Consent*; The Russell Sage Foundation: New York, 1988.
5. Stillman, R.J., III. *Preface to Public Administration*; St. Martin's: New York, 1991.
6. Lasswell, H.D.; Kaplan, A. *Power and Society*; Yale University Press: New Haven, CT, 1950.
7. Brewer, G.D.; deLeon, P. *The Foundations of Policy Analysis*; Brooks/Cole: Monterey, CA, 1983.
8. Anderson, J.E. *Public Policy Making*, 2nd Ed.; Holt, Reinhardt, & Winston: New York, 1979.
9. Jones, C. *An Introduction to Public Policy*, 3rd Ed.; Wadsworth: Belmont, CA, 1984.
10. Sabatier, P.A. Policy change over a decade or more. In *Policy Change and Learning*; Sabatier, P.A., Jenkins-Smith, H.C., Eds.; Westview Press: Boulder, CO, 1993; 13–39.
11. deLeon, P. The stages approach to the policy process: What has it done? Where is it going? In *Theories of the Policy Process*; Sabatier, P.A., Ed.; Westview Press: Boulder, CO, 1999; 19–32.
12. Sabatier, P.A. *Theories of the Policy Process*; Westview Press: Boulder, CO, 1999.
13. deLeon, P.; Martell, C.R. The policy sciences: past, present, and future. In *Handbook of Public Policy*; Peters, B.G., Pierce, J., Eds.; Sage Press: London, 2006.
14. Sabatier, P.A., Jenkins-Smith, H.C., Eds.; *Policy Change and Learning*; Westview Press: Boulder, CO, 1993; 13–39.
15. Mendoza, E.C. Usos y Costumbres en la Hechura de las Políticas Públicas en México. Límites de las Policy Sciences en Contextos Cultural y Politicamente Diferentes. *Gestión y Política Pública* **2000**, IX (002), 180–229.
16. Radin, B.A. *Beyond Machiavelli: Policy Analysis Comes of Age*; Georgetown University Press: Washington, DC, 2000.
17. Fischer, F. Beyond empiricism: policy inquiry in postpositivist perspective. *Policy Sci. J.* **1998**, 26 (1), 129–146.
18. Forester, J. *Deliberative Practitioner: Encouraging Participative Planning Processes*; The MIT Press: Cambridge, MA, 1999.
19. Hajer, M.A. Discourse coalitions and the institutionalization of practice the case of acid rain in great Britain. In *The Argumentative Turn in Policy Analysis and Planning*; Fischer, F., Forester, J., Eds.; Duke University Press: Durham, NC, 1993; 43–76.
20. Dryzek, J.S. *Discursive Democracy: Politics, Policy, and Political Science*; Cambridge University Press: Cambridge, UK, 1990.
21. Dryzek, J.S. *Deliberative Democracy and Beyond*; Oxford University Press: Oxford, UK, 2000.
22. Roe, E. *Narrative Policy Analysis*; Duke University Press: Durham, NC, 1994.
23. Janis, I.L. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*, 2nd Ed.; Houghton Mifflin Company: Boston, MA, 1982.
24. Pressman, J.; Wildavsky, A. *Implementation*, 3rd Ed.; University of California Press: Berkeley, CA, 1984.
25. Stone, D. *Policy Paradox: The Art of Political Decision Making*; W.W. Norton & Company: New York, 2001.
26. Kaufman, H. *Time, Change, and Organizations*; Chatham House: Chatham, NJ, 1991.
27. deLeon, P. Models of policy discourse: insights versus prediction. *Policy Sci. J.* **1998**, 26 (1), 147–162.

Standard of Award

Frayda S. Bluestein

*School of Government, University of North Carolina at Chapel Hill,
Chapel Hill, North Carolina, U.S.A.*

INTRODUCTION

The standard of award is an important aspect of the competitive bidding process in the public procurement process. Indeed, public procurement may be best known for its reputation as requiring the purchase of poor-quality goods or services under the “low bid” standard for awarding contracts. Whether or not this reputation may have been deserved at one time, a review of modern laws and regulations illustrates that the public procurement process has evolved into a much more quality-focused endeavor than is reflected in that common perception. Many public agencies include performance and quality requirements in specifications, coupled with a more generous standard for awarding contracts in order to secure responsible and responsive contractors through the competitive process.

OVERVIEW

The standard of award for public contracts may be found in federal, state, and local laws and regulations. The specific wording of the standard delineates the scope of authority a public agency has in awarding contracts that are subject to the bidding requirement. The standard of award determines what factors the agency may consider in choosing the successful bidder, and it affects the rights and remedies of the bidders who participate in the process. Examples of the wording that is used in standards of award include “lowest and best bid,” “lowest responsive and responsible bidder,” “responsible offeror whose proposal is determined to be the most advantageous to the [agency],” and “lowest bidder.”

In general, a standard of award that contains qualitative or subjective terms, such as “best,” or “most advantageous,” or a standard that specifically authorizes consideration of factors such as responsibility, quality, or performance, provides greater discretion to the awarding authority than a standard of award that requires only the lowest bid, without consideration of other factors. The trend in public procurement is to allow greater consideration of factors such as quality and performance over time, rather than strictly purchase price.

The standard of award may have several components—responsiveness, responsibility, and cost.

RESPONSIVENESS

Many standards of award incorporate the requirement that bids must be “responsive.” The requirement of responsiveness is implicit in the bidding process, even when not specifically stated as a part of the award standard. While the legal standards affecting the award of contract vary considerably among jurisdictions, the standard for determining responsiveness is nearly universal. A responsive bid is one that conforms substantially to the specifications and the invitation for bids and will, if accepted, form the basis for an enforceable contract (see Ref.^[1]). Responsiveness is a threshold determination in the bid evaluation process that each bid must satisfy in order to be eligible for award.

The public procurement process allows the public agency to accept bids that contain minor variations from specifications. These variations may be waived as long as the waiver does not give one bidder an advantage that is not enjoyed by the other bidders.^[2,3] This standard is generally enforced by administrative or judicial review, which may be sought by a bidder who is disadvantaged by the improper waiver of a deviation from specifications. While the requirement of responsiveness helps guarantee that bids will meet the needs of the agency as set forth in the specifications and invitation for bid, the need for fairness in the competitive process is also promoted by the responsiveness standard as enforced by the courts. The limitation on the ability of the agency to waive substantial deviations from specifications provides protection for the bidders who participate in the competitive process by ensuring that the terms of the solicitation are not, in effect, changed to the benefit of a particular bidder.

RESPONSIBILITY

Another term commonly found in standards of award is “responsibility.” In selecting the lowest “responsible” bidder, public agencies may consider “the bidder’s experience, skill, ability, business judgment, financial situation, integrity, honesty, possession of the facilities

necessary to perform the contract, previous conduct on similar contracts, reputation and record for reliability, as well as any other factors reasonably relevant to a bidder's successful performance if awarded the contract.''^[4] The awarding authority generally has broad authority to require bidders to submit information necessary for the agency to evaluate issues of responsibility, including financial statements, references, and records demonstrating capability and experience.

Some jurisdictions require that when a contract is awarded to a bidder who was not the lowest bid, the agency must provide the low bidder with notice of the decision and, in some cases, an opportunity for a hearing and to present evidence (see Refs.^[5,6]). This may reflect a concern that a determination as to the responsibility of a bidder may implicate the bidder's constitutionally protected due process rights. Judicial decisions have held that the bidder may be entitled to due process if a governmental agency's determination that the bidder is non-responsible would tarnish the bidder's business reputation and limit the bidder's liberty to pursue his or her chosen trade (see Ref.^[7]). Whether a particular decision implicates the liberty interest is not always clear, but the practice of providing notice and an opportunity for hearing may protect against potential claims under this theory. It may also help support the public agency's decision in the event it is challenged as being arbitrary or otherwise unlawful.

COST

A key purpose of competitive bidding is to ensure the prudent use of public funds in the procurement of goods and services. As such, it is important that the evaluation of cost not be limited to mere purchase price (see Ref.^[3]). Specifications and standards of award may incorporate consideration of value over the life of the goods or course of service, as well as considerations of quality, appropriateness, and usefulness for the desired purpose. In order to ensure fairness, however, the requirements of the agency and the bases for evaluation of cost and quality must be set forth in the specifications or other solicitation documentation. When a public agency awards a contract to a bidder who has not submitted the lowest bid, the basis for the decision should be documented by the awarding agency.

DISCRETION IN AND JUDICIAL REVIEW OF CONTRACT AWARD DECISIONS

An important issue that arises from the particular wording of a standard of award is whether or to what extent the standard allows the public entity to award a contract to a bidder whose bid does not offer the

lowest price. In evaluating the degree and type of discretion the awarding authority may exercise in evaluating bids, courts look to the specific wording of the standard of award. In some cases, a statute, regulation, or ordinance may list the specific factors that may be considered, such as inspection, testing, quality, workmanship, delivery, and suitability for the particular purpose. More commonly, the legal standard uses broader terms, such as quality and performance, which must then be interpreted. A narrowly worded standard, such as a "low bid" standard, may allow rejection of a low bid only if it is non-responsive or is otherwise ineligible for award. A more broadly worded standard may allow the agency to choose a higher-priced bid that offers a higher quality or performance to the agency. In such cases, the burden will generally be upon the agency, in the event of a challenge, or as part of the statutorily mandated procedure, to document the basis for the decision. In some jurisdictions, it is necessary for the agency to establish that the lowest bidder is not responsible, before an award may be made to the next lowest bidder.^[8]

The amount of discretion afforded by the standard of award also affects the extent to which a court, upon review in the event of a legal challenge, may scrutinize the agency's support for its decision. The greater the discretion afforded to the awarding authority in the standard of award, the more difficult the prospects for a disappointed bidder's challenge. When applying a standard of award that contains subjective standards, the awarding authority must exercise judgment, to which a court, in the event of a challenge, is likely to defer. This places a high burden on the challenger, who may be required to demonstrate fraud or an abuse of discretion in order to invalidate the award. A typical standard of review, as enunciated courts, is that the court will not interfere with the decision of the public agency unless the decision is arbitrary, made in bad faith, or constitutes an abuse of discretion.^[1]

The degree of discretion contained in the standard of award has been a factor courts have considered in determining what rights and remedies, if any, a bidder has with respect to the bid evaluation and award process.^[9] While in a few jurisdictions, a hearing and protest procedures are explicitly provided for by statute or regulation, others require recourse to judicial interpretation for guidance about bidders' remedies. Though most courts recognize a right to injunctive relief (allowing a bidder to seek a court order declaring the process invalid), it is more rare for courts to award lost profits or damages to disappointed bidders. Some courts have explicitly denied such recovery on the theory that because of the discretion involved in the contract award process, the complaining bidder has no legitimate claim that he or she would have been awarded the contract (see Refs.^[10,11]).

CONCLUSION

The standard of award, and its lawful application, reflect the dynamic tension in the dual purposes that underlie competitive bidding requirements in public procurement. These requirements are designed to promote value and economy, as well as fairness in the contracting process. In awarding contracts, public agencies must not exceed the degree of discretion afforded by the applicable standard of award, but must also apply to the best advantage of the agency the concepts of quality and responsibility. Finally, decisions must be documented in order to ensure fairness and promote openness in the contracting process, as well as to provide legal support in the event of a challenge.

REFERENCES

1. Evans, T.; O'Gallagher, J. *Eugene McQuillin: The Law of Municipal Corporations*, 3rd Ed.; West Group: New York, NY, 1999; § 29.65, § 29.73.
2. American Jurisprudence. Public works and contracts. In *American Jurisprudence*, 2nd Ed.; West Group: New York, NY, 2001; § 58.
3. National Association of State Purchasing Officials (NASPO). *State and Local Government Purchasing*, 4th Ed.; NASPO: Lexington, KY, 1994; 74–78.
4. Stevenson, S.M. *Antieau on Local Government Law*, 2nd Ed.; Mathew Bender & Co., Inc.: New York, NY, 2001; § 32.04[5].
5. La. Rev. Stat. Ann., § M38:2212 (West 1989).
6. Cal. Pub. Cont. Code, § 10306 (West 1985).
7. *Omaha v. Solomon*, 960 F.2d 31 (6th Cir. 1992).
8. *City of Inglewood–Los Angeles County Civic Center Auth. v. Superior Court*, 7 Cal. 3d 861, 103 Cal. Rptr. 689, 500 P.2d 601 (1972).
9. Isham, J.L. Public contracts: low bidder's monetary relief against state or local agency for non-award of contract. In *American Law Reports*, 4th Ed.; Lawyers Co-operative Publishing Co.: Rochester, NY, 1998; Vol. 65, 93 pp.
10. *Sowell's Meats & Services v. McSwain*, 788 F.2d 266 (4th Cir. 1986).
11. *Kim Construction v. Mundelein*, 14 F.3d 1243 (7th Cir. 1994).

Standards of Conduct

Jeremy F. Plant

*School of Public Affairs, Pennsylvania State University at Harrisburg,
Middletown, Pennsylvania, U.S.A.*

INTRODUCTION

Standards of conduct are statements about right behavior intended to guide the actions of individuals within organization, public jurisdictions, professions, or other bounded social institutions. Standards of conduct are usually justified by relating them to norms and values derived from religion, philosophy, political principles, law, or custom and tradition. They often include methods of enforcement and penalties for non-compliance. Standards of conduct have two common characteristics, whatever their basis and wherever found: They relate to the role performed by the individual within an organization or public jurisdiction and apply regardless of who occupies a position; and they are an effort to guide actual behavior, not just to identify basic norms and values.

DEFINING STANDARDS OF CONDUCT IN PUBLIC ORGANIZATIONS

Defining standards of conduct in public organizations raises several fundamental questions. What are the acceptable standards of conduct? Who shall determine what they are and by what authority? How will the standards be communicated to the members of the organization? What, if any, means of enforcing compliance with the standards will be used? Will standards state the minimal acceptable behavior or prescribe lofty goals or standards that may be hard to measure and attain?

One approach to defining standards of conduct rests on well-accepted parameters of behavior defined as corruption or corrupt acts. Corruption is usually proscribed by law and includes such behaviors as acceptance of items of value, bribery, nepotism, gross conflicts of interest, plagiarism, or conspiracy to violate laws or to prevent officers of the law from doing their work. In recent years, sexual harassment has been added to the list of actions deemed corrupt. Corrupt acts are such blatant violations of standards that they usually rest on unquestioned legal and moral bases.

Efforts to create a professional and non-partisan public service in the United States identified politics, especially partisan politics, as the major source of

ethical problems. The desire to insulate professional administration from political interference led to strict rules separating the realms of politics and administration. A central feature of such standards is the non-involvement of career administrators in political elections. In the reform movement that gave birth to modern public administration, this was a way to prevent partisan politics from interfering with the application of professional management techniques to governance. The Hatch Act, enacted in 1939 and substantially amended several times but still in force, addressed the problem somewhat differently. Its supporters saw in the growing numbers and influence of career public administrators, a possible political force that could wield undue power or disrupt democratic processes.

A second major source of ethical problems is the indistinct line between the public and private sectors. Taken together, these major sources of ethical problems have produced a bias in the United States toward formal standards of conduct that stress the need for political neutrality and the elimination of potential conflicts of interest as individuals move between public and private life.

More difficult to establish are standards of conduct that adapt general moral principles to the work of the public sector and to life in large, complex organizations. Standards of conduct defining corrupt behavior proscribed by law for a large, general population of individuals (e.g., all federal workers, or state and local employees in a state civil service system) may be of little help in guiding behavior of individuals who are trying to understand the nature of right action in complicated organizational and programmatic settings. For this reason, formal codes of conduct or codes of ethics, which have as their intent the application of general rules or values to the workplace, have become increasingly common in recent years.^[1]

CODES OF CONDUCT AND ETHICS IN PUBLIC ADMINISTRATION

Codes of conduct have a long history in the public sector. One of the longest-running uses of a code of conduct is that of the International City/County

Management Association (ICMA), which first established a code in 1924 and has adapted it several times since to meet changing conditions. Federal employees were first covered by a code in 1958, established by concurrent resolutions of Congress and strengthened by President John F. Kennedy by executive order in 1961. In the past three decades, states have been active in codifying standards of behavior for their employees, with around 40 states now employing some sort of code to identify standards of conduct. Local governments have also been active in promulgating codes of conduct.^[2] Professional associations whose membership is composed of public officials, in part or in toto, are another active source of codes specifying right behavior.^[3]

No single factor or explanation accounts for the wave of codifying permissible behavior that has swept over the public sector in the past quarter century. Declining faith in government or in bureaucratic organizations, media attention to corruption in political life, the fear of government power, demands of political elites for high levels of administrative accountability, and the use of codes by professions to increase their power in society have all been offered as reasons for the growing interest in identifying standards of conduct in public life. Codes are often adopted or modified in response to a specific scandal or problem, oftentimes without regard for consistency or logic. Whatever the impetus to codify is, codes are rarely if ever based on solid empirical evidence suggesting that the problem of corrupt or unacceptable behavior is increasing or posing a serious threat to the integrity of society.

BUREAUCRACY AS THE BASIS OF STANDARDS OF CONDUCT

Large, formal organizations have long been seen as a way of promoting ethical conduct and discouraging unethical behavior. Organizational structure can be a mechanism for ensuring compliance with ethical standards, especially those requiring accountability to higher authorities, transparency of decision making, and procedural fairness.^[4] In recent years, such organizational devices as internal audit units, ethics committees, ombudsmen, and ethics hotlines have increased the capacity of many public organizations to monitor the actions of individuals.

Organizations are also seen as a source of unethical behavior. Rules intended to guarantee fairness and accountability may become a means for individuals to avoid responsibility for their actions or may become ends in themselves, without concern for the higher values they are intended to promote. Organizations may develop a culture that allows ethical indiscretions to be overlooked or even provides reinforcement for

unethical behavior. The complexity of organizational processes may make it difficult for individuals to see how their work contributes to organizational goals—or why an occasional ethical lapse may have consequences outside the organization.

Standards of conduct that address proper interpersonal relationships within organizations, and between members of an organization, clients, and citizens, are increasingly common in both the public and the private sectors. They often stress the need for civility toward others and may identify types of language and behavior that are likely to be considered offensive to others. Harassment of others on the basis of sex, sexual orientation, race and ethnicity, age, or physical attributes such as appearance or weight, may be proscribed by explicit organizational rules or be inferred from general statements of expectations for civil behavior within the organization.

Unlike business, public administration has few standards of conduct that are developed by a single organization. Most standards result from efforts to codify right behavior for all organizations within a political jurisdiction (e.g., the Commonwealth of Pennsylvania, or the U.S. government). Others may be borrowed from professional groups that have members outside, as well as within a particular organization.

PROFESSIONALISM AS THE BASIS OF STANDARDS OF CONDUCT

Professionals are those individuals who apply expert knowledge, usually gained through higher education and often guaranteed by licensure, in the performance of their work. Professionals have become pervasive in the public sector; their presence in the public workforce requires altering traditional ideas of personnel policy, organizational control, and standards of conduct.

Professionalism affects standards of conduct in many ways. Professionals share a common identity, which is based not just on shared knowledge but also shared norms and values. They interact on a regular and lasting basis, often through associations that represent their interests and serve as a forum to share knowledge, insights, and experience. Professionals often are guided and governed by codes of ethics that apply general norms of right and wrong behavior to the work context.

Professional codes of conduct may exist independent of general ethics laws or public codes enacted by political institutions, or they may be required as a condition of public employment. They differ from codes promulgated as law or executive order by political institutions in several important ways. They often represent a voluntary commitment by the individual to

adhere to norms and acceptable behaviors. They may be enforced in ways established by the profession itself (e.g., committees established by a professional group to hear complaints and fix penalties, such as disbarment by a state bar association). They attempt to translate general ethical principles of right action to the context of the work performed. They may attempt to preclude even the appearance of wrongdoing, or include statements of ideal behavior, to increase public trust and confidence in the profession.

Professional associations representing well-established and specialized occupations usually find it easier to develop codes and apply them consistently than do associations that bring together individuals of varying educational and professional backgrounds. The American Society for Public Administration (ASPA) has had a code of ethics since the 1980s, but questions linger about its efficacy in actually guiding behavior. The ASPA represents students, academics, and practitioners, and it is hard to find common language to guide the actions of such a disparate audience.^[5]

DEMOCRATIC VALUES AS BASIS OF STANDARDS OF CONDUCT

Standards of conduct derived from professionalism focus on the meaning of general ethical principles in the application of knowledge to work situations. The particular obligations of public employment require that attention be given to standards of conduct derived from democratic values. This requires that fundamental principles of democratic governance be identified and translated into norms of conduct that guide the behavior of public administrators.

Beginning in the 1960s, an important body of literature in the field of public administration has addressed the problem of standards of conduct derived from democratic principles. Three major streams of thought have emerged. The first is based on the writings of John Rawls and is concerned with the advancement of the basic principles of liberty and social justice. A second stream looks at American constitutionalism as the basis of overarching principles to guide administrative behavior. The third approach focuses on individual virtue and the behaviors derived from virtue in the context of democratic governance. Virtue in public service requires a commitment to benevolence, honor, compassion, and trust in others.^[6]

Standards of conduct based on democratic principles often rest on commitments to basic concepts such as liberty, equality, and fairness. The challenge facing makers of codes is to find ways to operationalize such general democratic concepts to guide behavior and to deal with the trade-offs between conflicting values (e.g., the arguments for and against affirmative action,

based on differing notions of freedom and equal treatment).

Kernaghan argues that public service values should be the basis of codes.^[7] Kernaghan studies values statement in four Westminster nations (those with a British form of parliamentary government); Canada, Australia, New Zealand, and the United Kingdom. Based on his research, he identified four categories of public service values: ethical, relating to the integrity and fairness of individuals; democratic, relating to the rule of law, accountability, openness; professional, concerned with effectiveness, quality, and excellence; and people, for whom caring, tolerance, decency, and benevolence are expected in the behavior of public servants. Kernaghan suggests that the success of the values statement is related to the importance given to it by leaders, who should be exemplars of the values they expect to be integrated to the work of public organizations.

STANDARDS: PROSCRIPTIVE OR PRESCRIPTIVE?

A debate continues among scholars and practitioners on the relative value of standards of conduct that state explicitly what an individual may not do—*proscriptive standards*—vs. standards that state aspirational goals, or *prescriptive standards*. Closely associated with this debate is the related issue of whether standards and codes of conduct should be enforced, or should they simply rest for their implementation on the voluntary acceptance of those to whom the standards are addressed.

Example: One goal of standards is to encourage individuals in the public service to develop trust and confidence in others. Is it better to promulgate a code that pledges one to work actively to foster trust in the public service, or to pledge one to do nothing that will lessen trust in the public service, or to develop a series of specific constraints (e.g., dollar limits on gratuities that can be accepted or behaviors that can be considered conflicts of interest)? The first approach is clearly affirmational and prescriptive; the second is prescriptive but stated in less positive or idealistic terms; the third is proscriptive and based on a more negative view of human nature.

Example: The ICMA has had a working code of ethics since 1924. The ICMA code has been revised several times, including the addition of guidelines to help members understand the application of rules in typical work situations. The code is enforced by committees of ICMA members, with several types of sanctions that can be imposed. By all accounts, the membership of the association takes the code very seriously as a guide to action and as a statement of the values and norms that unite the profession. In 1987,

the association added a declaration of ideals that balances the negative, proscriptive nature of the code of ethics with a statement of the need for ICMA members to be committed to humane, equitable, and democratic values. While not enforceable, the declaration of ideals reinforces the commitment of the profession of local government management to the broad goals of democratic governance.

DO STANDARDS OF CONDUCT GUIDE ACTIONS?

How standards and codes of conduct actually guide behavior has become a major issue of concern for the field of public administration.^[8] For standards to influence behavior, the standards have to be promulgated in a timely and effective manner and explained in such a way that they are of use to decision makers. Training sessions may be required to show the application of standards in work situations. The literature on public-sector ethics is not clear on how well public administrators understand standards of conduct or how strong a factor they are in actually producing desired behaviors and attitudes.

Criticism of legalistic approaches to standards of conduct remains strong, even as more and more jurisdictions are adopting such requirements. Much of the criticism rests on the understanding that the work of public administrators involves judgment and discretion and cannot be fully programmed by formal standards. In this view of governance, it is necessary to inject administrative discretion with a sense of obligation to democratic values, as difficult as it may be to stipulate exactly what they require in actual work situations or how they may be enforced.

Others argue that individuals must exercise their ethical standards in the workplace on a regular basis, through actions that involve judgment and responsibility.^[9] Individuals should be shown how their actions relate to broader ethical and democratic values. The development of mature ethical judgment is shown to be of value to the organization, the public, and the individual by virtuous leaders who themselves inject the details of work with informed ethical actions. Through such examples, subordinates are given a better understanding of what it means to be ethically aware. This leadership by example can work with or without explicit codes or standards of conduct.

Final assessment of the importance of standards of conduct, in whatever form they may appear, rests on the added value they provide in democratic

governance. If they help individuals to understand better the organizational, professional, and systemic importance of adhering to consistent ethical behavior, they will lead to more efficient and effective performance and enhanced public trust and appreciation of public affairs. If they inhibit or stifle reasoned ethical judgment, or provide a convenient way for politicians to scapegoat career public administrators for failures of policy, or if they convince qualified individuals not to pursue careers in the public service, their costs will clearly outweigh any benefits they may provide.

REFERENCES

1. Blake, R. The nature and scope of state government ethics codes. *Public Prod. Manage. Rev.* **1998**, 21, 453–459.
2. Bruce, W.P. Codes of ethics and codes of conduct: perceived contribution to the practice of ethics in local government. *Public Integr. Annu.* **1996**, 23–30.
3. Plant, J.F. Codes of ethics. In *Handbook of Administrative Ethics*; Cooper, T.L., Ed.; Marcel Dekker Inc.: New York: NY, 2001, 309–333.
4. Appleby, P.H. *Morality and Administration in Democratic Governance*; LSU Press: New York: NY, 1952.
5. Van Wart, M. The sources of ethical decision making for individuals in the public sector. *Public Admin. Rev.* **1996**, 56, 525–533.
6. Frederickson, H.G. *The Spirit of Public Administration*; Jossey-Bass: San Francisco, 1997.
7. Kernaghan, K. Integrating values into public service: the values statement as centerpiece. *Public Admin. Rev.* **2003**, 63, 711–719.
8. Bowman, J.S.; Williams, R.L. Ethics in government: from a winter of despair to a spring of hope. *Public Admin. Rev.* **1997**, 57, 517–526.
9. Cooper, T.L. *The Responsible Administrator*, 4th Ed.; Jossey-Bass: San Francisco, 1998.

BIBLIOGRAPHY

1. American Society for Public Administration. <http://www.aspanet.org> (accessed June 2001).
2. Commonwealth of Pennsylvania. <http://www.state.pa.us> (accessed June 2001).
3. Ethics Updates. <http://ethics.acusd.edu/index.html> (accessed June 2001).
4. Federal Inspectors General. <http://ignet.gov> (accessed June 2001).
5. International City. <http://www.icma.org> (accessed June 2001).
6. U.S. Office of Government Ethics. <http://www.usoge.gov> (accessed June 2001).

State Administration of Mental Retardation and Developmental Disabilities Policies

Paul J. Castellani

*Department of Public Administration and Policy, University at Albany,
State University of New York, Albany, New York, U.S.A.*

INTRODUCTION

There are several key issues for states in the administration of services for people with developmental disabilities. These issues involve defining the scope of responsibility; controlling the high costs of developmental services; determining whether to administer developmental services separately or, if not, under which alternative framework; designing substate mechanisms for managing a complex array of publicly and privately provided services; and dealing with federal court decisions affecting states' ability to limit access to Medicaid-funded services.

responsibility in this area. Few state administrations (or substate entities) employ single-point-of-access mechanisms that apply uniform criteria defining disability. Moreover, determining eligibility, assessing need, and establishing priorities for access to services are key decisions often left to the private agencies that deliver developmental services. This translates into wide variations among (and often within) states in the numbers and types of individuals served and the costs of these services.^[1,2]

CONTROLLING THE HIGH COSTS OF DEVELOPMENTAL SERVICES

The closure and downsizing of large state-operated institutions and creation of community-based residential and day programs was the most prominent feature of developmental services in the field since the 1970s. Federal funding, primarily through Medicaid, underwrote much of the costs of this transition, supporting the remaining large institutions, as well as financing a substantial proportion of the community programs—the former as large Intermediate Care Facilities for the Mentally Retarded (ICFs/MR) and the latter as small ICFs/MR. Community day services, such as day treatment facilities, clinics, and other programs, were also reconfigured to qualify for federal financial participation.^[3,4] More recently, this federal funding has shifted from the ICF/MR program to the Home and Community-Based-Services (HCBS) waiver.

Medicaid is the primary source of funding for developmental services, and its share of the overall costs of Medicaid is large and growing. Of the approximately 36 million people in Medicaid, about 1.7 million are those with developmental disabilities—4.9% of the total—and triple the approximately 506,000 people with developmental disabilities receiving Medicaid in 1975—2.3% of the 1975 total. Medicaid recipients with developmental disabilities (4.9% of the total) account for 15.7% of total federal-state Medicaid spending.^[1] Controlling the large and growing cost of developmental services has been a central concern for both state and federal budget officials and policy makers.

DEFINING THE SCOPE OF STATE RESPONSIBILITY

The prevalence of severe mental retardation in society (approximately 0.7% of the population) means that states must administer programs for numerous individuals and their families. This is an important issue for state administration because developmental disabilities, by definition, manifest themselves at birth or early in life, are lifelong, and usually involve comprehensive and costly services. Although mental retardation is the most prominent of the developmental disabilities (usually accounting for about 80% of a broad definition), other disabilities such as cerebral palsy, epilepsy, autism, and numerous other neurological impairments are also included. There is wide variation among states in which developmental disabilities are in the scope of the primary state agency or scattered among other agencies. The scope of state responsibility is also complicated by functional definitions of these disabilities, contained in federal and state statutes, which add limitations in major life activities as well as environmental, familial, social, and situational factors to the categorical definitions.

Now that services are less often delivered in state-operated institutions than in thousands of community residential and day settings operated by various public and private agencies, states have great difficulty in controlling other crucial processes that affect their

STATEWIDE ADMINISTRATION: CHOOSING AN APPROACH

In the late 1970s and early 1980s, advocates for people with developmental disabilities were successful in getting separate agencies created to administer developmental services in some states. However, the majority of state developmental services programs are submersed within larger administrative agencies. In many states, the developmental disabilities bureau, office, or division is linked to a mental health unit, and occasionally units dealing with aging, substance abuse, children, and people with other disabilities. These administrative units are embedded in various Departments of Mental Health/Mental Retardation, Health, and Social and Human Services, and often at the second or third levels within the organizational hierarchy. Moreover, various administrative restructurings in the 1990s, under the rubric of reinvention and cost cutting, were universally in the direction of consolidation.

Despite the large numbers of people served in developmental services programs and the relatively substantial costs of these services, in most states, developmental disabilities does not have the cabinet status of programs serving many fewer people and having less substantial operational and fiscal consequence.

Although much of the attention of developmental services advocates and consumers has been on the bureaucratic location of the developmental disability agency, the location of responsibility for Medicaid rate setting and administration has been a less-noticed but important feature in recent reorganizations. In several states, the primary developmental disabilities agency operates the state's large institutions and publicly funded community services, but oversight and funding of private ICFs/MR are the responsibility of separate health financing agencies. The separation of administration, state and Medicaid funding, and rate setting complicates the administration of developmental services in numerous states.

Underlying these diverse approaches to statewide administration of developmental services is the debate about which services framework should guide a state's approach to developmental services.^[5,6]

Supported Living

Concern about the large and growing costs of Medicaid for people with developmental disabilities, as well as the pressures of reinvention, devolution, cost savings, and consolidations, were important forces largely from outside the field of developmental disabilities that resulted in changes in states' administration of developmental services. These external forces were made more potent because they were consistent with demands for change from within the field.

As policy makers and funders grew concerned about the rapidly rising costs of Medicaid for developmental services, advocates in the field pointed out that, although Medicaid was funding more group homes and community-based services, these were largely professionalized, medically oriented programs. Demands for less clinical and more personal supports, direct supports to consumers and families, and more self-direction were central in the approach that became known as supported living or self-determination. Dovetailing with policy makers and funders desires for cost constraints, more funding was shifted from the Medicaid ICF/MR program to the Medicaid HCBS waivers. These waivers were used to fund independent case managers, job coaches, personal care attendants, home modifications, cash subsidies, rent and housing subsidies, and other direct supports to individuals and their natural care givers.

Managed Care

At the same time supported living was being proposed as the new paradigm, managed care has also been preferred as a comprehensive solution for state administration of developmental services. Although advocates saw the capitation mechanisms of managed care as a potential vehicle for reforming the service system, state executives and legislators saw managed care as a strategy for Medicaid reform, cutting the size and growth of costs in an area that had grown disproportionately expensive to the size of the clientele.

Long-Term Care

Long-term care is not a prominent reform theme in this field, but despite the closure and downsizing of large state-operated institutions since the 1970s, large congregate settings are still a major locus of service for people with developmental disabilities. In 1996, there were 59,726 individuals residing in large state-operated institutions, and 79,441 people with developmental disabilities were living in other institutions—38,438 in nursing homes, and 41,003 in private and other large institutions.^[2]

Forensic Services

Throughout the era of the shift from large state institutions to community programs, proponents of continued use of public and private long-term care facilities were largely parents and public employees in those large facilities who raised issues of safety, the availability of medical care, and job security.^[7] In the mid-1990s, the issue of safety reemerged from the

perspective of the community. Megan's Law in New Jersey and the 1997 Supreme Court ruling in *Henricks v. Kansas* that upheld the Kansas Sexually Violent Predator Act, which allowed predatory sex offenders to be paroled from jail to mental hospitals or kept in mental hospitals, refocused attention on individuals who had "challenging behaviors." States have consequently increased the size and built new "secure units" within their developmental services programs, and this enhanced forensic responsibility had become an important feature of states' administration of developmental services.

Paradigms or Boutiques

Medicaid funding through large and small ICFs/MR has imposed an overarching hybrid medical-long-term care framework for developmental services since the mid-1970s. The HCBS waiver provides states the alternative of funding less medicalized services outside the confines of the ICF model, and many advocates for a supported living approach see the waiver as a vehicle for what is termed a new paradigm. However, states are more likely to face the challenge of managing various frameworks for developmental services that arise from different policy choices, as well as important distinctions among the characteristics and needs of people with developmental disabilities.

MANAGING A COMPLEX ARRAY OF PUBLIC AND PRIVATE SERVICES

The state and federal governments are the source of the overwhelming proportion of the costs of developmental services. In 37 of the 51 state (and District of Columbia) jurisdictions, local governments bear none of the direct costs of these services. Local governments in the handful of states with large fiscal contributions play important roles in decisions about funding, access to services, and operation of community developmental services. However, county governments that make no, or negligible, contributions still play important roles in many states, often passing state and federal funding through contracts and making key decisions about which agencies get funding, what services are funded, and who receives access to services. This results in wide variations within many states in the availability and accessibility of developmental services.

In addition to local governments, substate specialized developmental services agencies play important roles in several states. In some states, state regional offices play important roles in decisions about which agencies get funded, what services are funded, and who receives access to services. In other states, local

or regional organizations are non-profit or hybrid public-private organizations, where state and local governments share some governance power with formal parent, advocacy group, and consumer representation. These substate agencies may provide other services or contract with local service provider agencies. They may also act as a single point of entry into the service system, which is a key element of local control.

Administering and Coordinating Lifelong and Changing Services

Because developmental disabilities become manifest at birth or infancy, providing, paying for, or ensuring health and medical services are issues for state administration at the early stages in the life. In addition to health and medical services, until the individual reaches school age, states usually fund, certify, oversee, and manage various early intervention, preschool, and specialized developmental services delivered by various state and local government, school district, and private providers. The financing of early childhood services is a patchwork of Medicaid, private insurance, and state and local government funding, depending on and individual's age and need for services and other eligibility criteria, which change as the individual ages and/or requires different types of services.

Local school districts are mandated by the federal statute, the Individuals with Disabilities Education Act, to provide a free and appropriate public education. Although school districts assume the primary responsibility until the individual "ages out," typically at age 22, the underfunded costs of educational and support services borne by local school districts are a perennial state budget issue. State administration and coordination of support and ancillary service providers of services to school-age individuals is an additional challenge to state education and developmental disabilities agencies.

For individuals younger than age 22, the primary state developmental services agency usually plays a secondary role—certifying early intervention providers, sometimes providing funding, delivering or funding family support services, and playing a role in coordinating services for children younger than age 21. However, when an individual with a developmental disability "ages out" of school, another distinct panoply of eligibility criteria, funding mechanisms, provider organizations, and services comes into play. The mandates of the Individuals with Disabilities Education Act cease; waiting lists are encountered; funding mechanisms change; and services fall within a new state administrative framework: the various state-level departments, divisions, offices, and programs. From the perspective of the state developmental services

agency, the individual who seeks access to adult services is likely to have a 20 or more year service history. State developmental services agencies must deal with parents with experience with mandated educational services, eligibility based on age, and other entitlements and expectations that may not be part of a state's adult developmental services program.

Providing, Funding, and Overseeing Developmental Services

In every state, there is some version of the following organizational landscape: state agency-operated large institutions; state or local government agency-operated community services; and private non-profit and for-profit agencies operating large institutions, group homes, community day programs, and family and individual support programs. The mix varies widely from state to state, represents a combination of strategic or incremental policy choices and private practices, and is undergoing substantial changes in some states. With the privatization and devolution of state operated services, budgeting, rate setting, contract administration, quality assurance, and outcome measurement for private agency contracting represent substantial administrative changes from the mechanisms and procedures of direct government operation.^[8,9]

Most of the attention to these problems has been on the non-profit sector, where the administrative problems of contracting with a fragile non-profit sector and the entrepreneurial growth of some non-profit agencies have been the focus.^[8,10] A more recent phenomenon in the developmental disabilities field has been the growth and entry of large for-profit corporations. Some of these corporations operate hundreds of group homes and day programs in several states and specialize in developmental services. The growth of managed care has brought numerous generic managed care organizations into the field, and although the overall penetration remains low, developmental disability advocacy and provider organizations see these corporations as having the potential to effect substantial changes in the organizational landscape of developmental services. From the standpoint of a free state administration, the challenges of managing networks of small, often fragile provider agencies are obviously different from dealing with large, multistate, multiprogram non-profit organizations and for-profit corporations.

The experience of the past 30 years has been one of managing highly regulated networks: certifying providers, shoring up fiscally fragile agencies (large and small), attempting to ensure some equity in availability of services within and among locales, and managing access of individuals into provider organizations.

The simultaneous growth of small supported living agencies, access of consumers to private market and generic sources of support, large multistate developmental services corporations, and managed care organizations presents an exponential increase in complexity for state administration in budgeting, rate setting, contract administration, quality assurance, constituency relations, and policy making.

COURTS AND STATE ADMINISTRATION OF DEVELOPMENTAL SERVICES

The final major issue in state administration of developmental services is the renewed role of federal courts in this arena. Suits brought by advocates in federal courts during the 1970s and 1980s played an important role in the closure and downsizing of large state-operated institutions. The decision in *Sullivan v. Zebley* was critical to an expansion of Supplemental Security Income eligibility for children and had a significant impact on increasing the numbers of individuals states were required to serve.

After a decade or more of remaining on the sidelines, federal courts reentered the arena on the so-called "waiting list" issue. First in Florida, and later in Massachusetts, federal courts have agreed with plaintiffs' claims that states were violating their rights by failing to meet their obligations under federal statutes to furnish services covered under its Medicaid plan "with reasonable promptness to all eligible individuals" (*Doe v. Chiles*, 42 USC 1396(a)(8)). The effect on administration of developmental services is that states may no longer limit the entitlement for Medicaid-funded services (representing 35% to 60% of total state expenditures) by claiming a lack of funding for adequate programs and services. By the early 2000s, constructing a waiting list plan sufficient to keep the federal courts at bay became a major issue for virtually every state developmental services administration.

REFERENCES

1. Braddock, D.; Hemp, R.; Parish, S.; Westrich, J. *The State of the States in Developmental Disabilities*, 5th Ed.; American Association on Mental Retardation: Washington, DC, 1998.
2. Prouty, R.; Lakin, K.C. *Residential Services for Persons with Developmental Disabilities: Status and Trends Through 1996*; University of Minnesota, Research and Training Center on Community Living, Institute on Community Integration: Minneapolis, MN, 1997.
3. Boggs, E.; Lakin, K.C.; Clauser, S. Medicaid coverage of residential services. In *An Analysis of Medicaid's Intermediate Care Facility for the Mentally Retarded*

- (ICF/MR) Program; Lakin, K.C., et al., Eds.; University of Minnesota Center for Residential and Community Services: Minneapolis, MN, 1985.
4. Castellani, P.J. *The Political Economy of Developmental Disabilities*; Paul Brookes Publishers: Baltimore, 1987.
 5. Bradley, V.J., Ashbaugh, J.W., Blaney, B.C., Eds.; *Creating Individual Supports for People with Developmental Disabilities: A Mandate for Change at Many Levels*; Paul H. Brookes: Baltimore, 1994.
 6. President's Commission on Mental Retardation. In *The National Reform Agenda and People with Mental Retardation: Putting People First*; U.S. Department of Health and Human Services: Washington, DC, 1994.
 7. Castellani, P.J. Closing institutions in New York state: implementation and management lessons. *J. Policy Anal. Manage.* **1992**, *11*, 593–611.
 8. Castellani, P.J. Managing alternate approaches to the provision and production of public goods: public, private, and non-profit. In *Handbook of Local Government Administration*; Gargan, J., Ed.; Marcel Dekker: New York, 1997.
 9. Kettl, D.F. The myths, realities, and challenges of privatization. In *Revitalizing State and Local Public Service*; Thompson, F.J., Ed.; Jossey-Bass: San Francisco, 1993.
 10. Smith, S.R.; Lipsky, M. *Non-profits for Hire: The Welfare State in the Age of Contracting*; Harvard University Press: Cambridge, MA, 1993.

State and Local Public Pension Fund Management

Jun Peng

*School of Public Administration and Public Policy, University of Arizona,
Tucson, Arizona, U.S.A.*

INTRODUCTION

Public pension funds constitute a very important part of state and local government finance. They are important for at least two reasons. First, they hold an enormous amount of assets. According to the Census Bureau's annual survey of state and local finance, the total value of assets in public pension funds amounted to \$2.2 trillion in fiscal year 2002, covering 17 million members in the public sector.^a Thus the mismanagement of pension assets will affect the well-being of 17 million current and retired public sector employees. Second, pension funds are closely linked with the government operating budgets from which the pension contributions come. Through this connection, mismanagement of pension funds will therefore also affect other aspects of government finance such as the provision of public service. It is therefore critical to have a good understanding of pension management. This article serves to explain the most fundamental aspects of pension management, especially with regard to the measurement of pension fund liability and the main elements of prudent pension management in the public sector.

DEFINED BENEFIT AND DEFINED CONTRIBUTION

The first important thing to understand is the two very different types of retirement programs provided by the government: defined benefit (DB) and defined contribution (DC). In DB plans, government employers guarantee a certain level of retirement benefits to employees when they retire. In DC plans, employers guarantee a certain level of contribution, usually as a percentage of salary, to employees' individual retirement funds while they are working, but are not responsible for their retirement benefits when they retire. Instead, the employees have to be responsible for own future retirement benefits. Although DC has attracted more attention in recent years as many states

set up new DC plans and employees can switch to DC plans from DB plans,^[1] DB remains the predominant plan type among state and local government workers. In 1998, there were 12.98 million members in state and local public DB plans,^[2] accounting for 80% of the 16.15 million members in all state and local public pension plans in that year.^[3] Therefore the remaining of this article is only concerned with management issues related to DB plans.

The next important thing to understand is the funding methods for DB programs. There are two basic ways of funding DB programs: pay-as-you-go (PAYGO) and advance funding. PAYGO means that the current year's benefits to retirees are paid out of current year's contribution from employers and employees. Advance funding means that while an employee is still working, the employer and/or the employee will set aside a certain amount of money every year so that by the time the employee retires, a sufficient amount of funds has been accumulated to pay for all his future benefits. It is quite obvious that at the federal level, social security, although not entirely a retirement program, is funded on a PAYGO basis, as current workers' social security taxes are used to pay for current retirees' social security benefits. The looming social security crisis comes largely as a result of not having enough taxes coming in to pay the retirement benefits in the future. Unlike the federal social security program, state and local pension programs are funded using the advance funding method. With advance funding, the key issue is knowing the amount of funds that needs to be accumulated by the time the employee retires and the amount of funds that needs to be set aside every year so that the required amount will be accumulated by the time he retires. Understanding this issue is the key to understanding almost all other important aspects of public pension management.

ACTUARIAL EVALUATION METHOD

First, how do we know how much to set aside by the time the employee retires? To understand this, we first need to know how the future retirement benefits are calculated. An employee's future benefit is based on a formula like this: final salary $\times$ years of

^aThe data can be found at the Census Bureau's website <http://www.census.gov/govs/www/retire02.html>.

Table 1 Iowa public employees' retirement system funding status

Fiscal year	Actuarial value of assets (AVA)	Actuarial accrued liability (AAL)	Percentage funded	Unfunded actuarial accrued liability (UAAL)
2003	16,120,476,011	17,987,374,960	89.62%	(\$1,866,898,949)

service $\times$ multiplying factor. An example is if an employee's final salary is \$50,000 and has worked for the government for 30 years. Suppose the multiplying factor is 2% of an employee's final salary for each year of service, then his annual pension will be \$30,000. Second, pension sponsor needs to make an assumption about how long he will receive the pension after he retires. Last, because of the time value of money, a pension paid 30 years after he retires is worth a lot less than a pension of the same amount paid the year he retires. The pension sponsor needs to discount and find the present value of all future pension benefits by the time he retires.^b It is this present value of all future pension benefits that becomes the amount to be accumulated.

Second, how should this present value of all future benefits be funded over the working life of the employee? The key principle here is that the funding of benefits should be related to the period in which benefits are earned. There are several actuarial cost methods to distribute the cost over time, such as entry age normal, projected unit credit, and aggregate cost. Of these, the most commonly used is the entry age normal method. Under this method, the present value of future benefits is allocated as a level percentage of the individual's projected compensation between the time he starts working and the time he retires. The purpose of such a method is to spread evenly over time the cost of funding. The part of this total cost allocated to the year in which the service is provided is considered that year's normal cost, and the part allocated to future years is considered future normal cost. Once a person has already worked for sometime, then he has earned (or accrued) some benefits which the pension plan has to pay. To evaluate whether the pension for an employee has been funded sufficiently at any time before he retires, the present value of all his future benefits at the time of evaluation is divided into two parts: the present value of benefits already earned (or accrued) as a result of the service provided prior to the date of evaluation, which is now the pension plan's liability, and the present value of benefits yet to be earned. The first part should be covered by the assets already held in the pension plan and the second part should be covered by the present value of all future normal cost.

^bThe discount rate used for such discounting will be discussed later in this section.

Now we can use these basic concepts to understand the funding status of a pension plan at any particular time. Data in Table 1 were taken from Iowa's Comprehensive Annual Financial Report for FY2003. Such format is a standard way of reporting funding status of a pension plan.

In this table, actuarial accrued liability (AAL) is the difference, as of the actuarial evaluation date, between the present value of all future pension benefits and the present value of all future normal costs. AAL can roughly be interpreted as the present value of future benefits already accrued and thus is the liability of the plan. Then AAL is compared to the actuarial value of asset (AVA) on hand. The difference between the two is called unfunded actuarial accrued liability (UAAL). Dividing AVA by AAL, we also get a funding ratio of the plan. If AVA is greater than AAL, then the plan is overfunded. UAAL will be negative and the funding ratio will be greater than 100%. If AVA is less than AAL, then the plan is underfunded. UAAL will be positive, and the funding ratio will be less than 100%. In Iowa's case, because AVA is less than AAL, the plan is underfunded. It has a UAAL of US\$1.87 billion and a funding ratio of 90%.

To fully understand the actuarial evaluation method, two more elements need to be discussed: actuarial valuation of asset and assumptions underlying the valuation of liabilities.

The actuarial value of asset is different from the actual value of the asset in a pension fund. Inasmuch as most of the assets in a pension funded are invested in the stocks and fixed-income securities, the value of the assets can fluctuate tremendously from year to year because of the fluctuation in the financial market. Simply comparing AAL to the actual value of the asset will result in big swing in the funding status. To reduce such volatility, a smoothing technique is adopted in valuing the asset. Such technique involves calculating a 4- to 5-year moving average of investment return when valuing the asset. For example, if using a 4-year moving average, then only 25% of the investment gains or losses each year over the past 4 years will be recognized in valuing the assets for the current evaluation year. The rationale is to smooth out the ups and downs in the financial market, leading to a more stable investment return over a 4- to 5-year period. Therefore in some years, the actuarial value can be higher than the actual value of asset, whereas in other years, it can be lower than the actual value of the asset.

It can already be inferred to some extent from the previous discussion that the calculation of future benefits involves a lot of assumptions. These assumptions fall into two categories, economic and demographic. Economic assumptions include rate of investment return, inflation rate, and salary growth. Demographic assumptions include mortality rate and disability rate, among others.^c Of all these assumptions, the assumption on the rate of investment return is the most important one by far. The rate of return determines how fast the asset is projected to grow in the future. It is also the rate used to discount the future benefits and future normal costs to the present value. Why should the rate of return be used as a discount rate for calculating the present value of future benefits? That is because allowing the present asset to grow at a certain rate to reach an amount in the future to pay the pension benefits is the same as discounting that future pension payment to the present value using the same rate in order to compare with the value of the asset in the pension today. This assumption is important because different rates can lead to very different present value of future pension liabilities. For the same amount of future pension liabilities, an assumption of a higher rate of return (or discount rate in this case) will lead to a smaller present value of pension liabilities. Because it is the present value, rather than the future value, of pension liabilities that matters in evaluating the funding status of a pension plan, a smaller present value of pension liabilities would leave the plan in a better funding status than otherwise. When evaluating the soundness of a pension plan, one of the key factors a rating agency will look at is the assumed rate of investment return.

Other assumptions are also important in determining the value of future benefits. For example, salary growth projection will determine what a person's final salary will be. Inflation projection will determine what cost of living adjustment for current retirees will be. Mortality rate will determine how long a retiree will receive pension benefits.

PENSION FUND MANAGEMENT

With the basic understanding of the actuarial valuation of pension fund, we can proceed to look at some broader issues concerning pension fund management. The central theme in pension fund management is to make sure that it is sufficiently funded at any time and if UAAL emerges, a plan should be in place to shrink it and eventually eliminate it. There are two basic ways of paying off UAAL: amortization and pension obligation bonds (POBs).

The first and also the most commonly used method, amortization, is the same as paying off a mortgage or car loan. Inasmuch as UAAL is a financial liability of the government, it is no different from other types of debt or loans. To spread out the burden of paying off UAAL, actuarial requirement typically demands that UAAL be amortized over a 20- to 30-year period. Just like an individual who has to pay interest on the mortgage loan he takes out, the government also has to pay interest on UAAL when paying it off, with the interest rate being the assumed rate of return. The reason a government has to pay interest to the pension plan on its UAAL is because UAAL is the present value of future benefits not covered by current assets. Any portion of UAAL that is not paid off will therefore grow at its discount rate, which is the rate of return. Therefore the UAAL next year will be equal to the unpaid portion of UAAL this year plus the growth on the unpaid portion. This growth is the interest on UAAL. UAAL can also be thought of as the amount of asset that should have been in the pension plan at present. Had this amount been in the plan and invested, it would have grown at the assumed rate of return and earned investment income. Therefore the government has to return not only the asset (the amount equal to UAAL) to the pension plan but also the lost investment income (the amount equal to the interest on UAAL).

The second and less common way is to issue POBs to pay off the UAAL at once and then the government will pay debt service on POBs. POBs does not impose a new debt burden on the government, because it essentially turns a debt owed to the pension plan into a debt owed to the investors in POBs. There is only one factor that should determine whether amortization or POBs should be used: interest rate. With amortization, the government has to pay interest at the rate of return to the pension plans. If the government can issue POBs at a rate below the assumed rate of return, then the issuance of POBs results in interest cost savings to the government. While the assumed rate of return is usually based on historical return and does not change frequently, the interest rate on POBs is determined by the prevailing market interest rate and the issuer's credit rating. With the very low interest rate environment in recent years, the rate on POBs has also fallen considerably, resulting in opportunities for issuing POBs. While POBs can lead to interest cost savings, it is not without any risk. For example, if the proceeds of the POBs are invested and do not earn the assumed rate of return, then a new pension liability is created.^d

^cFor a more detailed discussion of assumptions, please see Ref.^[4]

^dFor a more detailed discussion of the pros and cons of POBs, please see Ref.^[5]

While understanding how to pay off UAAL is important, it is more important to understand why it occurs in the first place. Understanding the various factors for causing UAAL will shed light on how to prudently manage public pension fund. As discussed in the "Actuarial Evaluation Method," the calculation of UAAL consists of three factors: government contribution, asset in the pension plan, and the calculation of future pension benefits. The reason for underfunding then has to be found among these three factors.

Annual pension contribution consists of one or two parts. The first part is the normal cost, which is the portion of the future benefits assigned to the year in which it is earned. This should apply to every government with a pension plan. The second part consists of amortization of UAAL if a government has one. If normal cost is not fully contributed every year, then it will result in underfunding. If the amortization portion of the contribution is not fully met, then underfunding will worsen rather than improve. Therefore making the required contribution is one key element in making sure the pension fund is properly funded; if it is already underfunded, then such underfunding shrinks over time.

Once the contribution is made, it becomes the asset of the pension plan and will be invested. Investment management thus becomes the second element in ensuring a pension is properly funded. In their book on investment management process, Trone et al.^[6] list five key steps in a prudently managed investment process. These five steps are analyzing the current position, designing the optimal portfolio, developing an investment policy statement, implementing the investment strategy, and monitoring the performance of your investment portfolio. While all of these are important steps in making sure the investment is properly managed, what is of particular relevance to this article is designing an optimal portfolio. As pension fund investment has to meet the required rate of return, an optimal portfolio means designing an asset allocation

strategy that will allow the pension plans to achieve this rate of return with minimum risk (or least variation from this expected return). Asset allocation refers to dividing pension assets among a variety of investment categories such as equity, bonds, and real estate. For prudent pension management, such asset allocation strategy should be neither very conservative nor very aggressive. A very conservative strategy, such as not allocating a sufficient amount for equity, will make it very difficult for the pension plans to achieve the required rate of return, resulting in funding shortage. A more aggressive strategy, which tends to invest a very large portion in the stock market, may also lead to substantial decrease in pension funding level during a financial market slump.

On the liability side of the equation, the calculation of future pension benefits can change. Such change comes when government officials increase the level of pension benefits, possibly as a result of negotiations with public sector unions. Because pension benefits are not paid right away, it is potentially easier for politicians to promise an increase in pension benefits than an increase in wages, especially when the pension fund is very well funded. Therefore when government officials increase pension benefits, they need to be mindful how that will affect the pension funding status. If an increase in benefits is not matched by an increase in contribution or an increase in investment return, then a funding shortage can occur.

Because of these three factors, a pension plan that is fully funded today does not necessarily mean it will be fully funded tomorrow. Because of the method UAAL is calculated, to maintain its full funding status, the government still needs to contribute the required amount, earns the required rate of return on investment, and should not make changes in benefits without careful analysis of the impact on the pension funding status.

To complete the discussion of public pension funding status, it is useful to see how public pension

Table 2 Public pension funding ratio and asset allocation

	1990	1992	1994	1996	1998	2000	2002
Funding ratio	80.2	82	84.9	88.7	95.2	103.8	96.1
<i>Asset allocation</i>							
Domestic equity	37.8	39.3	40.2	42.0	47.9	45.0	41.1
International equity	2.0	3.7	6.5	10.5	12.0	15.3	13.9
Domestic bonds	44.4	41.4	35.0	34.3	28.6	26.1	33.2
International bonds	0.8	1.8	2.4	2.5	2.0	1.8	1.5
Mortgages and real estate	5.6	6.3	5.4	4.6	4.1	4.8	4.0
Short term	5.8	4.6	4.2	2.9	2.2	2.2	2.2
Other	3.6	2.9	6.3	3.2	3.2	4.8	4.1

funding level has progressed over the years. Every other year since 1990, the Public Pension Coordinating Council (PPCC) has been conducting survey of public pensions. The data on funding level and asset allocation are presented in Table 2.^e

It is obvious from this table that public pension funding has improved substantially since the early 1990s. The funding ratio increased from 80% in 1990 to 104% in 2000, meaning the pension funds overall were overfunded in that year. It is not a mere coincidence that this improvement happened to go hand in hand with an increase in asset allocation for equity, both domestic and international. That allocation increased from under 40% in 1990 to about 60% in the late 1990s. This increase in allocation for equity happened amid one of the greatest stock market booms in history, thus substantially increasing the value of assets in pension funds. It is no wonder that the 6-year stock market slump starting in 2000 also substantially decreased the value of these assets and lead to deterioration in pension funding level in 2002.

LOOKING FORWARD

Looking into the future, public pension funds will assume more prominence in government finance than in the past. As mentioned in the “Introduction,” the annual pension contribution has to come out of the government operating budget. It is therefore in direct competition with other vital public services for the scarce financial resources. Because of the phenomenal investment return in mid to late 1990s, the government could afford to both reduce pension contribution and still maintain a full funding status. The subsequent stock market crash, however, ended such a “holiday” for pension contribution. As can be seen from Table 2, the average pension fund is now underfunded. After years of decreasing pension contributions, many states now are seeing substantial increase in pension contributions, at a time when many states can least afford to do so because of the fiscal difficulty. Because of the underfunding, the government has to pay not only its normal cost but also the amortization of UAAL. The worst has yet to come. Because of the smoothing

technique used in valuing the assets in the pension funds, much of the stock market loss between 2000 and 2002 has yet to be filtered into the valuation of the assets. For example, if using a 4-year smoothing technique, a portion of the loss of 2002 will remain in the valuation until 2005. That means the underfunding situation will continue to deteriorate for several more years before it gets better, and the government pension contribution will continue to increase. To make matters worse, the sustained over 20% investment return in mid to late 1990s in the stock market probably will not come back any time soon and the investment return in the future more likely will only match the historical average return. Therefore, on the one hand, the government tries to make up for the investment loss; on the other hand, it also has to make sure that its future investment will meet the required rate of return. Otherwise, even greater funding gap will occur. With the future investment environment not as favorable as the one in the 1990s, it is all the more important for government to stick to the basic prudent pension management practices as laid out in this article.

REFERENCES

1. Fore, D. Going private in the public sector: the transition from defined benefit to defined contribution pension plans. In *Pensions in the Public Sector*; Olivia, M., Hustead, E.C., Eds.; University of Pennsylvania Press, 2001.
2. U.S. Department of Labor, Bureau of Labor Statistics. In *Employee Benefits in State and Local Governments*; 1998. Available at <http://www.bls.gov/ncs/ebs/sp/ebb10018.pdf>.
3. U.S. Census Bureau available at <http://www.census.gov/govs/retire/ret98t5.xls>.
4. Hustead, E.C. Determining the cost of public pension plans. In *Pensions in the Public Sector*; Olivia, M., Hustead, E.C., Eds.; University of Pennsylvania Press, 2001.
5. Burnham, J.B. Risky business? Evaluating the use of pension obligation bonds. *Gov. Finance Rev.* June 2003, 19 (3), 12–17.
6. Trone, D.B.; Allbright, W.R.; Taylor, P.R. *The Management of Investment Decisions*; McGraw-Hill: New York, 1996.

^eThe data in 1990 through 1998 were collected from PPCC pension survey reports published in *Government Finance Review*. Data in 2000 were available at PPCC's website <http://ppcc.grsnet.com>. Data in 2002 were actually collected by the National Association of State Retirement Administrators and the National Council on Teacher Retirement. The summary of the survey can be found at <http://www.nasra.org/presentations/brainardplenary.pdf>.

State Enterprise Zones

Ryan J. Watson

Middletown, Maryland, U.S.A.

Terry F. Buss

*Center for International Affairs, National Academy of Public Administration,
Washington, District of Columbia, U.S.A.*

INTRODUCTION

President Reagan first introduced Enterprise Zones in the United States in his Urban Jobs and Enterprise Zone Act of 1980. Congressional interest, debate, and legislation concerning EZs continued over the years, but no federal EZ program ever materialized.^[1] But, state officials loved the idea of building political capital by providing unbudgeted financial assistance to underdeveloped communities. By 1995, 2,840 zones had been established in 34 states: “no two states’ enterprise zone programs are alike . . . unlike many other common state incentive programs that have been patterned after prominent federal income tax programs.”^[2] Tax incentives provided the most attractive incentives in recruiting businesses (see [Table 1](#)).

HOW DO STATES JUSTIFY TAX INCENTIVES?

EZ tax incentives have many advocates.^[3] Public officials fear that their state or community will lose business to areas with more favorable business environments. Decision-makers want to rescue failing firms and shield businesses from competition.^[4–6] Drawing in businesses from other areas or facilitating the start up of new firms can boost the economic well-being of a community.^[6–9] Unsure about the intentions of private businesses, states offer tax incentives to hedge their bets.^[5,6] Since most states and communities already offer tax incentives, policymakers simply follow the herd and offer similar incentives.^[10]

Localities frequently perceive incentives as free money. The federal government underwrites some tax incentives and state taxpayers pay for state tax incentive costs, so localities receive free money.^[11] Pervasive government intervention in markets means that tax incentives can go unnoticed,^[12] especially since EZ incentives are generally not part of the state budget process. Tax incentives are viewed as a net benefit since the initial loss of tax revenue is eventually recaptured directly or indirectly-through taxes and growth. Market failures in state economies can be corrected

by judiciously using incentives.^[13,14] But in reality, politicians mostly use EZ programs to appease constituencies or advocates for distressed places.

WHAT IS THE STATUS OF EZ PROGRAM EVALUATION?

States rarely perform evaluations to determine the effectiveness of EZ programs or to answer economic development questions.^[15–17] A National Association of State Development Agencies survey of states concluded that states had not conducted rigorous cost-benefit evaluations of the incentives they offered. Likewise, the Council of State Governments’ annual incentive survey found that only a few states used a formal cost-benefit model to gauge the impacts of tax and financial incentives.^[18] The National Conference of State Legislators concluded: “Few states know the exact amount they spend on economic development initiatives.”^[19] A survey conducted by Lohman, et al.^[2] found that while zone administrators believed that EZ incentives increased jobs and investment in zones, few states could offer adequate statistics to measure the effectiveness of these programs. The disregard for cost-benefit analyses is a significant factor in making EZ programs ineffective.

In his literature review, James^[20] concluded that there has been no thoroughly acceptable evaluation of the impacts and cost-effectiveness of an EZ program (see also Ref.^[21]). Nonetheless, existing studies offer important insights.

ARE EZ TAX INCENTIVES GOOD ECONOMICS?

EZs are largely unassailable in a political sense. Republicans favor them because they promote economic development using market-based incentives rather than entitlement programs, and Democrats support EZs since they represent a major source of social welfare funding for disadvantaged communities. EZ tax incentives are also good politics because policymakers

Table 1 Examples of state EZ incentives

- State income tax deductions, credits, or exemptions
- Local income tax reductions, credits, or exemptions
- Infrastructure improvement
- Property tax breaks
- Unemployment tax credits
- Sales tax refunds, credits, or exemptions
- Regulatory relief
- Job training assistance and subsidies
- Deduction for net interest income coming from loans to qualified zone firms
- Credit for wages paid to “disadvantaged individuals”
- Low interest loans
- Day care tax credits
- Business development assistance
- Contributions to enterprise zone associations
- Permission to carry forward 100 percent net operating loss for many years
- Reimbursement of insurance costs
- Expensing all or a portion of qualified property

Sources: Refs.^[2,23,24]

can redirect resources to stimulate economic growth and development in the poorest regions of their states, which theoretically increases long-term state tax revenues.

All political factions use tax codes to benefit certain projects or sectors. Businesses are strong advocates of EZ tax incentives, and the taxpayers who bear the financial burden of these incentives are largely unaware of their existence. EZ failure can be blamed on business cycles, market forces, dysfunctional business behavior, or in extreme cases on poor people in distressed places (e.g., high crime). When the economy is booming, policymakers can easily claim political dividends by saying that EZs contributed to the positive economic climate. These claims may be untrue, untested, or exaggerated, but citizens have no reliable way of evaluating these statements.

Setting aside politics, it is worth considering whether EZs are good economics. Economists offer various criteria that can justify public investment in EZs. In practice, the criteria outlined in the tax literature appear to be frequently ignored. Courant^[22] suggests that if zone policies address market failures such as the inefficient utilization of resources or overly localized economies, zone incentives can increase economic efficiency. If investment is relocated from local labor markets with low unemployment to ones with higher unemployment, incentives may generate efficiency gains as underutilized resources are mobilized. Efficiency gains may also result if reductions in

unemployment produce positive externalities such as reductions in welfare dependence or crime. Bartik^[14] argues that net capital investment may be somewhat irrelevant, because redistribution, even within a state, may be an end in itself. Investment, through tax incentives, may allow individuals to build job skills and experience that enhance their long run employability.

There is a large body of research that says EZs are not effective economics. Peters and Fisher^[23] examined 75 EZs in 13 states by using a “hypothetical firm” computer simulation. They applied tax incentives to financial statements that they generated for hypothetical firms. Peters and Fisher came to largely negative conclusions about how well EZs work, as well as finding that the goals and policies of EZ programs are often confusing. Knowing that housing value growth rates are a quantifiable measure of EZ programs’ effectiveness, Engberg and Greenbaum^[24] conclude that EZs do not usually increase the growth rate in property values. In fact, zones that had high vacancy rates before being designated an EZ actually experienced a decline in the property value growth rate. Netzer^[25] concludes that economic development tax incentives including those in EZs are, for the most part, neither good nor bad from the standpoint of economic efficiency. With imperfections in economies due to government interference at all levels (see also Ref.^[12]), offering incentives does not parallel the efficiency operations of private competitive markets.

Research generally fails to measure how much market failures justify tax incentives. Skeptics believe that most tax incentives are poor economics.^[26] Advocates believe that tax incentives work or do little harm, and that diverting economic development funding to assess them is a waste of money.^[12] The majority of the research supports the skeptical view.

DO EZs STIMULATE GROWTH?

Some studies attribute economic growth to EZs, but more studies convincingly find negative or inconclusive results (e.g., Ref.^[1]). Numerous studies of the same zones come to opposite conclusions, while some researchers reverse their positive conclusions in follow-up studies. Many correlations between taxes and economic growth are methodological artifacts (see Refs.^[26]).

Multi-State Studies

Several researchers have undertaken multi-state EZ studies, the most extensive and rigorous among them are Peters and Fisher^[27] and Greenbaum and Engberg.^[28] Peters and Fisher studied manufacturing

employment in 75 zones found in 13 states, comparing zone incentives and non-zone incentives to businesses. They found that during the 1990s, non-zone incentives grew in importance relative to EZ incentives, effectively lessening the potential impacts of zones. They concluded "EZ incentives tend to favor capital-intensive over labor-intensive industries; usually cause losses to public coffers and have very little impact on new investment; and do little to improve the job prospects of residents in the zones."

Greenbaum and Engberg^[28] studied zones in California, Florida, New Jersey, New York, Pennsylvania and Virginia. Results showed that "the zone incentives do not significantly improve housing market, income, or employment outcomes either in the zones themselves or in adjacent neighborhoods."^[28] They hypothesize that the lack of impact may be attributable to 1) zone programs crowding out other beneficial policies; 2) zone incentives may subsidize businesses that lack long-term viability; or 3) zone designation may carry a stigma.

Indiana Zones

Papke^[29] initially found mostly positive impacts for Indiana EZs. Indiana EZs permanently increased inventory values to 8%, or \$3.2 million, more than would have been the case without the zone. Papke^[29] also found that unemployment claims declined by 19%, or 1,500 people, for the local labor market following zone designation. However, value of machinery and equipment declined by 13%, or \$5 million.

Papke^[30] subsequently reversed her positive conclusions, saying that EZs lost an average of about 2,300 people, or 674 households. Per capita income fell 2%. Zone unemployment only dropped from 9.3% to 8%. The percentage of people working in the zone as a place of residence rose only 4%. Firms created 2,897 new jobs in Indiana zones in 1988, with 15% going to EZ residents. Zone residents earned about half as much as non-resident EZ employees. Annual cost of an Indiana EZ job was \$4,564, or \$31,113 per zone resident. Zones lost more population and experienced a decline in per capita income compared to income growth in non-zones, while unemployment only fell slightly in zones, and fewer zone residents actually worked in zones.

New Jersey Zones

Urbanomics conducted a cost-benefit study of businesses in 11 New Jersey EZs (see also Ref.^[31]). Analysts defined benefits as jobs, payroll and production, both directly and through multipliers. The EZ program cost taxpayers \$51.6 million, but induced \$106.5 million in

subsequent tax revenues, totaling \$267.4 million with multipliers. The direct tax benefit-to-cost ratio was 2.2:1, and the total tax benefit-to-cost ratio was 5.2:1. Cities that had grown slowly before designation grew faster than comparison cities once the zone was in place.

Boarnet and Bogart,^[32] using an econometric model with a panel study design, tested Rubin and Trawinski's conclusions^[31] by studying 28 municipalities eligible for zone designation and 7 municipalities that actually became EZs in 1984. Boarnet and Bogart "found no evidence that the urban enterprise zone program in New Jersey had a positive effect on total municipal employment, on sectoral employment or on municipal property values."^[32]

Illinois Zones

Sridhar^[33] studied 49 Illinois zones. Sridhar found that economic development incentives, on average, yielded net benefits 10 times higher than tax abatement costs. For new jobs created, earnings were highest in low-unemployment zones and lowest in high-unemployment zones. But when earnings were made net of reservation wages, which are the wages necessary to induce people into the labor force, net benefits and benefits per job were greatest in high-unemployment zones. While net benefits and benefits for each job relocated were highest in high-unemployment areas, they were negative in low-unemployment areas. Therefore, positive net social benefits would exist if all jobs were redistributed from average-unemployment areas to EZs.

McDonald^[34] argued that Sridhar's work was invalid because it assumed that EZ programs were responsible for all jobs created in a zone. Lambert and Coomes^[35] studied one Louisville EZ where more than \$560 million in federal and state funding went towards the expansion of Louisville's airport, causing United Parcel Service (UPS) to increase its operations within the EZ. When evaluating Louisville's EZ, Lambert and Coomes write, "Most of the EZ benefits claimed by the proponents are linked to the airport expansion. UPS and other airport-related firms did take advantage of EZ tax advantages; however, it would be far-fetched to claim that the EZ program caused the local economic benefits."^[35]

California Zones

Dowall^[36] analyzed 13 California zones. Shift-share analysis attributed virtually all employment growth from 1986 to 1990 in zones to countywide growth and industrial mix. Zones actually retarded employment by 5,578 jobs.

Maryland Zones

The General Accounting Office (GAO)^[37] studied three Maryland EZs from 1980 to 1987, finding no effect on employment or welfare dependence of workers. Job creation in the zones did not affect costs incurred in subsidizing firms. GAO reviewed comparative case studies, state program evaluations, and broad reviews of program results, concluding that studies either had flaws or failed to focus on program-related changes in employment. Flaws included: 1) reliance on data of unknown or dubious quality such as surveys of zone administrators; 2) measurement of program effects too soon after zone designation; and 3) failure to use an adequate baseline for attributing observed effects to possible influences.

Ohio Zones

Hill^[38] studied Ohio's Amended Substitute Senate Bill 19, which reformed the state's EZ program. Hill examined numerous publications and key informant interviews to assess Ohio's EZ program. He concluded that the program is an ad hoc business tax reduction, not true tax reform. The program puts most distressed areas of the state at a competitive disadvantage. Even though the program intends to respond to interstate competition, its major impact was to stimulate intra-state tax competition. The program's exaggerated benefits did not result in net new job formation.

Rural Zones

Few studies examine rural EZ effectiveness, and findings conflict (see Refs.^[39–41]). Louisiana zones were declared ineffective.^[42] Robinson and Reeder^[40] conducted the best research on rural EZs. They reanalyzed the rural zone data in the 1996 HUD survey of zone administrators. Rural zones created or saved 2 jobs per 100-zone residents as compared with 3 jobs per 100 residents in metro zones. Only one of seven zones reported saving any jobs. Rural zone development came almost entirely from expansions within the zone. Most were manufacturing companies. Three-fifths of jobs created went to zone residents and one-half went to unemployed people. Weaknesses in the Robinson and Reeder^[40] study include reliance on zone administrator surveys and an inability to determine whether jobs would have been created in the absence of zone incentives. Using the number of jobs saved to measure success is also problematic.

Urban areas often dominate funding for EZ programs. In Florida, 83% of EZ tax incentives went to just three urban zones, while less than one percent of the EZ funding went to the state's twenty rural

EZs.^[43] Rural and small-town EZs also experience difficulty in the implementation and administration of EZ incentives. While Florida's Miami-Dade County spent approximately \$325,000 for the administration of its EZ program in the 1998-1999 fiscal year, many coastal communities within the state put less than \$1,500 each towards EZ management in the same year.^[43]

Federal Studies

In the 101st Congress, legislators proposed six EZ bills. Bills provided labor subsidies to reduce wage costs by almost 14% if half of total wages were devoted to economically disadvantaged workers. Some firms might increase EZ investment between 8.5% and 20%, employment between 6.5% to 14%, and output between 7.5% and 15%. But tax-induced increases in investment, employment, and output in EZs would largely be offset by declines elsewhere.^[44] EZ investments could earn up to 39% lower pre-tax rates of return than are available from alternative investments, representing a reduction in national income. HUD researchers concluded that job creation could not be attributed to EZs; other factors were responsible.^[1]

ARE EZ TAX INCENTIVES WASTED?

Literature is divided on whether investment would have occurred in the absence of an EZ, although the best research suggests that it would have. This literature relies heavily on surveys of firms and administrators, as well as administrative data.

Surveys of Firms and Zone Administrators

Dowall's^[36] study of California EZs showed one-half of eligible zone firms did not use incentives. Another fourth tried to access incentives but were unable to do so. 67% of businesses using zone programs stated that traditional factors such as markets, transportation, and worker availability influenced their location or expansion decisions, not EZ incentives. Importantly, many businesses did not discover tax incentives until after making their decisions. Elder and Cohen^[45] had identical results in their Illinois zone study. 77% of Louisville's EZ firms that used the most expensive and common EZ benefits made their decisions based on factors other than EZ incentives.^[35]

GAO's Maryland study^[37] used interviews with several hundred employers to explore the importance of locational factors. GAO presaged Dowall's^[37] findings above. At least 60% rated market access, community characteristics, site characteristics and government cooperation as important in their location decision.

Transportation, real estate, financial health of region and quality of life were rated as important by at least 50% of employers. Financial incentives were important to only 14%.

Rubin and Trawinski's^[31] study of New Jersey zones, by contrast, revealed greater impact. When companies were asked about location decisions, 32% reported that EZ benefits were the only or primary reason for their decision to expand or locate within zones, 38% suggested that EZ benefits provided a secondary reason to consider the zones, and 30% felt that EZ benefits had no impact.

HUD's interviews with zone administrators found that property and corporate income taxes were ranked first in importance by two-fifths of respondents. This contrasts with GAO's conclusions (as reported in Ref.^[1]).

Firms differ in their consideration of EZ incentives when making decisions. Firms may misrepresent their motivations for using incentives: some may not need them, but still justify why they used them; some may not need them, but want to show support for zone administrators or the economic development department; and some may truly need them. Skeptics are not convinced that researchers have sorted out these motivations.^[46]

Administrative Record Studies

Dowall's^[36] report on the results of the California State Franchise Tax Board tax credit study concluded that incentives are not important. Board records showed that 1,719 jobs were created by granting tax credits to zone businesses, a figure representing only 5% of the jobs reportedly created in zones by zone administrators. EZ employment from tax credits contributed only 6% of total net increase in zone employment between 1986 and 1990. Therefore, EZ administrators may exaggerate job creation and retention in order to bolster program support. California EZ businesses claimed \$10.6 million in tax credits between 1986 and 1990, while the total net taxable income of zone firms was \$2.7 billion: tax credits address a small part of business costs.

DO EZS EXACERBATE TAX WARS AMONG STATES?

Disadvantaged communities offer incentives to attract firms to EZs, but these have a limited effect on interstate competition. Few firms relocate. Effective tax rates overall created competition mostly within rather than across states.^[47] The average effective tax rate in cities outside zones was 9.1% as compared to 7.3%

inside zones. Average state and local tax burden on new investment, then, was only 19% lower. Considerable variation occurred across cities. Outside zones in low-tax cities, the effective tax rate was 5.6% and 14.7% in high-tax cities. Inside zones, rates dropped to 3.7% to 13.2%, respectively.

Fisher and Peters,^[47] using their hypothetical firm approach, offer the best evidence about EZs. Cities with EZs usually offered more generous incentives than cities without zones, but "enterprise zones are little more than geographically targeted versions of standard state and local economic development programs."^[47] However, state corporate income tax credits were more prevalent and more generous in EZs than statewide, and policymakers employed jobs credits twice as often as investment credits within EZs (see also, Ref.^[48]). Zones tended to offer incentives worth two to three times those available statewide.^[47] Within EZs, zone incentives accounted for one-third of the incentives offered.

Most jobs in EZs come from expansions of existing zone businesses and from new businesses, rather than relocations.^[21,49] In a study of 357 zones in 17 states, 26% of 1,623 zone firms making new investments were new firms. 8% were new branch plants of non-zone firms and 9% were relocations from outside the zone. "If EZs had been established to nurture new business development, the incentive programs would have been structured quite differently and an emphasis would have been placed on small business incubators, technology transfer programs, management assistance and venture capital provision."^[47]

HOW WELL DO EZ INCENTIVES WORK?

Literature suggests that targeted populations and businesses do not necessarily benefit from EZs. Frequently, incentives negatively impact zones.

Incentive Beneficiaries

Job tax credit programs may not induce jobs, because many firms cannot fully use them. Most job tax credits have a statutory ceiling and are one-time credits.^[47] For 13 states with job tax credits in Fisher and Peters' study, firms in 6 states used 100% of credits available. But in the other states, firms used only 67% of the credit on average. Job tax credits did not induce employment as expected. Job tax credits may have no price effect at the margin, meaning that all other costs being equal in other locations, these taxes may have little impact on firm costs of business.

People who ought to benefit from EZs may not. Shocks or subsidies in labor markets only temporarily

reduced unemployment in targeted areas. In-migration of workers to a targeted area quickly eliminated benefits to original residents. In the long run, new in-migrants to a targeted area took many newly created jobs.^[50] Bartik,^[51] in reviewing 18 studies, calculated that 60% to 90% of jobs created by employment programs go, in the long-run, to in-migrants rather than the targeted beneficiaries. Ladd (Ref.^[52] see also Ref.^[53]), in her literature review of EZs, surmised that place-based subsidies and even people-based, place-based subsidies failed to raise employment levels and well being of targeted residents.

Minority businesses were not attracted to EZs. Glover^[54] surveyed zone firms and administrators, finding that only 5% of firms were minority-owned.

Capital Investment Incentives

Not all incentives increase employment; some actually decrease it.^[22,30,48] Nationally, wide variation exists in the mix of labor and capital incentives inside and outside zones, ranging from 0% to 100% focus on labor.^[47] Typical capital incentives clearly have much larger effects on the price of capital goods than average labor incentives have on wages. The exemption of capital equipment from a 6% sales tax reduces its acquisition cost by 4% to 4.5%. A 1% state investment tax credit reduces the price about 0.65%. Average jobs credit per job, represents only 0.2% to 0.5% of present value of wages over 10 years. Maximum jobs credit represents from 1% to 1.6% of present value of wages over this period. Incentives that lower the price of capital goods have both an output effect (production and employment increase because costs are lowered) and a substitution effect (capital is substituted for labor). If the substitution effect is stronger, a capital incentive could reduce employment. Loh^[55] looked at county-level growth in Ohio in the 1980s as a function of economic incentives, concluding that capital subsidies to private businesses were more effective than either labor subsidies to private businesses or capital subsidies to communities. However, Loh's work has been roundly criticized.^[47]

EZ Size

If every community is an EZ, the program breaks down. For example, an existing EZ in Jefferson County, Kentucky was expanded due to political pressures. The expanded EZ included new zone areas that did not fulfill the minimum socioeconomic requirements put forth in Kentucky's EZ legislation, but the overall zone's economic vital signs were still low enough to qualify.^[35] Arkansas and Kansas have EZs

that cover the whole state.^[2] In 1995, South Carolina designated 90% of the state an EZ. Some states designated entire counties. Critics suggest that this misconstrues the intent of geographically small area targeting. Small area targets do not significantly affect economies, either positively or negatively, so partial equilibrium models apply. When zones get too large, full equilibrium models apply. Full equilibrium suggests that transfers of wealth from one area to another reduces wealth in the entire economy (see Ref.^[48]).

REFERENCES

1. Birdsong, B. *Federal Zones*; Urban Institute: Washington, DC, 1989.
2. Lohman, R. et al. State enterprise zone programs: a survey of the benefits (part I). *J. Multistate Tax. Incent.* June 2002, 12 (3).
3. Key, K.; Smith, J. Trends in state and local economic development incentives. *J. State Tax.* 1996, 15, 1–13.
4. Burnier, D. Becoming competitive: how policy makers view incentive-based development policy. *Econ. Dev. Q.* February 1992, 6, 1.
5. Noto, N. Trying to understand the economic development official's dilemma. In *Competition Among States and Local Government*; Urban Institute: Washington, DC, 1991.
6. Wolman, H. Local economic development policy. *J. Urban Aff.* 1988, 10, 19–28.
7. Ledebur, L.C.; Woodward, D. Adding a stick to the carrot. *Econ. Dev. Q.* 1990, 4, 15–25.
8. Spindler, C.J.; Forrester, J.P. Economic development policy. *Urban Aff. Q.* 1993, 29, 28–53.
9. Clingmayer, J.; Feiock, R. Distribution and redistribution in economic development. *J. Polit.* 1995, 57, 508–520.
10. Gilbert, J. Selling the city without selling out. *Urban Lawyer* 1995, 27, 1–13.
11. Watson, D.J. *The New Civil War*; Praeger: Westport, CT, 1995.
12. Wiewel, W. Response. *Econ. Dev. Q.* 1999, 13, 357–364.
13. Baum, D.N. The economic effects of state and local business incentives. *Land Econ.* 1987, 63, 348–360.
14. Bartik, T.J. *Who Benefits*; Upjohn Institute: Kalamazoo, MI, 1991.
15. Buss, T.F.; Yancer, L.C. Cost-benefit analysis: a normative perspective. *Econ. Dev. Q.* 1999, 13, 29–37.
16. Hartzheim, L.A. State tax incentives. *J. State Tax.* 1997, 15, 51–64.
17. Bartik, T.J. Better evaluation is needed. *Econ. Dev. Q.* 1994, 8, 99–106.
18. Chi, K.S.; Leatherby, D. *State Business Incentives*; Council of State Governments: Lexington, KY, 1997; 8 pp.
19. National Conference of State Legislatures. *A Review of State Economic Policy*; NCSL: Washington, DC, 1997; 1 pp.

20. James, F.J. The evaluation of enterprise zone programs. In *Enterprise Zones*; Green, R., Ed.; Sage: Newbury Park, CA, 1991; 225–257.
21. Wilder, M.; Rubin, B. Rhetoric versus reality. *J. Am. Plan. Assoc.* **1996**, *62* pp. 4, 473–91.
22. Courant, P. How would you know a good economic development program if you tripped over one? *Nat. Tax J.* **1994**, *47* (4), 863–891.
23. Peters, A.H.; Fisher, P.S. *The Effectiveness of State Enterprise Zones*; W.E. Upjohn Institute for Employment Research: Kalamazoo, MI, 2002a.
24. Engberg, J.; Greenbaum, R. State enterprise zones and local housing markets. *J. Housing Res.* **1999**, *10* (2), 163–187.
25. Netzer, D. An evaluation of inter-jurisdictional competition. In *Competition Among States and Local Governments*; Kenyon, D., Kincaid, J., Eds.; Urban Institute: Washington, DC, 1991; 110–132.
26. Buss, T.F. The effect of state tax incentives on economic growth and firm location decisions: an overview of the literature. *Econ. Dev. Q.* **2001**, *15*, 90–105.
27. Peters, A.H.; Fisher, P.S. *State Enterprise Zone Programs: Have They Worked?* W.E. Upjohn Institute for Employment Research: Kalamazoo, MI, 2002.
28. Greenbaum, R.; Engberg, J. An evaluation of state ez policies. In *Public Policies for Distressed Communities Revisited*; Redburn, F.S., Buss, T.F., Eds.; Lexington Books: Lanham, MD, 2002; 33–64.
29. Papke, L.E. *Tax Policy and Economic Development*; Working Paper, Boston University, 1991.
30. Papke, L.E. *What Do We Know About Enterprise Zones?* National Bureau of Economic Research: Cambridge, MA, 1993. Working Paper #4251.
31. Rubin, M.; Trawinski, E. New Jersey's urban enterprise zones. *Urban Lawyer* **1991**, *23*, 461–471.
32. Boarnet, M.G.; Bogart, W.T. Enterprise zones and employment. *J. Urban Econ.* **1996**, *40*, 198–215.
33. Sridhar, K.S. Tax costs and employment benefits of enterprise zones. *Econ. Dev. Q.* **1996**, *10*, 69–90.
34. McDonald, J.F. Comment. *Econ. Dev. Q.* **1997**, *11*, 222–224.
35. Lambert, T.E.; Coomes, P.A. An evaluation of the effectiveness of Louisville's enterprise zone. *Econ. Dev. Q.* **2001**, *15*, 168–180.
36. Dowall, D.E. An evaluation of California's enterprise zone programs. *Econ. Dev. Q.* **1996**, *10*, 352–368.
37. General Accounting Office. *Enterprise Zones: Lessons From the Maryland Experience*; GAO: Washington, DC, 1988.
38. Hill, E.W. *Tax Abatement: War Within a State*; 1994. Unpublished manuscript.
39. U.S. department of housing and urban development. *State-Designated Enterprise Zones*; HUD: Washington, DC, 1986.
40. Robinson, K.L.; Reeder, R.J. State enterprise zones. *Rural Dev. Perspect.* **1991**, *7*, 30–35.
41. Wortman, M.S. Impact of entrepreneurship upon rural development. In *Rural Development Research*; Rowley, T.D. et al., Eds.; Greenwood Press: Westport, CT, 1996.
42. Nelson, A.C.; Whelen, R.W. Do enterprise zones make a difference? In *A Paper Presented to the Urban Affairs Association*; March 9–12, 1989.
43. Perez, E. Flaws seen in enterprise zones' focus. *Wall Street J.* **April 19, 2000**.
44. Zimmerman, D. *Federal Tax Incentives for Enterprise Zones*; Congressional Research Service, U.S. Library of Congress, June 15, 1989; 89–371E.
45. Elder, A.; Cohen, I. *Evaluation of Effectiveness and Efficiency of Enterprise Zones in Illionois*; Illinois State University, 1988.
46. Netzer, R. Comment. In *New England Economic Review*; Federal Reserve Bank of Boston: New England, Mar/Apr 1997; 131–137.
47. Fisher, P.S.; Peters, A.H. *Tax and Spending Incentives and Enterprise Zones*; Federal Reserve Bank of Boston: New England, Mar/Apr 1997; 109–137.
48. Gravelle, J.G. *Enterprise Zones*; Congressional Research Service, Library of Congress, June 3, 1992. CRS Report 92–476S.
49. Erickson, R.A.; Freidman, S.W. *Enterprise Zones*; Center for Regional Business Analysis, Penn State University, 1990.
50. Blanchard, O.; Katz, L. Regional evolutions. *Brookings Pap. Econ. Act.* **1992**, *1*, 1–75.
51. Bartik, T.J. Who benefits from local job growth. *Reg. Stud.* **1993**, *27*, 293–311.
52. Ladd, H.F. Spatially targeted economic development strategies. *Cityscape* **1994**, *1*, 193–218.
53. Erickson, R.A.; Syms, P.M. The effects of enterprise zones on local property values. *Reg. Stud.* **1986**, *20*, 1–14.
54. Glover, G. Enterprise zones: incentives are not attracting to spur urban economic development. *Rev. Black Polit. Econ.* **1993**, *22*, 73–79.
55. Loh, E.S. The effects of jobs-targeted development incentive programs. *Growth and Change* **1993**, *24* (4), 365–383.

State Government Administration in a Federal System

Timothy D. Mead

*Department of Political Science, University of North Carolina at Charlotte,
Charlotte, North Carolina, U.S.A.*

INTRODUCTION

American federalism has never been static. Yet, each ebb and flow of the relative weight of the national government and the states in policy making has left the national government stronger than it was before. The rise of a national economy, the increasing technology of warfare and national defense, and the development of policy problems only subject to resolution on a large scale have all contributed to the secular increase of power in the hands of the national government.

SIGNIFICANCE OF STATES IN THE U.S. FEDERAL SYSTEM

By definition, a federal system is a method of governance with multiple levels of government. Often the argument is made that the particularly successful innovation in governance struck off at the Constitutional Convention was a federal structure. Yet, as a political matter, the Founding Fathers had no choice other than to recognize the existence of the states. The colonies, which became states, were already in place and, during the period of the Confederation, had substantial independence from one another. In 1787, a federal system by one form or another was a *sine qua non* of a national government.

This is not to say that the role of the states did not get attention at the Constitutional Convention. On July 17, 1787, the Virginia Plan, advocated by the largest and most powerful of the states was debated. It would have permitted Congress “To negative all laws passed by the several States contravening in the opinion of the Nat: Legislature the articles of Union.”^[1] Only two states—Massachusetts and Virginia—voted for this provision.^[1] Further, the Constitution as ultimately adopted over 200 years ago guarantees the integrity of the states by providing that no state shall be divided without its approval.

While the provision that Congress could veto state laws did not pass, another provision, the Supremacy Clause, made the states subject to national law. Indeed, a case could be made that the Supremacy Clause made state laws subject to veto by the judiciary rather than the legislature. Early in our constitutional history, the

capacity of the national government to establish its laws within the states was challenged by the Kentucky and Virginia Resolves, which would have permitted the states to determine whether Congress had exceeded its powers. Ultimately, a “great war” was fought to resolve the relative balance of the national and state governments.

Once states are admitted to “this Union” by Congress, they stand on an equal footing vis-à-vis all other states. As a constitutional provision, this has meant that there have been no “second class” states, thus assuring that territories would seek statehood rather than independence.

THE RISE OF A FEDERAL POLICE POWER

The Tenth Amendment to the Constitution appears to be a significant grant of power to the states. It provides, “The powers not delegated to the United States by the Constitution, nor prohibited by it to the States, are reserved to the States respectively, or to the people.” Popular discussion of the Tenth Amendment often asserts that it reads “expressly delegated,” yet both houses of Congress rejected the insertion of “expressly.” As the Supreme Court noted in *United States v. Sprague*,^[2] the Tenth Amendment “added nothing to the instrument as originally ratified.”

Yet, as a written document, the Constitution created a government of limited powers, that is powers that are granted to the national level of the federal system. A general power “to regulate the public health, welfare, morals and convenience,” often called the police power, was not included in the grant of powers to the national government. Roughly, from the death of Chief Justice John Marshall in 1835 to the New Deal, the absence of a federal police power was often argued as the basis for the exercise of various state police powers.

Dual federalism, the notion that the Constitution created two separate, coequal sovereignties, was the dominant constitutional argument. National and state governments were each perceived to be supreme within their individual spheres. The powers granted to one level were limits on the other. *Hammer v. Dagenhart*^[3] represented the legal highwater mark of dual federalism when the Supreme Court struck down a

congressional enactment banning the interstate shipment of goods produced by child labor. Dual federalism was specifically rejected by the Supreme Court in *U.S. v. Darby*,^[4] when *Hammer v. Dagenhart* was overruled and the Court accepted the extension of national labor laws to businesses that had an impact on interstate commerce. In *Darby*, Justice Stone asserted that the Tenth Amendment was a “truism that all is retained which has not been surrendered.”

Over time, however, a federal police power has developed, seriously eroding the powers of the states in the federal system. Several streams of constitutional development have contributed to the rise of a federal police power and a concomitant diminution of the discretion of the states.

Among those streams is the interpretation of the “necessary and proper” clause^[5] to mean “convenient, or useful, or essential” to carrying out one of the enumerated powers.^[6] Thus, the national government may exercise a power not in the list of powers granted that may be implied from that list. The significance of this interpretation in the development of the federal system cannot be overstated.

Another stream in the development of a federal police power is the expansion of the power “to regulate commerce . . . among the several States.”^[7] In a landmark decision, Chief Justice Marshall held that the power to regulate commerce, to “establish the rule,” was not limited to the transfer of goods across state boundaries but included the entire transaction.^[8] This decision created the legal foundation for a national economy. As the Industrial Revolution transformed the American economy, the grounds for national, as contrasted to state, administration of the economy was laid. National authority over the economy rapidly led during the 1930s to national regulation of a variety of activities previously administered by states alone.

Further, the power to tax and spend^[9] has served as a tool in the development of a national police power. The power to tax, for example, has led to a host of special provisions in the tax code to encourage individuals and businesses to engage in practices deemed to be in the public interest; some call these provisions loopholes, while others find them appropriate incentives. Examples include tax deductions for home mortgages, credits for funding research, and a host of others. The power to spend has permitted Congress to enact provisions of appropriations designed to encourage—some would say require—recipients of national government funds to modify personnel practices, establish speed limits on highways, establish minimum age requirements for the consumption of alcohol, and permit right turns on red-traffic-signal-controlled intersections.

Critical, from the perspective of the rise of a federal police power as a constraint on state administration, was the adoption of the Fourteenth Amendment.

It provides, in part, “nor shall any State deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws.” While early interpretations of the Fourteenth Amendment were restrictive of its application,^[10] subsequent decisions upheld national regulation of the purity of foods and drugs, the hours and wages of labor, access to the courts, and a panoply of services that the Founding Fathers could not have anticipated.

While there are other streams that contribute to the development of a national police power, these cited are most critical from the perspective of state administration. And the development of a national police power is significant from the perspective of state administration because that combination of powers has shifted the balance of policy-making responsibility, and the parallel administrative responsibility, from the states to the national government. The upshot of the development of a national police power has been that in many ways, the states have become administrative agents of national policy.

Mandates, national requirements that states do something in particular or do things in particular ways, spring from the national police power. Mandates may take the form of direct orders, for example, to observe the requirements of the Occupational Safety and Health Act protecting workers. Mandates may be crosscutting regulations that affect all, or virtually all, national programs, for example, requirements of Environmental Impact Statements before national funds can be spent. Mandates may be crossover sanctions requiring that states do something before national funds may be spent on something else, for example, the various highway safety rules that states must adopt before they may spend national highway construction funds; such mandates, of course, are only apparently optional. Or mandates may be preemptive, in which the national government adopts a standard, for example, for the permitted amount of arsenic in drinking water but allows states to adopt a more rigorous standard.

In one fashion or another, there are hundreds of such national mandates. Those most onerous from the perspective of states are the unfunded mandates, national requirements that states do one thing or another that will require the expenditure of funds by the states without reimbursement by the national government.

NATIONAL ADMINISTRATION AS A MODEL FOR STATE ADMINISTRATION

National administrative practices have often served as models for state practices. Early in our national

history, administration of all government programs was done by patronage; a variety of rationales were offered to support such administration.

Multiple forces came together to replace administration by patronage with administration by a career civil service. In 1883, the national government adopted the Pendleton Civil Service Act, which created a career merit system for employment by the national government. Following the model of the national government, roughly three-quarters of the states have adopted requirements of a career civil service. In the balance of the states, programs supported by federal funds are required to hire and fire according to standards of a career service. Partisan loyalty as a requirement of state employment was rejected by the Supreme Court in *Rutan et al. v. Republican Party of Illinois*.^[11]

Another example of the national government as a pacesetter for state administrative practices is the 1946 Administrative Practices Act. This act established standard procedures for federal agencies, including guidelines for rule making, public hearings, judicial review of administrative decisions, and a host of other practices that are now taken for granted as appropriate administrative practice. In more recent times, the Privacy Act of 1974 and the Freedom of Information Act of 1966 have been amendments to the national Administrative Practices Act. Following the lead of the national government, all states have adopted an administrative practices act, which includes various elements of the national legislation.

Further, as the national government has adopted one or another of the “hot” practices of public administration, Program-Planning-Budgeting-System in the 1960s, Zero-Based Budgeting in the 1970s, Management by Objectives in the 1980s, and Reinventing Government in the 1990s, many states have followed suit.

RECENT HISTORY OF ADMINISTRATIVE FEDERALISM

Since World War II, every presidential candidate save John Kennedy in 1960 and Lyndon Johnson in 1964— independent of political party—has made the case that the national government was too intrusive and that states should be strengthened. The irony, of course, is that the long-term trend toward greater national influence has continued. As a current example, President George W. Bush has urged greater state responsibility for the conduct of primary and secondary education and a series of national tests to assure the accountability of state educational success.

President Reagan, perhaps, had the most clearly defined concept of federalism of all the postwar presidents. His concept was a return to the doctrines of dual federalism, described above. His “New Federalism”

proposal for “swapping” of programs with the collective states and “turnback” of programs he believed the national government had usurped over time was not adopted by Congress. Yet many elements of it were included in the Omnibus Budget and Reconciliation Act of 1981, including the consolidation of 57 national grant-in-aid programs into nine block grants. And the Reagan perspective set the tone of the debate over federalism for the balance of the twentieth century and into the twenty-first.

“Devolution” has been the standard position of national politicians. What devolution has come to mean is that the national government will provide block-grant funding to states and sufficient oversight and administrative support to states to assure that national requirements are met, for example, that civil rights protections remain in place; yet, the states will be free to adopt diverse and innovative practices.

Perhaps no policy area better demonstrates the devolution thrust in national-state relations than the Personal Responsibility and Work Opportunity Act of 1996. This legislation, in the words of President Clinton, “ends welfare as we know it.” Rather than the categorical grants to states provided under the New Deal era Aid for Families with Dependent Children (AFDC), the new program provides a block grant to the states, and the states are to address issues of family poverty in individual and innovative ways. Unlike AFDC, the new legislation limits the time that dependent families may remain recipients to five years and requires that after two years of receiving assistance, adults would be required to find work or participate in state-certified work programs. A generation of program evaluation will be required before an assessment of this particular devolution can be made; it is, certainly, the central piece in the devolutionary thrust.

PROSPECTS

Among the futile enterprises in American politics and administration is the debate concerning the “intention of the Founding Fathers.” Imagine the amazement of the representatives to the Constitutional Convention if they were told that they needed to anticipate the regulation of images sent through the air and received by electronic boxes in the living rooms of virtually 250,000,000 Americans! Or that the national government would need to take the lead in research to fight a devastating attack to the human immune system that would arrive from Africa!

The dual federalism assumption that national and state responsibilities could be easily and understandably separated was not demonstrable during the heyday of dual federalism. At the peak of the dual federalism period, multistate corporations were demanding that

the national government regulate commerce to protect them from the vagaries of regulation by the states.

Yet the debate continues. In the last two decades, the driving force of change in the federal relationship has been the availability of federal funds. A strong case can be made that the objectives of the Reagan New Federalism were reached not so much through the transfer and swapping of functions as through the reduction of national revenue caused by the 1981 Economic Recovery Tax Act and the ensuing declines in appropriations. In turn, state governments were compelled, for political reasons if not administrative, to assume funding those programs they thought most important to continue.

In mid-2001, the future of federalism appeared to depend on the use of the projected national government budget surplus and the long-term implications of the tax cuts enacted in early 2001. The tragic events of September 11, 2001, however, set off a series of events that consumed the projected national government budget surplus.

And the policy debate that ensued illustrates how easily the national government is compelled by political demands to accept responsibility for elements of shared national-state responsibility. For example, during the debate over airport security, that there was a national government obligation to assure public safety—a police power—was a given. The controversy was whether that meant that airport security would be provided by federal employees or that the national government would merely establish standards and assure that they were reached.

This example merely captures the dynamic of the federal administrative history. Problems arise. Politicians respond, as is appropriate in a democratic society. National and state governments share, based on the availability of resources, political factors, and diverse other considerations, responsibility and accountability for resolution of the particular problems. In a dynamic federal system, no distribution of power in the federal administrative system and the role of the states in it can be permanent.

REFERENCES

1. Madison, J. Debates in the federal convention of 1787. In *Documents Illustrative of the Formation of the American States*; U.S. Government Printing Office, 1927; 390–391.
2. *United States v. Sprague*. (282 U.S. 716 [1931]).
3. *Hammer v. Dagenhart*. (277 U.S. 251 [1918]).
4. *U.S. v. Darby*. (312 U.S. 100 [1941]).
5. Constitution of the United States, Article I, Section 8, Clause 18.
6. *McCulloch v. Maryland*. (4 Wheat. 316 [1819]).
7. Constitution of the United States, Article I, Section 8, Clause 3.
8. *Gibbons v. Ogden*. (9 Wheat. 1 [1824]).
9. Constitution of the United States, Article I, Section 8, Clause 1.
10. *Slaughterhouse Cases*. (16 Wall. 36 [1873]).
11. *Rutan et al. v. Republican Party of Illinois*. (110 S. Ct. 729 [1990]).

State Government Administration of Rural Development Policy

Michael McGuire

Department of Public Administration, University of North Texas, Denton, Texas, U.S.A.

INTRODUCTION

State government administration of rural development policy refers to the activities directed toward the economic and cultural development of rural communities that are undertaken by administrators in state government departments, agencies, and affiliated organizations. These activities may include providing resources such as finances, information, and expertise; direct service delivery; indirect service delivery by coordinating development efforts across multiple communities and/or through various state agencies; designing comprehensive development strategies by linking related programs; and establishing relationships with administrators in both federal and local government, and with non-governmental actors in communities. No single administrative activity is more important than another, and the degree to which state government administers development policy through these activities will vary according to the needs and capabilities of individual communities.

DEFINITIONS

The term “rural” describes any city, town, township, village, county, or unincorporated community that is not located within a metropolitan statistical area, as defined by the United States Office of Management and Budget and applied to United States Census Bureau data. Rural communities thus include small-to medium-sized non-metropolitan localities of up to 25,000 in population, as well as the majority of communities that are inhabited by fewer than 10,000 persons. Rural communities are unique, not just in terms of size and location; these communities are more likely than urban and suburban communities to face challenges related to structural changes in the economy, such as low skill and resource densities, lack of specialized expertise and information, isolation, and limited access to urban amenities. Although not all rural communities are alike, these internal forces have contributed, in general, to dwindling populations, relatively low wages, above average incidences of poverty, and a declining institutional infrastructure.

“Development policy” refers to the totality of programs designed to generate an acceptable quality of life for rural residents. A state’s rural development policy is the package of goals, strategies, policies, and programs that in some way promotes the growth and development of rural communities.^[1] Some programs focus on directly stimulating or maintaining business activity and/or employment, whereas others are designed to improve the physical, social, and cultural attributes of a community as a means to indirectly affect income and employment. Fundamental programs addressing taxation, regulation, and education comprise a state’s rural development policy, as do more targeted programs that compensate for perceived market failures, such as training, business financing, direct investment, trade assistance, and promotion. Rural development policy is an arena in which the federal government traditionally played the leading role, although the focus of federal rural policy historically has been on agriculture rather than on the larger issue of enhancing the physical, social, cultural, and economic well-being of rural communities. In response to federal disengagement beginning in the late 1970s and resource deficiencies across all levels of governments, state governments have sought to broaden their development focus to include the revitalization of rural cities and regions.

REASONS FOR STATE INVOLVEMENT

State government is in the best position administratively and politically to affect the development of rural communities. First, in contrast to the first three-fourths of the twentieth century, the national government has not recently played a leading role in rural development. As the condition of rural communities worsened during the 1980s and early 1990s, interdependencies between rural communities and the overall state economy became apparent. Almost by default, the well-being of rural residents became the responsibility of state government. Second, states know better than national or local governments the extent to which rural problems affect the economy and culture of the entire state. Even in states with prosperous and growing urban economies, rural problems can become a drag

on the state economy. States are able to foster the critical economic and social linkages across rural communities, and between urban and rural communities, without which the entire state economy could suffer; this is the “weakest link” argument. Market failure in rural areas may be so severe that state intervention in rural areas can improve the efficiency of the state’s economy.^[2] Third, success in preserving rural communities will ultimately rest on genuine local preferences orchestrated through coordinated collective action. States are able to tailor policies to specific local circumstances, to provide assistance to communities seeking to take charge of their future, and to design policy responses around local plans and strategies.^[3]

CHARACTERISTICS OF STATE GOVERNMENT ADMINISTRATION OF RURAL DEVELOPMENT POLICY

State administrators do not “do rural development,” but rather they undertake activities and tasks as a means to affect the development process in rural communities. The focus of state activity since the 1980s has been on designing the administrative processes and structures that are most appropriate for the unique features that differentiate rural development policy from other state policy areas. The administrative issue of determining how best to deliver development policy for rural communities has in many ways transcended the evaluative issue of what programs work best for rural development; much is known about the processes of development, but states are learning the importance of constructing the proper public technology for delivering goods and services to communities.^[4] Certain characteristics of state rural development policy distinguish it from other policy activity.^[5] These characteristics accurately depict the challenges faced by state administrators.

Rural Strategy is Crosscutting

One unique quality of state rural development policy is the way that place, rather than socioeconomic conditions, drives the content of rural policy in the states. What constitutes rural development is only partially a function of poverty or education levels; some states dedicate more resources and energy to rural development activities simply because there are more rural jurisdictions in that state. Although rural policy cannot be framed in terms of a choice between people or places, the most successful state efforts in developing rural communities have been place oriented to the extent that activities are designed to facilitate the ability of communities to plan and implement

development programs on their own. These efforts involve not only the development of the local economy but also improvement of local physical infrastructure, strengthening of local institutions, and enhancement of the social and organizational infrastructures.^[6] Persons ultimately receive the benefit of state action, but the place is the primary unit of analysis for state administrators.

Because rural problems can be identified as “rural” by definition, rather than in terms of particular functions, the responsibility for developing rural communities does not rest with any single agency, department, or office, or even a single set of designated agencies. Indeed, many models of structuring rural development policy can be found among the states, none of which can be considered any more effective than the other since the constituents and circumstances of rural residents vary significantly throughout the country.^[2] For example, states such as Iowa and Nebraska, which have approximately one-half of their population living in non-metropolitan areas, administer rural development through specialized divisions located within a larger department of economic development. The state of Texas, which has the largest total rural population of any state, but constituting only 15% of its population, has an Office of Rural Affairs through which general economic development programs—revenue bonds, enterprise zones, and Main Street, for example—are targeted to rural areas. The divisions in Iowa and Nebraska employ larger staffs and include more rural-oriented programs than does the office in Texas. On the other hand, Vermont, which has the highest percentage of all state residents residing in non-metropolitan areas, but ranks 39th in total rural population, has some specialized rural programs, but these are administered through the Department of Economic Development like all other urban development programs.

One factor contributing to the multiple administrative locations of rural development policy in the states is the nature of the policies that actually affect rural communities. Rural and urban communities are not easily distinguished with standard economic and social measures, so state policy makers often have little incentive to design a host of rural-specific programs. Crosscutting policies such as taxation, education, regulation, and health care are standard functions of state government that are not explicitly directed to, but still significantly impact, rural communities. However, it is difficult to unravel the part of these strategies and programs that specifically benefit rural areas. State administrators in many different departments forge special rural initiatives, such as the offices of rural health or special rural education programs and institutes located in state universities, while also applying generic programs that benefit rural areas.

Information is the Most Important Resource

A second distinguishing characteristic of state administration of rural development policy is the prominence of information in the total array of policy-making resources. Contrary to the traditional view of rural development as being income and/or agricultural support, an effective rural policy focus is based on important non-financial resources: information, specialized expertise, and the capacity needed to mobilize public and private resources.^[7] A significant portion of the goods and services provided by the states is thus information based. Even in programs providing financial assistance to rural communities, there is a large information-based, technical assistance component.

One of the most essential tasks undertaken by state administrators in rural development is gathering information about the conditions of rural communities and assessing the capacities of the institutions and people in these communities. By providing such information, administrators attempt to reduce uncertainty felt by communities and to raise the level of information about the applicability and merit of rural development activities. State policy makers and administrators provide information so that local administrators are able to more capably fit development policy activities with the problems and objectives of the community. In most cases, the informational demands are so great that state officials must provide road maps to guide local officials through the bewildering maze. In addition to providing information directly to communities, effective state administrators will guide prospective lenders to the right sources, make contacts, help in preparation of applications, advocate on behalf of communities, and follow through on the progress of local activity.

Information is also necessary to coordinate the various rural programs. State administrators interact with communities on behalf of the programs administered for other state agencies. For example, a state administrator in an economic development agency understands that communities may seek basic information and an interpretation, or attempt to make a program adjustment, regarding a specific funding program.^[8] As a broker of information, that administrator will work with loan and grant program officers, other state administrators, and the community representative. Similarly, a community request may be external to the development agency, for example, for road funding from the state highway department or for a construction permit from the state environment agency. In such a case, it is necessary for the state rural coordinator help make the contact, assist in applications, advocate, and follow through.

Many of the most important tasks undertaken by state agencies include developing leadership in rural communities, helping a community develop an

understanding of its local economy, conducting institutional and environmental evaluations, and assisting a community in establishing priorities for the purposes of designing and implementing strategic development plans.^[9] These rural capacity programs are generally operated out of state economic and community development departments. Many state government efforts include programs in leadership training, which focus on developing organizing skills, increasing residents' substantive knowledge of development, and exposure to "best practices" experiences in other communities; community assessment, which provides training in environmental needs assessment, goal setting, consensus building, developing visions, and undertaking a strategic planning process; and institution building through small grant programs that support local non-profit development agencies delivering services to rural residents.

Collaborative Administration

Another important characteristic of rural development policy that has become increasingly salient over time is the state's reliance on multibureaucratic, multiorganizational, and multisectoral mechanisms for producing, providing, and administering development services to rural communities. Faced with the need to provide financial resources to enhance local revenues, as well as information and expertise to compensate for the complexity and uncertainty inherent in the policy area, the state rural development administrator has become dependent on many other actors in and outside of the community. States recognize that dependencies exist among a host of private and public organizations for scarce economic resources that have, in turn, induced a new set of governing responsibilities for affecting the process of development. The governance instruments through which states and rural communities cities now plan, design, and execute development policy are complex multiorganizational networks built upon the idea of collaborative problem solving.^[1] Such networks are not merely mechanisms for coordination, but arrangements for solving interorganizational problems that cannot be achieved by single organizations.^[10] Thus, the most salient administrative and institutional arrangements in state rural development policy are not functional and agency driven, but instead are often based in public-private and state-local linkages.

Rural development is not a specific service designed and administered directly by government personnel. Intergovernmental activity for state program officials increasingly means establishing linkages among resources and programs through task forces, partnerships, and networks composed of both state and

non-state employers. The intergovernmental linkages that are required for formulating and implementing economic policy have resulted in a public administration involving multiple persons from multiple agencies in multiple jurisdictions administering multiple policies and programs. Bureaucratic structures still characterize the agencies and organizations involved in rural development, but these bureaucracies do not act alone. Instead, multiple bureaucracies collaborate on projects, design policy together, develop joint strategies, adjust mechanisms for operation and implementation, and formulate some means for determining whether expected results are being accomplished.^[11] The linkages that comprise the networks designed for one particular strategic purpose are often not the same as those in networks addressing another purpose.

EXAMPLES OF MULTIAGENCY, INFORMATION-BASED COLLABORATION

There are numerous institutional arrangements representative of the information-based, multiorganizational nature of state rural development. The Mississippi Development Authority, the department charged with administering economic and community development in a state with the seventh largest number of rural residents of any state in the country, offers a large number of information-based programs for communities that involve linkages across agencies and sectors. These programs include the Competitive Community Program, which incorporates elements of community analysis, group skills enhancement, and strategic planning into a program involving local chambers of commerce and volunteers. The Community Development Fund administered by the Iowa Community and Rural Development agency provides financial and technical assistance to groups that emphasize local partnerships and regional approaches as part of a local development strategy.

Intergovernmental partnerships known as state rural development councils (SRDC) have become a national model for procuring and sharing information, while jointly designing development strategies. Begun with 8 pilot councils in 1990, 40 councils now operate in the United States. Each SRDC involves representatives from federal, state and local government, tribal governments, and non-profit and for-profit organizations. State rural development councils play a significant role in helping states define its rural development mission and advising state agencies on how best to carry out that mission. Such councils advocate and recommend programs designed to foster community and economic development initiatives in rural areas, provide a voice for the needs and interests of rural residents and businesses, and complement local efforts to improve the

quality of life in rural areas. The SRDCs receive small operating budgets from the federal government and state-in-kind matches provide for a very small staff.

States incorporate SRDCs into the policy making and administrative structure in various ways. In most states these councils act in concert with state agencies but are separated administratively from the state government structure; such councils are established as non-profit corporations. Others may be affiliated with various agencies, such as Iowa's Rural Development Council, which is housed at the Iowa Department of Economic Development within the Division of Community and Rural Development, and California's RDC, which is located administratively within the state's Trade and Commerce Agency, with the secretary of trade and commerce serving as chairperson of the council.

CONCLUSION

Public administrators working in rural development at the state level face daunting challenges that administrators in other policy areas do not face: Action is determined by geographic rather than purely socio-economic factors; several agencies are involved in administering policy, often without an explicit, comprehensive direction; information is the most important resource in demand by communities and supplied by state administrators; and numerous organizations across governments and sectors are necessary for adequately designing and delivering development goods and services to rural communities. Administrators establish linkages with other governmental jurisdictions and multiple communities while administering numerous programs housed in many different agencies. Service design and delivery is increasingly through partnerships of public and private actors, rather than by a dominant state agency in possession of all policy-making resources. As a result, rural development administrators at the state level find themselves working with and through other agents to help communities help themselves.

REFERENCES

1. Radin, B.A.; Agranoff, R.; Bowman, A.O'M.; Buntz, G.C.; Ott, S.J.; Romzek, B.S.; Wilson, R.H. *New Governance for Rural America: Creating Intergovernmental Partnerships*; University Press of Kansas: Lawrence, KS, 1996.
2. Sears, D.W.; Redman, J.M.; Gardner, R.L.; Adams, S.J. Designing successful state rural development strategies. In *Economic Development Strategies for State and Local Governments*; McGowan, R.P., Ottensmeyer, E.J., Eds.; Nelson-Hall Inc.: Chicago, 1993; 74-90.

3. John, D. *When Does a State Need a Rural Development Policy?* Aspen Institute: Washington, DC, 1991; Unpublished paper.
4. Fosler, R.S. State economic policy: the emerging paradigm. *Econ. Dev. Q.* **1992**, 6 (1), 3–13.
5. Agranoff, R.; McGuire, M. The administration of state government rural development policy. In *Handbook of State Government Administration*; Gargan, J.J., Ed.; Marcel Dekker, Inc.: New York, 2000; 385–420.
6. Bradshaw, T.K. Multicommunity networks: a rural transition. *Ann. Am. Acad. Polit. Soc. Sci.* **1993**, 529, 164–175.
7. John, D. Where is rural policy headed? *Ann. Am. Acad. Polit. Soc. Sci.* **1993**, 529, 22–33.
8. Agranoff, R.; McGuire, M. Expanding intergovernmental management's hidden dimensions. *Am. Rev. Public Adm.* **1999**, 29, 352–369.
9. Luke, J.S.; Ventriss, C.; Reed, B.J.; Reed, C. *Managing Economic Development: A Guide to State and Local Leadership Strategies*; Jossey-Bass: San Francisco, 1988.
10. Agranoff, R.; McGuire, M. After the network is formed: process, power, and performance. In *Getting Results Through Collaboration: Networks and Network Structures for Public Policy and Management*; Mandell, M.P., Ed.; Quorum Books: Westport, CT, 2001; 11–29.
11. Agranoff, R. Rural enterprise alliances: challenges to public management. *Int. J. Public Adm.* **1998**, 21, 1533–1575.

State Government Administration, Impacts of Reform Movements

Cynthia J. Bowling

Department of Political Science, Auburn University,
Auburn, Alabama, U.S.A.

INTRODUCTION

Since at least the 1880s, governmental reform has been a prominent theme in mainstream politics, as well as in public administration theory and practice. From beginnings in the Progressive Reform movement has come a never-ending quest for effectiveness, efficiency, economy, and/or equity in government operations. This course continues today in the most recent wave of reform, commonly known as “reinventing government.”

Once known as the “fallen arches” of the federal system, the American states have gone from being considered only a “funnel” through which national programs and policies reached localities, to “governing in new and traditional spheres and responsibilities and . . . doing a better and better job at it.”^[1] States are now considered policy innovators and laboratories where administration is more professionalized, streamlined, and performance-focused than ever. This new state persona is the result of three waves of reform movements. In fact, Bowling and Wright characterize the impacts of state reform as an “administrative revolution.”^[2] In this article, three historic periods of reform and the impacts (broadly defined) they have made on state government are discussed.

WAVES OF ADMINISTRATIVE REFORM: REFORM MOVEMENTS AND THEIR IMPACTS ON STATE GOVERNMENT

Over the last hundred years, three waves of reform have swept into state government: *Rising Tides* (the period from 1915 to the early 1960s), *Cresting Waves* (1965 to late 1980s), and *A New Current* (the 1990s to the present). For each period, the major reform impetuses and a number of potential impacts are discussed within five broad categories: 1) *constitutional/statutory*; 2) *structural*; 3) *gubernatorial powers*; 4) *personnel*; and 5) *procedural*. Table 1 displays each of these movements, related goals and catalysts of reform, and the major changes occurring within each wave.

1915–1965: Rising Tides

The first reform movement was based on orthodox administrative principles and scientific management advocated by the Progressive Reformers. Responding to patronage, spoils, graft, and corruption occurring in the latter half of the nineteenth century, the Progressive Reform movement ushered in both political and administrative changes in local, state, and national governments.

The Progressives had a difficult path before them. Governors were traditionally weak figureheads. State agencies, in an effort to curtail corruption, were increasingly headed by independently elected executives and by boards and commissions. The constitutions of many states were antiquated and, when rewritten, contained detailed agency functions and structures. With constitutionally based organizations, other agencies had to be created separately by statute whenever new needs arose. In essence, state government was uncoordinated, unconsolidated, and uncontrolled.

According to tenets of scientific management, if government can create the correct structure or organization for administrative operations, efficiency and accountability will inevitably follow. The work of Woodrow Wilson, Frederick Taylor, and Luther Gulick is indicative of the search for the principles and institutions that would provide this efficiency. President Taft’s Economy and Efficiency Commission in 1912 spurred many states to action. One of the first persons to describe state reorganizations, A. E. Buck,^[3] identified a number of standards for reform that are indicative of the administrative principles advocated for state government: 1) concentration of authority and responsibility (especially under the governor and appointed department heads); 2) integration of functions into departments; 3) fewer boards and commissions for administrative purposes; 4) coordination of staff services; 5) independent auditors; and 6) recognition of a governor’s cabinet. The main focus during this time period was reorganization reflecting these standards.

Table 1 Reform movements, catalysts, and impacts

Reform movement	Reform catalysts	Main impacts of reform movement				
		Constitutional and statutory	Structural	Gubernatorial powers	Personnel	Procedural
Rising Tides (1915–1965)	• Progressive Reform movement • Scientific management • Taft Commission • Brownlow Commission • Hoover Commissions	• Reform primarily by legislative statute	• Consolidation of agencies by function • Statutory limits on number of state agencies	• Introduction of short ballots • Increased appointment powers • Lengthening of governors' terms • More centralized budget processes	• Limited gubernatorial staffs • Slow increase in civil service coverage for state employees	
Cresting Waves (1965–1980s)	• Diffusion of “first-wave” reforms across American states • Increased scope of state responsibilities	• Major constitutional amendments or new constitutions in 40 states	• Less constitutional control over agency structures • Continued consolidation by function; creation of superdepartments	• Gained ability to order executive branch reorganizations • Increased appointment and removal powers	• Increase in size of governor and legislative staffs • Continued increase in use of merit systems • Rapid improvement in quality and diversity of agency heads	• Varied and dispersed innovations in budgeting processes and management styles
New Current (1990s–Present)	• Faltering economy • Reagan devolution • Budgeting and management innovations • Publication of <i>Reinventing Government</i>		• Continued executive branch reorganizations across the states			• Reinvention initiatives • Strategic planning • Privatization or “contracting out”

James Garnett^[4] reports that this reform movement began impacting state government in 1915, primarily through a series of reorganizations spanning the next five decades. Three main reorganization efforts impelled these changes. First, although not implemented, New York produced plans to restructure executive operations in 1915, elements of which diffused across several states. In 1917, Illinois became the first state to enact major executive branch reorganizations. Twenty-five other states reorganized at least part of their executive branches over the next two decades.^[1] The second catalyst for change was the federal report on administrative practices and principles released by Luther Gulick and the Brownlow Commission in 1937. Immediately following its release, another 12 states were involved in reorganization and reform efforts.^[1] Finally, after World War II, reformers began focusing not just on proposals arising from the national government but also began a search for the states' own prescriptions. Many states enacted "little Hoovers" based on the most recent federal studies of administrative practices (the Hoover Commissions in 1949 and 1953). Reports and innovations began to spread from state to state. The adoption rate of reform proposals studied by Garnett almost tripled after World War II compared to those initiated in the earlier part of the century.^[4]

The majority of these early reorganizations was instituted by *statute*, with constitutional amendments taking much more effort to pass. Consequently, the impact of reforms was limited in scope.^[4] According to Garnett, moderate *structural* changes took place. The number of agencies was reduced in many states. For instance, in 1926, a constitutional amendment and statutory changes in New York resulted in limiting the number of departments to less than 25. Just prior to this change, there had been over 150 different agencies. Consolidation of function and authority occurred, merging agencies into statutorily defined departments. However, these reorganizations also had the effect of institutionalizing some department and agency structures within constitutional or statutory guidelines, allowing only limited discretion for reorganizations that became necessary with the Great Depression and World War II.

There were also limited improvements in the *gubernatorial capacity* to lead state agencies. One advocated reform was the short ballot, which limits the number of major departmental executives elected separately from the governor. Tennessee, New York, and Virginia were most "progressive" with the short ballot and consolidation, with partial changes in at least eight other states.^[1] This trend was quite limited, though, with a number of the states still electing several major department heads today. For instance, at least nine different high-level executives are elected separately from the

governor, including a powerful lieutenant governor, in Alabama. In other states, the governor did gain the ability to make a larger number of executive appointments at the department level, and the number of elected boards and officials governing agencies began to slowly decrease.

Other formal powers of the governorship increased during this time as well. By 1955, almost one-third of governors had four-year terms, instead of the one- or two-year terms prevalent earlier in the century.^[5] This allowed the governor to govern more and campaign less. The governor was also gaining the ability to control the budget process, with budget creation and implementation increasingly centered in the governor's office. During this time period, however, *staffing* the governor's offices and departments with adequate numbers of qualified persons remained somewhat problematic. The staffs remained small, limiting the ability of the governor to control and coordinate the departments' activities despite the more centralized budget processes. At the end of this wave of reform, the average size of the governor's staff was still only 11 persons.^[5]

Finally, professionalism in state administration improved very slowly in the first half-century. In 1938, over 55 years after the passage of the Pendleton Act, less than one-third of the states had a merit-based personnel system,^[3] though professionalism continued to steadily improve. By 1960, about half of state employees were covered by some sort of civil service system.^[1]

1965–1980s: Cresting Waves

The impact of the first movement was limited in scope and adoption but set the stage for the next wave of reform and reorganization. The reforms that began in some states in the first half of the century, with variations in degree, began to occur across the country. It seems as if the "scientific management" goals of the previous wave had finally, fully poured into state government. These reforms were culminating in what Bowman and Kearney have described as the "resurgence of the states."^[1] This resurgence occurred not only in administrative capacity but also in the scope of state policies, programs, and responsibilities. As the states were called on to implement federal policies in the 1960s, and then as policy making and program implementation devolved in the early 1970s, reforms rapidly developed.

Perhaps the preeminent impact of this wave of reform was the *constitutional* changes that occurred. Legislative statutes had been responsible for most of the reforms and reorganizations in the first half of the century. Between the mid-1960s and the

mid-1980s, however, almost 40 states experienced sweeping constitutional changes or the ratification of new state constitutions.^[1] Most constitutions were streamlined and clarified, reflecting aspects of the revised Model State Constitution created by the New York Municipal League. A major impact of these transformations was an increase in the institutional powers of the governor and legislature, while erasing specific administrative provisions from the bounds of state constitutions. Further, Bowman and Kearney report that the constitutional changes triggered other internal administrative adjustments across the states.^[1]

Improved organizational structures also appeared. Between 1965 and 1975, 20 states enacted a major reorganization of the executive branch and another 24 reorganized at least one agency or department.^[1] During this time, many states consolidated agencies by functional area into 12 to 20 superdepartments headed by governor-appointed executives.^[1,2]

Even more important was the increased ability of governors to order executive branch reorganizations themselves. In 1956, only two governors had this power. As part of the movement to strengthen gubernatorial influence over executive branch activities, governors were increasingly allowed to reorganize by executive order. By 1994, 22 governors had gained this capability.^[5] Several states also increased the governor's appointment and removal capacity. With the addition of removal powers, the governor gained another potential tool for influencing state administration and policy. In 1980, about half of major state administrators were appointed by the governor. Though Bowman and Kearney^[1] and Beyle^[6] found appointment power limited by the long ballots present in most of the states, sunset provisions and other legislation began decreasing the numbers of separately elected boards and commissions.

The *professional* resources of the governors' offices also increased. In the mid-1980s, governors' staffs averaged 25 persons;^[1] by 1994, that average had increased to 54.^[6] Larger staffs added the ability to plan, organize, and coordinate agencies, as well as provide information to the governor in a timely manner. Additionally, planning agencies were added to these staffs. In 1960, 37 states had planning departments but only 3 of those were located in the governors' offices. By the mid-1970s, all states had planning departments; 29 were directly under gubernatorial control.^[5]

The quality of personnel improved as well. The most rapid improvements in both education levels and diversity were between 1964 and 1984. Bowling and Wright^[7] report that the number of state agency heads with a graduate degree rose from 40% in 1964 to almost two-thirds of administrators in 1984. The percentage of agency heads without a college degree

dropped from 34% in 1964 to only 8% in 1984. The numbers of minority administrators rose from about 2% in the 1960s to 10% in 1984. Minority representation has remained at this level since the early 1980s. The proportion of women executives has steadily increased from only 2% in 1964 to over 22% in 1994. Additionally, over three-quarters of all state administrators are now covered by a civil service merit system.^[2]

Although these reforms and reorganizations cannot guarantee better governance, they are the necessary tools. Institutional and structural reforms and improved staffs have provided greater administrative capacity. State governments used these resources to attack new policy issues, share programmatic innovations across the states, increase resources available to local governments, and administer programs more efficiently and effectively. By the end of the 1980s, almost a century of reforms had created a government ready and able to serve its people, and to undergo the most current reform wave—reinventing government.

The 1990s and Beyond: A New Current

The advancement toward performance accountability, privatization, and economy was born of necessity. State governments faced faltering economies, President Reagan's increased devolution of programs, less federal monies, and "cutback management" in the 1970s and 1980s. The net result of these factors was a collective need to do more, do it better, and do it cheaper. The most recent wave of reform in state government seeks to improve governmental performance and responsiveness through strategic planning and budgeting, decentralization of decision making, evaluation, and privatization. Osborne and Gaebler's *Reinventing Government*^[8] is perhaps the most identified source of reform principles within this movement. During the 1990s, the "impact" of programs, administration, and reforms became a leading issue for states. The variety of changes spawned in the previous two decades culminated in this new current of reinvention.

Reinvention is differentiated from previous movements by the types of reform advocated. Dan Durning suggests that two major reform efforts are present in state government today.^[9] One method of reform is the structural reorganizations that have been present for the last 100 years. However, even as states continually reorganize administrative structures, and governors still seek to increase their formal powers, reinventing government is *procedural* and process-oriented. Although governors are implementing both types of reform in the name of efficiency and effectiveness, reinvention is currently the most discussed and debated panacea.

The preeminent aspects of reinvention are “catalytic” government and “steering rather than rowing.”^[8] This calls for government to set the scope and expectations for programs and services and then find the most efficient and effective method for delivery. This especially involves increasing competition in service provision and urges privatization if possible. Additionally, reinvention includes aspects of total quality management, a management philosophy based on a customer orientation, flattening of decision-making hierarchies, and coordination among different units of administration. Each prescription seeks to redefine the processes used to accomplish governmental goals, not make formal structural changes or increase executives’ powers.

Using data from the American State Administrators Project, Brudney, Hebert, and Wright^[10] and Burke and Wright^[11] studied the extent state agency heads reported implementing reinvention initiatives over the last decade. Surveying agency executives in over 90 agencies across all the states, Burke and Wright report that by 1998, the following recommendations had been partially or fully implemented by a majority of state agencies: 1) training programs to improve customer service (83% of agency heads surveyed); 2) strategic planning (87%); 3) quality improvement programs aimed at empowering employees (79%); 4) benchmarking (73%); 5) creation of systems to measure customer satisfaction (64%); and 6) decentralization of decision making (60%). Other reinvention goals, particularly those increasing administrative discretion, were adopted to a lesser extent: 1) reduction in hierarchy (44%); 2) increased discretion in procurement (45%); 3) simplification of human resource rules (34%); and 4) greater discretion to “carry over” funds (23%).

Further evidence of reinventing government is the noticeable movement in states toward privatization or “contracting out.” Burke and Wright^[11] found that almost a quarter of state agency heads reported privatization of major programs within their agency and almost three-fourths reported contracting with other organizations for the delivery of some services. These efforts are improving state government. Over half of administrators polled perceived improved service quality through privatization, while only 9% thought service quality decreased. Both the continuing expansion of “contracting out,” as well as the satisfaction levels indicate that at least some of these programs are improving the effectiveness of state government.^[11]

Another aspect of reinvention is customer/client orientation. Reforms are targeting this goal through training programs to create more responsiveness and mutual respect in administrative transactions with the public. Further, streamlining of paperwork and procedures improves citizens’ interactions with government programs. Citizens are receiving more responsive and

convenient services, often in a shorter time frame (see Osborne and Gaebler,^[8] Goodsell,^[12] for many positive experiences). These changes are at least partly responsible for rising trust in state government. In 1998, over four-fifths of citizens trusted state government to take care of “important business,” an increase in trust from almost every age, race, and educational group. The root cause of this trust, says Andy Kohut of the Pew Center, is the general view that “states have somehow become more competent and responsive to voters’ needs . . . (where previously) state governments were not seen as being progressive or responsible enough.”^[13]

An additional emphasis on state government performance is provided by *Governing* and Syracuse University in the Government Performance Project (GPP). They are exploring and “grading” the quality of five aspects of administration in all 50 states—financial management, capital management, human resources, managing for results, and information technology. In essence, the GPP is an attempt to measure the performance impact of administrative changes and processes utilized by state governments. Many states are doing remarkable jobs in general administration—the top three according to the GPP are Washington, Utah, and Michigan.^[14] Twenty-nine states received grades in the A or B range. However, there is variation across the states. Over a third of the states performed at only an “average” or “acceptable” level (grades in the C range) according to the GPP indicators.

CONCLUSION

This discussion brings us full circle to the question of impact. What has the last century of reform wrought? The Rising Tides of reform movements in the first part of the century brought to the states changing institutions and structures of administration. This early Rising Tide crested into the rapid and widespread reforms of the 1960s, 1970s and 1980s, furthering scientific management principles. The states solidified these orthodox principles with the creation of superdepartments, growing gubernatorial leadership, and more professional personnel. With these structures in place, a new focus on efficiency and, more importantly, effectiveness emerged in the 1990s. Many states still reorganize often, almost with every change in leadership. However, with the focus on performance and accountability that emerged in the 1990s comes the growing recognition that reorganization and reform alone does not “automatically” create a better state government. Even when reform movements do not seem to create better performance and/or governance, a more important feature may simply be the constant calls for

reform that rally state officials to continuously seek efficiency, effectiveness, economy, and equity.^[15]

REFERENCES

1. Bowman, A.O'M.; Kearney, R.C. *The Resurgence of the United States*; Prentice Hall: Englewood Cliffs, NJ, 1986.
2. Bowling, C.J.; Wright, D.S. Public administration in the United States: a half-century administrative revolution. *State Local Gov. Rev.* **1998**, 30 (1), 52–64.
3. Buck, A.E. *The Reorganization of State Governments in the United States*; Columbia University Press: Morningside Heights, NY, 1938.
4. Garnett, J.L. *Reorganizing State Government: The Executive Branch*; Westview Press: Boulder, CO, 1980.
5. Beyle, T.L. Being Governor. In *The State of the States*, 3rd Ed.; Van Horn, C., Ed.; CQ Press: Washington, D.C., 1996; 77–107.
6. Beyle, T.L. The Governors. In *Politics in the American States: A Comparative Analysis*, 7th Ed.; Gray, V., Hanson, R., Jacob, H., Eds.; CQ Press: Washington, D.C., 1999; 191–231.
7. Bowling, C.J.; Wright, D.S. Change and continuity in state administration: administrative leadership across four decades. *Public Adm. Rev.* **1998**, 58 (5), 429–444.
8. Osborne, D.; Gaebler, T. *Reinventing Government*; Penguin Books: New York, 1992.
9. Durning, D. Governors and administrative reform in the 1990s. *State Local Gov. Rev.* **1995**, 27 (1), 36–55.
10. Brudney, J.L.; Hebert, F.T.; Wright, D.S. Reinventing government in the American states: measuring and explaining administrative reform. *Public Adm. Rev.* **1999**, 59 (1), 19–30.
11. Burke, B.; Wright, D.S. Reassessing and reconciling reinvention in the American states: exploring state administrative performance. *State Local Gov. Rev.* **2002**, 34 (1), 7–20.
12. Goodsell, C.T. *The Case for Bureaucracy: A Public Administration Polemic*, 3rd Ed.; Chatham House Publishers: Chatham, NJ, 1994.
13. Boulard, G. Applause, sweet applause. *State Legis.* **1998**, 24 (7), 18–23.
14. Barrett, K.; Greene, R.; Mariani, M. Government performance project: grading the states 2001. *Governing* **2001**, 14 (5), 20.
15. March, J.Q.; Olsen, J.P. What administrative reorganization tells us about governing. *Am. Polit. Sci. Rev.* **1983**, 77 (2), 281–296.

State Government Adoption of “Managing for Results” Reforms

Donald P. Moynihan

Robert M. La Follette School of Public Affairs, University of Wisconsin-Madison, Madison, Wisconsin, U.S.A.

INTRODUCTION

This article examines the adoption of Managing for Results (MFR) reforms in U.S. state governments. MFR reforms have been adopted in all states, and involve the creation and dissemination of performance information, with the hope that such information will inform public decision-making. MFR has been popular for a variety of reasons, including its power as a political symbol. There is now unprecedented performance information about state governments. At this point, the main challenge for public officials is to find ways to ensure that this information is used.

PERFORMANCE INFORMATION SYSTEMS

For over a decade, the most popular public management reform in the state government has been the adoption of MFR reforms. From the early 1990s state governments embraced the idea of MFR. Of the variety of reforms associated with the reinvention movement, the most widely adopted in state governments were those related to strategic planning, customer service measurement, and performance measurement.^[1] By 2004, all states claimed to have some form of MFR system. Thirty-three states had MFR legislation, and the remaining 17 states had an administrative requirement. Of the 33 states that had adopted MFR requirements, all but three involved legislative changes that occurred after 1993.^[2]

MFR is a combination of strategic planning (systems that set organizational goals), performance measurement (systems that track and provide information on cost and accomplishments of government), and some form of strategic/performance management (systems that shape working relationships and structure discretion in a way consistent with organizational goals).

The standard critique of government bureaucracies is that they are too focused on rules, processes, and inputs rather than results. How can governments require agencies to manage for results? Up to now, state governments have largely limited themselves to

creating formal requirements that push agencies to produce performance information. The result is that state governments have created performance information systems.

The procedural requirements of MFR prompt strategic planning, performance plans, and reports of actual achievements. Typically, agencies are required to develop medium length (3–5 yr) strategic plans and a list of performance objectives and results on an annual or biennial basis. Once agencies generate the performance information, they typically submit it to the central budget office before publication. This is because the majority of states have tied the production of performance information to the budgeting process. As with other parts of the budget, the central budget office provides guidance on how performance information will be reported, including definitions and formats that agencies must adhere to.

The basic logic of MFR is illustrated in Fig. 1. The model assumes that MFR is a continuous system, where strategic planning creates goals and objectives, performance measurement tracks the achievement of these goals, and this information feeds into a variety of decision-making venues, which in turn will affect strategic choices to be made. Evidence from American state governments suggests wide acceptance (if not always actual implementation) of this model. States frequently refer to performance management systems, emphasizing that strategic planning and performance measurement are connected processes, and intended to affect decision-making, particularly budget and management decisions.^[3]

The Government Performance Project graded the quality of MFR efforts in all states in 1999 and 2001, and in 2005 graded both MFR efforts and aspects of information technology in a revised category called information. Table 1 provides these grades, and more detailed information can be found online at <http://results.gponline.org/>.

WHY IS MFR SO POPULAR?

There are a variety of possible reasons why MFR has become so popular in state governments. One is that

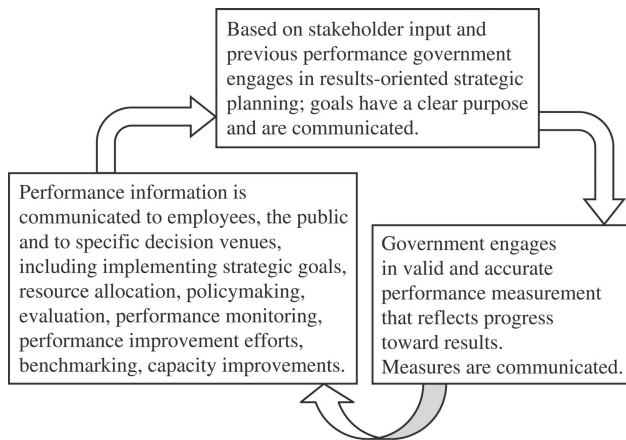


Fig. 1 An MFR system: integrating planning, measurement, and decision venues.

state governments have adopted such reforms because of budget deficits and are looking for a basis to cut programs. However, MFR reforms were mostly adopted during a period of unprecedented prosperity.

Another explanation is that adoption followed a pattern of institutional isomorphism,^[4] meaning that as soon as MFR was adopted among some pioneers and deemed to be a success, other states viewed it as important to their professional legitimacy and also adopted it. There is some circumstantial support for this theory. Along with some early adopters such as the state of Oregon, the most visible pioneer of MFR was at the federal level, which passed the Government Performance and Results Act in 1993. Many states took the federal model as a validation of performance management, and a model to emulate. MFR was actively disseminated and promoted by professional public administration organizations and other good government groups, such as the Governmental Accounting Standards Board, the National Academy of Public Administration, the National Council of State Legislatures, the International City/County Managers Association, and the Urban Institute. The effect was to make MFR increasingly important to the professional legitimacy of governments. Federal program requirements and bond agencies also encouraged the adoption of MFR. Some federal programs tie state funding to performance reporting, while bond agencies include performance information as a criterion for rating the credit-worthiness of a government project.

Elected officials were involved in the adoption process of MFR and so political considerations must be taken into account. As legislatures adopted MFR reforms, they largely neglected related reinvention proposals designed to liberate managers from

Table 1 Government performance project state grades for MFR/information

State	1999	2001	2005
Alabama	F	D+	C
Alaska	C–	C–	C
Arizona	B–	C+	B–
Arkansas	D	C–	C+
California	C–	C–	C
Colorado	C	C+	C+
Connecticut	D+	C–	C–
Delaware	B	B	B
Florida	B	C+	B
Georgia	C+	B–	B–
Hawaii	C–	C	D
Idaho	C–	C–	C+
Illinois	C	B–	C+
Indiana	C	B–	C
Iowa	B+	A–	B
Kansas	C	C+	B–
Kentucky	B	B+	B
Louisiana	B	B+	A–
Maine	C	C+	C+
Maryland	B–	B	C+
Massachusetts	C	C	C+
Michigan	B	B+	B+
Minnesota	B	B	B+
Mississippi	C	D+	C+
Missouri	A–	A–	A–
Montana	C	C	C
Nebraska	B–	B–	C+
Nevada	C	C	B–
New Hampshire	D+	D	C–
New Jersey	B–	B–	C
New Mexico	D+	C	B
New York	D+	C–	C+
North Carolina	B–	B	C+
North Dakota	D	C–	C
Ohio	C+	B	C+
Oklahoma	D+	D+	C
Oregon	B+	B	B
Pennsylvania	B–	B	B
Rhode Island	C	C	C+
South Carolina	B–	B	B
South Dakota	D	D	D
Tennessee	C	B–	C+
Texas	B+	A–	B
Utah	B+	B+	A–

(Continued)

Table 1 Government performance project state grades for MFR/information (*Continued*)

State	1999	2001	2005
Vermont	B–	B	B–
Virginia	A–	A–	A–
Washington	B+	A–	A–
West Virginia	C	C	C+
Wisconsin	C	C	B–
Wyoming	C	C+	C

constraints on the use of financial and human resources.^[5] Why? If we assume that elected officials look at the political costs and benefits of reforms, MFR is politically attractive because it communicates values to the public.^[6] The public sees that elected officials share their frustration with bureaucratic inefficiency, but have a rational plan to make government more effective and hold bureaucrats accountable for results. Such a reform proposal is difficult to resist. By interpreting MFR narrowly to mean performance information systems, elected officials limited political opposition. If they had decided to pursue broader reforms that included reorganizing personnel authority and eliminating traditional financial controls, this would have been a more contentious political battle, with strong opposition from public employee unions.

IMPLEMENTATION OF MFR

There remain numerous barriers to MFR implementation, including:

- Ensuring that performance data are timely and relevant
- Ensuring that lower level objectives and measures are consistent with strategic goals
- Ensuring that data are valid and reliable

Such barriers, while real, are technical in nature. Adequate resources, administrative energy, and information technology can overcome them. Central agencies are becoming better at identifying standards for reporting information, and agencies are becoming better at collecting performance information and undertaking strategic planning. States have increasingly devoted resources to performance auditing. Usually a part of the legislative branch, the performance auditors verify the reliability of the data that agencies report. States have also reduced the costs and increased the speed in collecting and disseminating performance

data by using the Internet, a good example of which is the Virginia Results website: <http://www.dpb.virginia.gov/VAResults>.

This progress means that states have become competent at the first two steps in the MFR system illustrated in Fig. 1: strategic planning and performance measurement. They have also increased their capacity to disseminate performance information, meaning it is usually available for decision-makers who are interested in it. Where state governments have struggled with MFR is in making sure that performance information actually shapes decisions.

Ensuring performance information use is the most difficult part of MFR, but also the most consequential. After all, what good are performance information systems if no one pays attention to them? The challenge for fostering performance information use is not primarily a problem solved by technical solutions. It requires changing ingrained patterns of decision-making that people are used to and comfortable with, which in turn requires changing organizational cultures and incentives. Because MFR has been largely associated with the creation of performance information systems, rather than as a package of more widespread organizational reforms, the likelihood of fostering such comprehensive changes is low.

Examples of state governments using performance information to make smart choices do exist.^[7] But such evidence appears to be anecdotal, and there is little evidence that this information is being used among governmentwide decision-makers in the governor's office or in the legislature. The most frequent use of performance information is to inform strategic planning, and to simply report the information to other actors.^[2]

Perhaps the expectations for MFR are too high. The enthusiasm to adopt such reforms was characterized by some unrealistic expectations about how performance information would change decision-making. Actual experience with MFR has tempered expectations. A more pragmatic goal is that performance information should inform, though not dominate, the dialogue that leads to decisions.^[8] Other factors, including political considerations, cannot, and should not be sidelined, because governmental decision-making is at heart a political process.

If we accept this view of MFR, we can begin to understand how performance information is used in the context of a dialogue between different actors. The actors involved will have different perspectives, which will influence what performance information they collect and highlight, and how they interpret and explain performance information to others. Agencies can be expected to use performance information to defend their policies and management, and press for

more resources. Central agencies will attempt to use performance information to monitor agency performance, police budget requests, and ensure consistency with the governor's policies. The legislature will use performance information to maintain oversight on the executive branch.

This difference in perspectives is wide enough that actors from different institutions are likely to disagree on the basic purposes of MFR and the problems it faces. A practical illustration of these different perspectives comes from a survey of state budgeters and managers.^[2] Budgeters perceived that the major problem for MFR is that managers and elected officials are not using the information, and that performance information does not carry enough weight in budget decisions. Managers were significantly less likely to see these as problems. Instead, they emphasized the problems associated with data collection, because they carry the reporting burden. Managers also pointed to outside factors affecting measurement, reflecting a concern about being held accountable for factors they felt they did not control. Both managers and budgeters perceived inconsistency between the performance measurement database and the accounting/budgeting database as being a major problem. This sounds like a technical problem, but partly reflects the different preferences of the executive and legislature. Managers in the executive branch create a performance management database that links to programs, while budgeters in the legislature are used to traditional account structures that may define programs in different ways. Neither side wants to change the structure for organizing programs that makes sense to them, which makes the process of aligning performance information with budget decisions that much more difficult.

While MFR was largely premised on the promise of better decision-making and greater government accountability, it is also worth considering some of the unanticipated benefits. MFR can be used as a lever to change organizational culture and motivate employees, improve internal and external communication, and foster organizational learning.^[2] These changes may not immediately and measurably improve performance, but they do enhance organizational capacity, and in the long run can be expected to improve organizational performance.

The uses of MFR tend to occur largely at the agency level, where there is less likelihood of performance data becoming a matter of political disagreement. Agency leadership plays a key role in determining whether an agency actively uses MFR for management, or simply complies with reporting requirements. When choosing whether to actively promote MFR or not, agency leaders will be guided by a consideration of the constraints and opportunities faced by their agency, as well as their own vision for

the future. If MFR furthers the vision of the agency within the constraints faced, they embrace it. If they viewed MFR as simply a reporting requirement that is likely to make little real difference to management or resources, leaders ensure pro forma implementation and little else.^[6]

CONCLUSIONS

A cynic might compare state government experience with MFR as akin to a hamster on a running wheel. The hamster is diligently trotting along, but appears to be going nowhere fast. This characterization is somewhat unfair, but contains a kernel of truth. State governments have become increasingly skillful at collecting and disseminating performance information. There is, to be sure, a wide variation in the quality of the information produced between states, and there is occasional backsliding within states. But all states are better off than they were a decade ago, and collectively, there has never been as much performance information of such good quality and as widely available as there is today. This, surely, is progress.

On the other hand, state governments are still struggling to reap the benefits of this performance information. Simply because the supply of performance information has improved, it does not mean that there is a strong demand for it, or that managers have changed their patterns of decision-making. This is the most difficult part of MFR, but also the part that will ultimately define whether a decade of reform was actually worthwhile.

REFERENCES

1. Brudney, J.L.; Hebert, F.T.; Wright, D.S. Reinventing government in the American states: measuring and explaining administrative reform. *Publ. Admin. Rev.* **1999**, *2* (1), 19–30.
2. Melkers, J.E.; Katherine, G.W. *Staying the Course: The User of Performance Measurement in State Government*; IBM Center for the Business of Government: Washington DC, 2004; <http://www.businessofgovernment.org/pdfs/MelkersReport.pdf> (accessed August 5, 2005).
3. Moynihan, D.P.; Ingraham, P.W. Look for the silver lining: managing for results in state government. *J. Publ. Admin. Res. Theory* **2003**, *13* (4), 469–490.
4. DiMaggio, P.J.; Powell, W.W. The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. *Am. Soc. Rev.* **1983**, *48* (4), 147–160.
5. Moynihan, D.P. Managing for results in state government: evaluating a decade of reform. *Publ. Admin. Rev.* **2006**, *66* (1), 78–90.

6. Moynihan, D.P. How and why state governments adopt and implement managing for results reforms. *J. Publ. Admin. Res. Theory* **2005**, *15* (2), 219–243.
7. Liner, B.; Hatry, H.; Vinson, E.; Allen, R.; Dusenbery, P.; Byrant, S.; Snell, R. *Making Results-Based State Government Work*; The Urban Institute: Washington, DC, 2001; <http://www.urban.org/UploadedPDF/results-based-stategovt.pdf> (accessed August 5, 2005); Governmental Accounting Standards Board. State and local government case studies and the effects of using performance measures for budgeting, managing and reporting. <http://www.seagov.org/> (accessed August 5, 2005); U.S. Government Accountability Office. *Performance Budgeting: States' Experiences Can Inform Federal Efforts*; GAO: Washington, DC, 2005.
8. Joyce, P.J. *Linking Performance and Budgeting: Opportunities in the Federal Budgeting Process*; IBM Center of the Business of Government: Washington, DC, 2003; http://www.businessofgovernment.org/pdfs/Joyce_Report.pdf (accessed August 5, 2005).

BIBLIOGRAPHY

1. Moynihan, D.P. Goal-based learning and the future of performance management. *Publ. Admin. Rev.* **2005**, *65* (2), 203–216.

State Government Policy Initiatives: Improving Access for the Medically Underserved

Mary Schmeida

Department of Nursing Research and Innovation, Cleveland Clinic, Cleveland, Ohio, U.S.A.

INTRODUCTION

This article discusses healthcare access barriers and state action for improving services to the medically underserved. Geographic isolation from health centers, lack of health insurance, and workforce shortage are major barriers. Many factors interact, making no single policy the only answer to the issue. Under fiscal constraint, states have been passing a patchwork of laws bridging the disparities between populations with and without healthcare access. We define the medically underserved community as rural or urban with professional shortages and/or populations with socioeconomic or demographic barriers to healthcare. Most underserved populations include rural residents, the minority, the elderly, emergency care recipients, and Medicaid eligible—poor, disabled, and children.

MEDICALLY UNDERSERVED POPULATIONS

History shows a growing maldistribution of physician specialists and health services in non-urban areas. Most graduating physicians choose to practice in areas with more clientele, higher incomes to offset school loans, and accessibility to advanced diagnostics. As a result, rural areas are at a disadvantage for care, as residents must travel a distance to see a physician or dentist. The most rural, poorest, and elderly of the states usually have a lower supply of physicians. This translates to better clinical outcomes for urbanites as metropolitan residents report fewer diseases. However, urban isolation affects many elderly and the chronically ill population. In fear of the external environment, the frail often choose to isolate despite proximity to specialists.

Healthcare for rural women and newborn is a concern. Rural counties have a higher rate of complicated pregnancies and infant mortality. Prevention of disease and early medical intervention can save lives, but is difficult for mothers living at a distance and lacking insurance. Also, rural vehicular death rates are higher owing to treacherous terrain, and lagged response time from distant emergency crews. Rural emergency care

recipients are disadvantaged as issues of personnel recruitment and retention, and funding challenge this needed public service.

Minorities have largely been medically underserved. Poorer health outcomes are seen in heart disease, stroke, cancer, diabetes, and kidney disease, with African-Americans leading other groups. Asian and Pacific Islanders lead in stomach cancer death, while Native Americans and Alaskan Natives have higher suicide rates. More Hispanic adults lack a usual place for health service than non-Hispanic black and white adults.^[1] Many health professionals lack cultural sensitivity training important for accurate medical evaluation and the treatment of ethnic/racial populations.

Barriers to Healthcare Access

Physician and nurse workforce shortage, lack of healthcare insurance, and geographic distance are primary barriers to healthcare. In 2001, Nevada, California, Texas, Utah, Idaho, Oklahoma, Arizona, and New Mexico had the lowest rate of active non-federal nurses. In 2002, active physician rate was lowest in rural and poorer states led by Idaho, Oklahoma, Mississippi, Nevada, and Wyoming. As resident population decreases so does the supply of physician specialists.^[2]

Professional workforce shortage

States are experiencing a shortage of medical professionals in non-metropolitan areas. In 2003, for example, primary care shortages in rural areas—especially for general physicians, pediatricians, and internists—were greatest in California, Arizona, and Pennsylvania.^[3] Access to mental health professionals is more limited outside urban settings, and shortage is greatest in Michigan and Kansas. Dentist shortages are also highest in rural areas.^[4] Conditions are likely to worsen; professional school enrollment in both fields lags retirement rates. Shortages of nurses are evident throughout the healthcare system. Hospital restructuring, reduced job satisfaction, and retirement (by 2010, some 40% of nurses will be aged 50 and over) are contributing factors.^[5] As the “baby boom” age

cohort retires, the consequences of the several health-care workforce shortages will be increasingly complex.

Lack of insurance

The lack of health insurance is a barrier to healthcare services. Healthcare costs are exceedingly expensive. People without health insurance (private or federal) or ability to pay for services do not have access to the healthcare system. Racial/ethnic minorities, the impoverished, and jobless commonly lack insurance or ability to pay for access. In 1990–2002, lack of insurance was greatest for people aged 18–24 yr and least for people aged 65 and older. In 2002–2003, people living inside central cities of Texas, New Mexico, and California had the greatest uninsured rate across the nation.^[6]

Many people with private health insurance also face healthcare access problems. Some health insurance plans are not adequate to meet the healthcare needs of the insured person. For example, some health plans will pay for preventative healthcare, but do not cover expenses for pre-existing medical conditions, leaving people underinsured for their chronic illnesses. Better health plans are available but with costly premiums.

State Government Policy Initiatives

Improving healthcare access is a priority for state policymakers and initiatives. States are responding to professional shortages using incentives to recruit and train professionals, improving insurance access through Medicaid waivers and private insurance regulations, and building statewide emergency medical systems (EMS). Telehealth cooperatives across states and institutions are bringing health services to rural and urban isolates through advanced technology.

Telehealth programs

Telehealth programs can improve health services to the medically underserved. Through advanced communication technology, providers at urban health centers can connect with needy patients at a distance. Persons can remain at home in their community and receive services even though the nearest specialists are miles away. Expensive start-up costs have led state agencies, private sector organizations, university, and medical institutions to pool resources and form cooperatives across states and regions, with some participating in international cooperatives. Programs are quickly growing in scope and considered cost-effective alternatives to professional shortages.

Considered a leader, California is participating in many telehealth initiatives. The “doc in the box”

program administered by the Department of Mental Health Riverside County has ranked first in mental healthcare across the states. It bridges the gap between emergency room services in urban and semiurban areas through televideo and phone. Funded through a state grant/contract with California, the Lake County telepsychiatry program is targeted at youth. The Telehealth Idaho research project provides both mental and dental rural healthcare services. Dentists can digitize and electronically transmit X-rays and still images, videoconference with distant specialists, and electronically expand the capacity of mobile dental clinics. The Kentucky Telehealth Network administered by local academic medical centers digitally connects rural public school clinics to primary care centers and institutions to service children with special needs. Telehealth can provide support to the chronically ill needing home healthcare to prevent hospital readmission.^[7]

State lawmakers are creating the legal environment for telehealth expansion by enacting a patchwork of enabling laws. Physician and nurse telehealth licensure regulations protect professionals from liability risks and govern practice. Health information privacy laws protect sensitive health information as it is accessed and transferred by professionals in patient management. Medicaid telehealth reimbursement laws expand professional interests to use telehealth for servicing the poor and disabled populations. Both Medicaid and Medicare policy have expanded eligibility for telehealth practice to allied health professionals.

Workforce recruitment and training

Financial incentives, such as school loan repayment and subsidies have been effective intergovernmental policies in recruiting professionals to work in underserved areas. Housed in the federal Bureau of Health Professions, the National Health Service Corps assists states to recruit primary care, mental health, dental, and nurse professionals into underserved areas through school loan repayment and scholarship assistance. Through a matching grant, states operate individual loan repayment programs. Michigan has a large State Loan Repayment Program aiding in the recruitment and workforce retention in professional shortage areas. Some recruiting programs are web-based, such as Med Job Louisiana, and target professionals of diverse populations to meet the needs of minority groups.

As with recruitment, training programs aim toward cross-cultural competence. Culturally sensitive professionals are able to respond to ethnic/racial concerns and contextual differences, important for accurate illness diagnosis and case management. This includes effective treatment of the rural culture, socioeconomically distinct from the urban. Also, training professionals

to practice several areas of medicine is a cost-effective solution. Cross-training the primary care physician about mental health diversifies those choosing to practice in professional shortage areas.^[8] Other state policy attracting physicians and nurses to underserved areas include subsidies for relocation start-up costs and the accompanying premiums, income tax credits, and out-of-state guest licenses.

In addition to loan forgiveness programs and recruitment campaigns, fixing the nurse shortage has been challenging. States and professional groups have focused on improving long-standing work condition problems by providing measures, such as safer work environment and patient workload, better compensation for overtime work, flexible schedules, and expanding the role of the nurse. As a response to their extreme nurse shortage, the California Nursing Education Initiative is building public-private collaborative relationships to improve the capacity of California's nursing education system^[9] as is more state funding.

State health insurance initiatives

State health insurance initiatives include Medicaid waivers and private insurance regulations. The majority is reforming insurance coverage through Medicaid reform initiatives and has a Medicaid system created under a waiver. These include grants, collaboration with other government and private agencies, premium cost sharing, and beneficiary empowerment. Using managed care arrangements, California proposed Medicaid expansions to include previously uninsured groups. These arrangements build program capacity creating efficiency, improving service management and enrollment. Other initiatives include beneficiary cost sharing of premiums. New Mexico's *Mi Via* Program initiative empowers beneficiaries through privatization. The state allocates a budgeted amount for healthcare spending to beneficiaries, giving them discretion in choice and control over their own services, instead of government making the consumer choices. Traditionally underserved populations including the elderly and mentally ill are eligible.^[10] In Louisiana's 2005 healthcare reform package communities can leverage local funds so as to use pooled funds for health needs.^[11] Some states have made collaborative arrangements with other public agencies to promote efficiency and better target services. Noted for best practice, Maryland's interdependent arrangement between the Medicaid and State Mental Health agencies has improved access to community behavioral services, decentralizing decision-making authority on day-to-day operations. New Jersey's collaborative arrangement between Medicaid and several human service divisions is also recognized for best practice in improving access to mental healthcare.^[12] However,

at times of fiscal stress, some states have reduced Medicaid benefits and applied eligibility restrictions even though enrollment has increased for the State Children's Health Insurance Program.

The states are also regulating the private insurance industry with a variety of laws to improve access to health benefits. Parity laws have become common. Commercial employers offering employee health benefits must provide insurance parity or equitable insurance coverage between medical disease treatment and treatment for mental illness and drug addiction. In the past, higher capitation rates for medical treatment left some with mental illness and addiction spending down to qualify for Medicaid benefits. Some states, however, have only passed watered-down versions of these regulations with loopholes for commercial employers. As states are looking to private insurance as one answer to improving access, some are mandating the private market as the primary insurance coverage followed by public insurance. Experiencing Medicaid budget constraints, promoting the use of private insurance first is gaining appeal in poorer states. Others are requiring insurance carriers to sell insurance to everyone applying for benefits despite their age and health status, and to make premiums more equitable across age groups, gender, and health conditions. Policies on refundable tax credits for health insurance for those covered by employer or private non-employer health plans are in debate.

Building statewide EMS

State officials are responding to federal pressure, morbidity and mortality rates, and citizen demand for statewide EMS that work. A continuum of laws has been passed with the majority funding EMS personnel recruitment programs, equipment, manpower training, transportation, communication technology, and public safety programs. As a venue for change, the state offices of rural health are catalysts in reforming the rural healthcare system, such as emergency care, managing interests and funding opportunities, diffusing policy ideas interstate and intrastate.^[13]

Building communication capacity between clinicians and EMS personnel is essential, and tele-emergency care will become more common. Early communication between emergency technicians with physician specialists located at hospital sites can improve clinical outcomes for injuries requiring quick medical intervention, such as stroke and heart attack. Advanced technology has the potential to improve rural and urban emergency treatment. Recognized as a best practice, Chicago's Fire Department is equipped with medical reporting technology as hand-held notepads improve early communication with emergency room personnel to save time and lives.^[14] Building the

information infrastructure is a nationwide effort and responsibility is shared among levels of government and the information industry. Given current terrorist issues, building effective EMS is a priority for disaster preparedness and response, a citizen concern. Stakeholders on the state–local level are voicing their values on emergency preparedness to state and national officials hoping to encourage policy development.

REFERENCES

1. U.S. Department of Health and Human Services. Summary Health Statistics for U.S. Adults: National Health Interview Survey, 2003. Series 10. No. 225, 2005.
2. U.S. Census Bureau. Statistical Abstract of the United States: 2004–2005. Available at: <http://www.census.gov> (accessed July 2005).
3. U.S. Department of Health and Human Services, Office of Rural Health Policy. In *Rural Healthy People 2010: A Companion Document of Healthy People 2010*; Government Printing Office: Washington, DC, 2003.
4. U.S. Department of Health and Human Services. In *Health, United States*; Government Printing Office: Washington, DC, 2001.
5. General Accountability Office. 2001. *Nursing Workforce Emerging Nurse Shortages Due to Multiple Factors*. Report No. GAO-01-944. Available at: <http://www.gao.gov> (accessed August 2005).
6. U.S. Census Bureau. 2004. Income, Poverty, and Health Insurance Coverage in the United States: 2003. Available at: <http://www.census.gov> (accessed July 2005).
7. Office for the Advancement of Telehealth. 2005. Grantees Directory. Available at: <http://www.telehealth.hrsa.gov> (accessed July 2005).
8. Office of the Surgeon General. 2005. Minority-Oriented Services. Available at: <http://www.surgeongeneral.gov> (accessed August 2005).
9. California State Assembly, Republican Caucus. 2005. Education. Available at: www.republican.assembly.ca.gov (accessed August 2005).
10. New Mexico Human Services Department. 2005. The New Mexico *Mi Via* Program. Available at: <http://www.state.nm.us> (accessed August 2005).
11. State of Louisiana. 2005. Governor's Health Care Reform. Available at: <http://www.dhh.louisiana.gov> (accessed August 2005).
12. National Governor's Association Center for Best Practices. 2005. Best Practices in State Medicaid and Mental Health Program Collaboration. Available at: <http://www.nga.org> (accessed September 2005).
13. U.S. Department of Health and Human Services. 2005. Rural Health Policy. Available at: <http://www.rural-health.hrsa.gov> (accessed July 2005).
14. National Governor's Association Center for Best Practices. 2005. Chicago Pilots Hand-Held Reporting Units for Paramedics. Available at: <http://www.nga.org> (accessed August 2005).

State Government, Aspects of Managing Strategic Change in

Paul C. Nutt

Fisher College of Business, Ohio State University, Columbus, Ohio, U.S.A.

Jeffrey A. Rosen

Gregory Paxton

School of Public Policy and Management, Ohio State University, Columbus, Ohio, U.S.A.

INTRODUCTION

Agencies and departments of state government are faced with many new demands that call for changes in their missions.^[1,2] The unique needs of public agencies^[3,4] require leaders to think about managing processes of change in ways that differ from those approaches used in the private sector. Under the current political climate, change in state government agencies will be continual, and agency leaders must be prepared to incorporate the flow of new demands. Managing such a process of change in a public agency can be a daunting task. This article builds on the strategic change concepts developed by Nutt and Backoff^[5] and shows conditions under which leaders can initiate the process of strategic change in a state agency along with some of the required moves. First described are the types of agencies that exist in today's public environment emphasizing the conditions that influence an agency's prospects for undergoing change. Then discussed are several strategies that fit agency types susceptible to change.

MANAGING THE CHANGE PROCESS

Managing a process of change requires state agency leaders to make several crucial moves. Among the first is to determine the agency's prospects for change. Once this has been done, the leader matches a proposed strategy with the expected agency environment. Leaders are offered some thoughts on how to do this.

Agency Positioning and the Prospects of Change

Many leaders of state agencies find themselves in situations of limited capacity and/or low responsiveness to expectations.^[6] Capacity has been eroded by years of budget cutting, limitations on prerogatives from court rulings, rule making by legislatures, assignment of

duties to other agencies during reorganization and chronic staff turnover. As capacity reaches a low ebb, complacency may set in, making change difficult. Demands are experienced at a rate that exceeds the agency's ability to understand the proposed actions or requests. Such an agency may have been reacting with a learned helplessness posture, in which elaborate rationales for avoiding action supplant building capacities.

Fig. 1 shows how state agencies can be classified according to their capacity and the extent to which prerogatives are dictated by outsiders, such as legislatures. Strategic change attempts to move a state agency toward increased capacities and increased control over their environment. This section discusses how state agencies get in such situations along with when it is possible to move agencies toward change. This is shown by moving up the diagonal, running from lower left to upper right, as indicated by the arrows in Fig. 1. The path along the diagonal toward a more proactive posture identifies conditions that make change feasible. When there is an imbalance between external control over actions and internal capacity, strategic change is apt to be blocked. A balance between control and responsiveness is required to make strategic change seem possible.

Agencies that Resist Change

Agencies have little motivation to change when internal competence and external control are seriously imbalanced. When either competence or control dominates, attention is focused on maintenance of competence and control and not on using competence to exercise useful control. Such agencies get located far off the diagonal describing the path of change in Fig. 1.

Professional agencies

Agencies that have high internal capacity, real or perceived by outsiders, with little external control are called professional. Activity is directed internally by

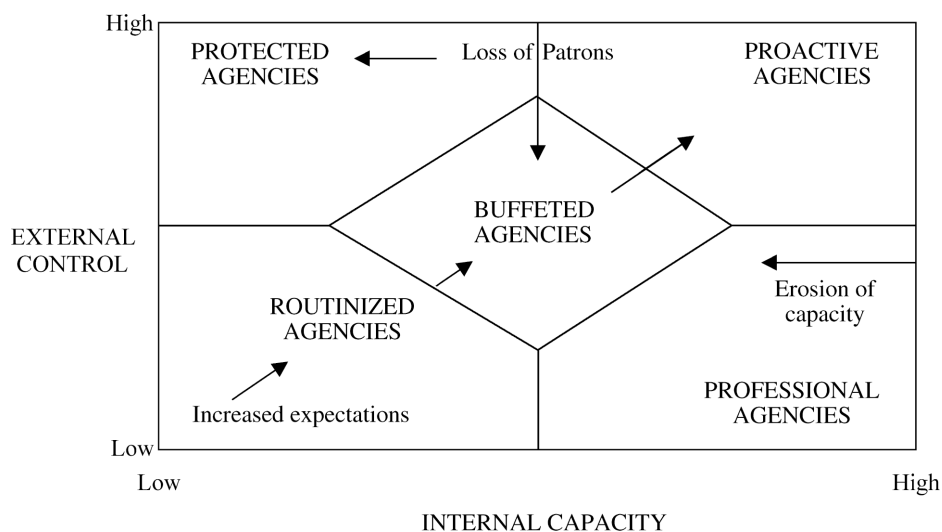


Fig. 1 Forces prompting change in state agencies.

professional values. Control is low because the political, legal, and economic structures that ordinarily direct and channel energies and ensure accountability are not in place. Typically, such an agency has protected its budget or resource base and its prerogative to act within a prescribed arena. State departments of taxation and the attorney general's office provide examples. Such agencies rely on highly skilled and self-regulated staff members to provide what professionals deem to be the proper response for classes of problems and issues. Other examples include state run mental health and acute care hospitals, university facilities, social work agencies, and any other agency dominated by a professional elite that controls the terms of services that are provided. Strategic change with professionals who seek to maintain this kind of control is not apt to be successful. The professional bureau or agency can ignore client or stakeholder preferences, the political agenda of reformers, and public support and may even be insulated from budget cutting. In some instances, the professional agency can even turn aside legal mandates and judicial orders. Such agencies have little motivation to change.

Protected agencies

Agencies that lack internal capacity but have a clear cut domain of action are termed "protected." Such an agency has exploited its leverage with a power elite to carve out a domain over which it has an exclusive right to act. Legislation may even be written to give exclusive rights or responsibilities. Protected agencies then take rigid control of the domains. The Iowa Legislature gave the University of Iowa Medical School exclusive rights to provide all tertiary care in the state. In a protected agency, old-boy networks that connect the agency to a power elite are carefully maintained. The network is used in this way to deflect

criticism and routine accountability and to harness the forces that would normally be used to monitor actions. Departments of development (DOD) are often viewed as essential by Republican legislatures and governors. Such an agency can be expected to carefully maintain such a network to harness forces that would normally call for accountability. When a critic asks how many new jobs have been created by a DOD and how much tax base has been given away, the network is mobilized to deflect the critic's questions or make them seem unwarranted or politically motivated. Control of turf in this way by agencies creates little incentive to improve capacity and makes the agency quite political.

Routinized agencies

Agencies low in competence and capacity quickly become routinized. To justify their existence, highly codified routines are developed and rigidly followed. If no one questions what such an agency produces, the agency can go about its processing with very little accountability. Examples include state highway patrols, contract activities in departments of transportation, claim processing in bureaus of worker compensation, and service departments such as travel, records and transcripts, and stores in universities. Without leverage, a critic gets nowhere because little is expected of the agency and even less has been provided to facilitate its mission. Former students who complain about poor service, such as delays in getting a transcript, are ignored by university leaders and oversight bodies that have more important things to do. Delays in receiving compensation benefits fall on deaf ears, charges of unfair or unethical contracting for highways are ignored, and questions about how the highway patrol spends the bulk of its budget go unanswered. In each case, oversight bodies have little motivation to question long-standing practices.

The Forces for Change

The protected, professional, or routinized agency can be pushed toward change by loss of control and/or a loss in capacity. When this occurs these agencies become buffeted and begin to field pressure to change. This movement can occur in three ways as shown by the arrows in Fig. 1.

Change for a professional agency

Strategic change in a professional agency is difficult even when forces calling for more control over the agency's actions are clear and compelling, unless the agency's capacity is thought to be declining. Such an agency is moved toward the diagonal in Fig. 1. If perceived competence falls perceptibly, such a drop will make strategic change both possible and desirable. For example, the FBI's reputation at the death of J. Edgar Hoover went into eclipse forcing the FBI to become more concerned about public relations and to reexamine its role. The strategy of a professional agency rests on maintaining its internal capacity. So a decline in competence (real or perceived) is a necessary precursor for change to occur in professional agencies prompting them to attempt to maintain an image of competency when new practices are being forged and adopted. If real or perceived competence declines, such an agency requires new positioning. Defending past practices coupled with small adaptations is carried out. The IRS, stung by reports that they unmercifully harassed tax payers, confiscated the bank accounts of children, and gave inaccurate tax advice, mounted strategy to do damage control and reform its procedures.

Change for a protected agency

Agencies that cultivate patrons and ignore building capacity end up in the upper left corner of Fig. 1. The strategy in such an agency is to maintain the old-boy network and other arrangements that protect turf. Real change is not apt to occur unless these arrangements erode. Erosion of turf or prerogatives can occur when agency patrons lose elections, leave office, or when an old-boy network ceases to function. This shift pushes the protected agency into the mainstream. The political agenda of newcomers or a shift in power among stakeholders signals the need for action. The clear and compelling call for better services and the like can no longer be ignored. To respond, the agency must increase its internal capacity, shifting it toward the diagonal in Fig. 1.

After a new governor took office, Ohio's Department of Public Welfare experienced such a shift, moving from a protected position to one of increased

expectations. The status quo was no longer acceptable and the need to change was accepted. The welfare department had to abandon a strategy in which clients were expected to respond as directed and move toward a posture that emphasized helping clients, which provided the impetus for strategic change. Increased awareness of low internal capacity often accompanies shifts in environmental control. Exposing this low capacity invariably forces rapid action to improve matters, increasing the need for and desirability of strategic change. Leadership transitions and increased consensus to adopt a new proactive stance often result.

Change for a routinized agency

State agencies with modest to low competence can be ignored for decades. Oversight bodies place few demands on them, so they go about their actives with little accountability until there is a shift in expectations for either competency or control or both. For instance, complaints about long waits and inept staff prompted change in a state bureau of motor vehicles. New funds were provided and new systems were designed to allow mail-in license plates renewals that separated taxpayers and staff, thereby reducing both dysfunctional contacts with the public and the time involved. Note how complaints had to become intense before action was taken that recognized the need to increase competence (new systems). Such attention is necessary, but sometimes not sufficient to move an agency out of its routine posture. The state bureau of motor vehicles soon reverted to its bureaucratic ways when funds to go beyond the mail-in license renewal program were cut off by a new administration. Bureaucratic routines arose to carry out the mail-in license renewals program that recognizes only clearcut mailings and tolerated no exceptions.

Promoting Change

Agencies on the diagonal in Fig. 1 are put into an unstable state when internal capacity and external control are more balanced. When a question about whether the organization's directions have or can produce hoped for outcomes is posed externally through the preferences of stakeholders, shifts in political agendas, erosion of public support, budget shortfalls or windfalls, or legal rulings, the agency is prodded into action. Such a prodding can also arise internally when there is a shift in consensus or changes in leadership. Internal and external prodding for agency transformation then make strategic change seem pragmatic and lower barriers to change erected by political or professional postures. Two new agency types are created: the buffeted and the proactive.

The buffeted agency

When a professional, protected, or routinized agency experiences a sustained erosion of competence and/or increased control, it becomes buffeted. The buffeted agency can get pulled about by fickle public support, changing legal mandates, shifting political agendas, and shifting preferences of clients or key benefactors. As mandates shift, budget support may shrink and constraints about the use of resources emerge. As a result, shifting priorities provide moving targets that can lead to frustrated staff and turnover of key personnel. Turnover forces large investments in training that further bleed away resources. But agencies experiencing such buffeting are susceptible to change.

A process of strategic change can help the buffeted agency come to grips with the forces at work in its environment. Capacity to act can be gradually built by finding ways to increase the agency's control over its environment, which in turn, can lead to both a happier and more challenged staff and can also reduce turnover. As capacity is built and environmental control established, the agency develops an increased awareness of the need for strategic change.

The proactive agency

Demands for high internal capacity balanced with high external responsiveness and the obligations inherent in this responsiveness creates a proactive agency. An agency in such a posture develops a shared understanding of needs and potential to act. Consider the Ohio Department of Natural Resources (DNR), which moved toward a proactive organization with high environmental responsiveness by strategic management. Key stakeholders who had expressed criticisms were folded into task forces to deal with important issues such as wildlife protection, natural areas, hunting and fishing rights, waterways, and recreation. Programs were identified and funds sought to operate them through licensing the use of wild areas, through park programs, and through new types of fishing and hunting licenses. The fees collected were turned back to fund program development, which produced more fee generating activities until two-thirds of the DNR's funding was outside the state's budget process. By responding to stakeholders in this way, the DNR became a self-sustaining proactive agency that is highly responsive to stakeholders, as well as high in action.

MATCHING STRATEGY TYPES TO AGENCY ENVIRONMENTS

Agency leaders managing a change process must identify their environment and match to them an

appropriate strategy. Needs and responsiveness to needs, which describe the environment in which all state agencies must operate, are often "socially constructed." They get fashioned by agency leaders and oversight bodies around beliefs about what people need and the extent to which an agency should respond to these needs. The agency leader or its oversight body may see these needs and their ability to respond to them realistically, or they may misrepresent them.^[7] Change is difficult when an agency leader wants to be more proactive than his or her oversight body sees as being appropriate or necessary or when the oversight body calls for acting on low priority or non-existent needs. Both situations create unstable states that cause agency leaders difficulty and will ultimately derail any change effort. When this occurs, the agency leader has several strategic choices. Within each enacted environment there are two strategic moves possible, one superior to the other.^[8-10]

Fig. 2 depicts the environments in which state agencies operate. Fig. 3 illustrates how state agencies can shape their strategy to fit their environment, indicating both effective and ineffective strategies. An agency's *responsiveness* to perceived pressures for action creates the environment in which they operate. *Need volatility* creates a loss of control, as the needs of the clients being served become difficult to predict and can increase with little warning. The lower left corner of Fig. 3 depicts an agency that is functioning without producing much useful action. The aim is to move such an organization as far as possible up the diagonal from lower left to upper right, gradually making it more proactive in responsiveness to emerging needs and fashioning arrangements to meet these needs. Instead of merely responding, a state agency can alter its relationship with other agencies and stakeholders that make up their environment. The expected level of responsiveness is modified as the organization makes these commitments. This can prompt a leader to adopt a *director*, *bureaucrat*, *accommodator*, or *compromiser* strategy.

Agencies that respond as if needs are at a low ebb are in a *placid* environment. A *drifter* strategy, a likely result of this environment, is often adopted by professional agencies in which goldbricking is widespread.

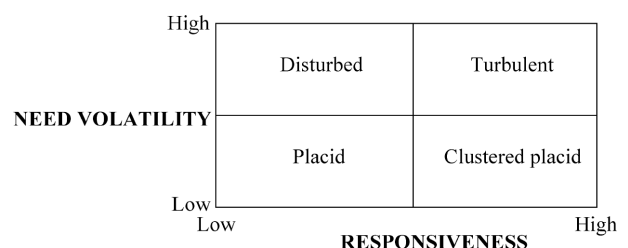


Fig. 2 Environmental types.

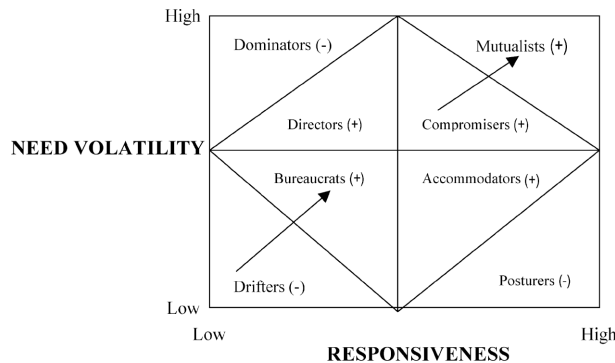


Fig. 3 Agency strategies.

A *bureaucratic* strategy moves such an agency to a minimally acceptable posture. An agency with this strategy takes a defensive posture to protect seemingly important practices through budget maximization and the creation of slack for inevitable budget cutbacks. The equivalent of protectionism in firms is sought. Agencies with high responsiveness to low demands for action prompt a *clustered placid* environment. Such agencies often adopt a *posturer* strategy where issues are brought to the attention of others for action. State agencies that bring problems to the state legislature for decision making display some of these attributes. For instance, the new head of the Ohio Bureau of Workers Compensation (BWC) posed questions about the adequacy of agency reserves and called for increased funding to bring them up. Legislators saw the request as a call for higher fees and/or more tax support, which was seen as irresponsible. The legislature had commitments to firms to keep fees low and to the electorate to keep a lid on taxes. The BWC leader was told to bring solutions, not problems, to the legislature in the future. A more effective approach applies an *accommodator* strategy to deal with each cluster by recognizing and dealing with the emergent and unique needs in each cluster. Priorities must be set to allocate available funds among the clusters based on the compelling nature of the needs in each cluster. Allocation is carried out in conference, where moves are made to create opportunities to make subsequent moves. This tactic is used in budget hearings, media reports, and the like to gradually reveal allocations and read reactions.

Disturbed environments arise when responsiveness has been low and the need for action abruptly increases. Agencies that deal with emerging needs with little or no new responses create a *dominator* strategy. The motivation here is to maintain discretion in the choice of action and divorce it from accountability. State departments of taxation use such a strategy. A better approach calls for the *director* strategy, in which new demands are dealt with by taking measured

action. Turbulent environments arise when responsiveness and the need for action both increase dramatically. This often prompts the *compromiser* strategy. The agency plays one constituency against another, meeting the needs of important constituents, needy constituents, or both when resources permit. Such a strategy was adopted by the BWC to prompt internal process changes that increased the speed of claims processing and reduced costs to allow employer fees to fall. A *mutualist* strategy is a better approach because it attempts to create collaborative arrangements to meet all the important needs that arise with umbrella organizations.

Mutualist strategy calls for novel arrangements to be created to meet emergent needs. A *mutualist* strategy calls for: 1) leaders to set the tone by subordinating personal and organizational interests; 2) a consortium that draws together the key stakeholders into an umbrella organization to address priority issues; 3) creating or shaping a vision to meet emergent needs; 4) seeking win-win arrangements for affected parties; and 5) promoting trust so stakeholders will cooperate toward resolving the issue, thereby guiding stakeholders away from competition and toward cooperation.^[5]

A *mutualist* strategy takes shape as a quest, which proposes new initiatives to create a grand vision. A *mutualist* strategy calls on stakeholders to search for new ways to deal with unseen realities and entices key people to collaborate in meeting the needs that prove to be crucial. To create major shifts in competence, the buffeted agency must move toward the *mutualist* strategy.

Strategy is used by an agency to focus effort into reaching for its visions or ideals by providing plans, ploys, positions, and perspectives.^[11] As a plan, a strategy offers a way to take action, such as finding new clients and services. Ploys are strategies invented to deceive an opponent. Note how ploys are not useful in *mutualist* strategy, but are essential to a *compromiser* strategy. Strategy as position seeks services that seem needed by reading environmental signals. Positioning, such as protecting budgets or becoming territorial by staking out a service area, often accompanies the *director* and *accommodator* strategies. As a perspective, strategy captures an organization's traditions and commitments, becoming a touchstone for future action.

CONCLUSION

When trying to introduce change into the public sector, strategic leaders face many daunting challenges. To improve the prospects of success, leaders must both understand the environment under which the agency

operates and be able to recognize situations in which change is feasible. Once a leader understands the agency's environment, leaders then can select a strategy that allows the agency to realize the most pervasive change possible. In this article, agency leaders have been provided descriptions of the various types of environments that exist in the public sector. Some of the needed steps have been laid out to match strategies to agency environments, denoting strategies that fit and those that have the best prospects of making meaningful changes. It is clear that the current political environment necessitates the changing of public agency missions and practices. The concepts described will hopefully offer agency leaders one way to think about the process of change.

REFERENCES

1. Osbourne, D.; Gaebler, T. *Reinventing Government*; Addison-Wesley: Reading, MA, 1992.
2. Wechsler, B.; Backoff, R.W. Policy making and administration in state agencies: strategic management approaches. *Public Adm. Rev.* **1986**, 321–327. July/Aug.
3. Bozeman, B. *All Organizations Are Public*; Jossey-Bass: San Francisco, 1987.
4. Rainey, H. *Understanding and Managing Public Organizations*; Jossey-Bass: San Francisco, 1989.
5. Nutt, P.C.; Backoff, R.W. Strategy for public and third sector organizations. *J. Public Adm. Res. Theory* **1995**, 5 (2), 189–211.
6. Nutt, P.C.; Backoff, R.W. Organizational publicness and its impact on strategic management. *J. Public Adm. Res. Theory* **1993**, 3 (3), 209–231.
7. Daft, R.; Weick, K. Toward a model of organizations as interpretative systems. *Acad. Manage. Rev.* **1984**, 9 (2), 284–295.
8. Emery, F.E.; Trist, E.L. Casual texture of organizational environments. *Hum. Relat.* **1965**, 18 (1), 21–32.
9. Miles, R.E.; Snow, C.C. *Organizational Strategy, Structure and Processes*; McGraw Hill: New York, 1978.
10. Agor, W.H. *The Logic of Intuitive Decision-Making*; Quorum: New York, 1986.
11. Mintzberg, H. Crafting strategy. *Harvard Bus. Rev.* **1987**, 65–75. July–August.

State Government: Complex Policy Network Management

Kathleen Hale

Department of Political Science, Auburn University, Auburn, Alabama, U.S.A.

INTRODUCTION

Complex policy networks are unique organizational arrangements in which public and private organizations work together to accomplish effective, coordinated action in furtherance of public policy objectives. These networks represent a distinct organizational response to the pressures of devolution and government reform. Complex policy networks extend the boundaries of intergovernmental relations beyond the public sector and incorporate private sector actors, most frequently non-profit organizations. As organizational arrangements that exist outside the traditional hierarchical authority and communication relationships of public agencies, complex policy networks place new demands on the skills and resources of public managers. Examples can be found across public policy domains; the strong local character of education and criminal justice administration reinforce the management complexities experienced in state governments. More broadly, complex policy networks present challenges to state governing authority and capacity in the public policy process.

NETWORKS AS UNIQUE ORGANIZATIONAL ARRANGEMENTS

The concept of complex policy networks has evolved through the study of networked arrangements in public policy implementation, their consequences for public administration and the impact on public management practices. Networks in policy implementation are understood as structures of interdependence involving multiple organizations or parts of organizations in which one unit is not simply the formal subordinate of another unit within the context of a larger, hierarchical arrangement.^[1] A complex policy network can be distinguished from other organizational arrangements in the lack of formal hierarchy and the lack of congruence of organizational mission among constituents.

In the context of state government, complex policy networks present cross-sectoral arrangements that include all or part of one or more public agencies and all or part of one or more private, typically non-profit, organizations. As such, complex policy networks

include the distinct characteristics of both public and private non-profit organizations. Vertical lines of authority and elected or other forms of representative governance coexist with volunteer boards of directors and community volunteers, each in pursuit of distinct organizational missions. A host of vertical, horizontal, formal and informal lines of authority transmit communication, spread responsibility and diffuse authority.

In complex policy networks, public policy objectives are shared in whole or in part by constituent organizations. Similarly, separate policy implementation frameworks are conceptualized for the network as a whole. Policy outcomes exist for the organizations in the network, the network itself, and the broader policy communities and political environments that touch the network.^[2]

EVOLUTION IN AN ENVIRONMENT OF INTERGOVERNMENTAL REFORM

Complex policy networks evolved over the past several decades against a backdrop of shifting pressures on intergovernmental relationships. Inflationary pressures on real estate and property taxes throughout the 1970s and 1980s fostered successful citizen campaigns to limit state taxes and state spending. During the 1980s and 1990s, studies of government inefficiency gained national prominence. The Grace Commission in 1984 reported that tremendous savings in public expenditures could be realized through private sector contracting and revised procurement protocols; the report fueled federal budget reductions and legislative initiatives to balance the federal budget. The 1993 National Performance Review and the reinvention movement promoted decentralization, flatter organizational hierarchies, and increased reliance on external service provision through contracting or privatization. The net result has been a growing preference within public administration for private sector models of management. Intergovernmental funding patterns also shifted in response to changing presidential agendas; common themes included the devolution of funding responsibility to the states and grant requirements for collaboration between public agencies and community service providers at the state and local level.^[3]

Combined pressures of devolution, reinvention and budget restrictions, states turned increasingly to the non-profit sector for direct delivery of services. Non-profit service providers are attractive to state governments for at least two reasons. First, non-profits can access private donations that could be used to leverage federal funds under the concept of matching grants. Second, non-profit organizations are also seen as trustworthy organizations and therefore may present less political or financial risk to state governments than for-profit service providers.^[4] Today, non-profit organizations are recognized as a vital element of policy relationships that touch state government.

Complex policy networks can be distinguished from other responses to devolution and the pressures of government reform. Networks move beyond the concept of privatization vehicles and outsourcing contracts through incorporation of greater numbers of actors and points of interaction between constituents. Networks incorporate actors across the intergovernmental spectrum but move beyond regional government arrangements by incorporating the dimension of public-private collaboration.^[5]

A final factor that contributed to the evolution of complex organizational arrangements in policy design and implementation has been the growing acknowledgment of certain public policy issues as essentially intractable or “wicked problems.”^[6] Policy solutions to “wicked problems” are formed through collaborative, often interdisciplinary, networks of stakeholders. As a response to these various pressures, new forms of organizational arrangements have developed in and around state government that depart from traditional, agency-based models of government service provision. State governments are now actively engaged in myriad networks of non-profit service providers, volunteers, advocates, and other units of government in order to design public policy and to accomplish policy objectives.

EXAMPLE IN THE FIELD OF EDUCATION POLICY

School-to-work programs illustrate the management challenges presented to state government by one type of complex policy network. School-to-work programs are established under the federal School to Work Opportunities Act of 1994 (STWOA). The goal of STWOA is the institutionalization of a new system to help students move from high school to productive careers. The Act (STWOA) is jointly funded and administered at the federal level by both the Department of Education and the Department of Labor, and federal grants are available to the states for design, implementation, and evaluation of school-to-work activities. Federal grant eligibility is predicated on construction

of state collaborative partnerships that represent the interests of a school-to-work system broadly defined. Specific constituencies within state government must be represented, including the governor, the state educational agency, and state officials responsible for economic development, job training, postsecondary education, vocational education, vocational rehabilitation, and the state human resource council. Collaborative partnerships also incorporate school administrators, parents, teachers, students, employers, business associations, organized labor, and community groups.

At the state level, a bureaucratic hierarchy is nominally imposed through the creation of a state school-to-work office. The state school-to-work office is embedded within the state department of labor or commerce, the governor’s office, or most typically, in the state department of education. The activities that must be coordinated for successful program implementation, however, involve diverse constituencies outside the jurisdiction of state government. Successful management in this context thus involves communication between representatives of formal and informal constituencies with distinct agendas. In some states, school-to-work collaborations emerged through the creation of new partnerships between educators and employers; in others, collaborations were created from existing bureaucratic arrangements such as employer councils formed under the Job Training Partnership Act, or state associations of regional groups of educators.

School-to-work policy also encourages local design and implementation to promote access as well as programming tailored to fit local needs. Federal policy does not mandate a specific institutional structure but broad accountability for program implementation and success rests with state government. Within this broad policy framework, state management efforts are diffused toward outside agencies and community groups. State government administrators invest considerable time in building and maintaining relationships and advancing policy objectives without hierarchical control or significant central management structure.

EXAMPLE IN THE FIELD OF CRIMINAL JUSTICE POLICY

Drug court programs present another dynamic example of a fluid arrangement of public and non-profit organizations that function in the absence of an established state hierarchy or centralized state management structure. Drug court programs or “drug courts” operate within the state court system across the states to provide early intervention and intensive substance abuse treatment for nonviolent drug offenders. Drug courts utilize intensive judicial intervention, early and

rigorous treatment, and in many cases, accelerated case processing within the state criminal justice system. As early as arraignment, defendants are screened and referred immediately to multiphase outpatient chemical dependency treatment programs, which include a strict regimen of contact with one or more treatment providers, frequent drug testing through urinalysis, frequent status hearings before a designated drug court judge, and participation in supportive services such as literacy and GED education, employment, and employability training. Criminal defendants are encouraged to participate in drug courts in exchange for expungement of the criminal record of the drug offense upon successful completion of the drug court program.

State drug court programs synthesize, manage, and monitor the efforts of law enforcement agencies and elected law enforcement officials, judges, elected executive and legislative offices, administrative agencies, and private non-profit service providers to accomplish specific policy objectives. Integrated and collaborative arrangements between public organizations from each of the three branches of government, administrative agencies, and service providers are required as a condition of federal grant funding under the Omnibus Crime Control and Safe Streets Act of 1994. Multiple public offices and agencies interact with multiple private service providers. Most service providers are non-profit or community-based organizations with distinct missions which may also include the advancement of religious tenets or faith-based philosophies.

Drug court models require extensive linkages, collaboration, and communication in order to operate. However, no formal hierarchical bureaucracy defines the complex policy network of the drug court. Local communities and the states are free to innovate in developing new organizational arrangements for governance and management of the drug court process.

COMPLEX POLICY NETWORKS AND MANAGEMENT CHALLENGES

As the examples in education and criminal justice policy illustrate, complex policy networks are not confined to a single state agency and are not fully subject to the governing authority or management of state government. Rather, complex policy networks exist in tension with the traditional hierarchical bureaucratic model of public administration. In contrast to the stable, long-term command-and-control relationships in the bureaucratic model, network relationships are far more spontaneous and are essentially transitory. In public management, networks overlap with bureaucratic hierarchies, thus management of complex policy networks is best considered in tandem with bureaucracy rather

than in opposition to it.^[7] This duality presents specific challenges for managers in state government in terms of governance, performance measurement, and the capacity of extended networks to deliver public services and public policy.

GOVERNANCE AND COMPLEX POLICY NETWORKS

Complex policy networks impact governance—the way in which government accomplishes its job. A significant body of policy implementation literature argues that the primary contribution of complex policy networks is to bring about policy failure through the pressures of various interest groups. A relatively recent line of research argues that complex policy networks can be studied specifically in terms of their ability to enhance policy success and may represent an alternative framework for analysis that is a more accurate representation of public policy implementation than top-down or bottom-up models currently in use.^[8] In extending the study of governance beyond the public agency context to the network, states will need to consider means of adapting the primarily vertical hierarchy of the public bureaucracy to organizational arrangements with considerably more extensive horizontal reach than that within the typical public agency. Such a task is made all the more challenging by the need for states to retain legitimacy as sovereign powers in organizational arrangements in which the distribution of power is highly diffuse.^[9]

ASSESSING CAPACITY AND THE ABILITY TO IMPLEMENT POLICY

Considering complex policy networks as a positive influence on policy implementation has extended the way in which state capacity is considered. Early assessments of state capacity for policy design and implementation focused on program management, policy management, and resource management as baseline elements necessary to accomplish effective public governance.^[10] Capability in program, policy, and resource management were considered in the context of traditional hierarchical bureaucratic arrangements; state agencies that possessed sufficient capacity were those that demonstrated successful levels of performance in the management of programs, policies, and resources. Successful bureaucracies were also dependent on the presence of specific functional capabilities within the ranks of public personnel.^[11]

Capacity also depends on the presence of organizational characteristics that enhance the likelihood of successful performance in policy implementation.

Intergovernmental management literature recognizes a process framework and skills that can be expected to activate and sustain intergovernmental arrangements, including authority bonds, exchange relationships, trust, and the effective use of information; in the increasingly networked environment, these skills have become critical aspects of the public management environment.^[12] Stakeholder support, mission valence, and leadership are now understood to be important characteristics of successful intergovernmental relationships.^[13] Factors that suggest success for complex policy networks include institutional arrangements built upon trust relationships between stakeholders, information exchange across constituencies, the effective use of information, and staff dedicated to those principles.

Despite the identification of personal and institutional characteristics that can be expected to enhance policy implementation and policy success, measuring the capacity or success of complex policy networks in implementing policy has proven to be elusive. Although the network itself is salient, systematic descriptions of the arrangements and roles of public and non-profit constituents vary widely, as do the descriptions of methods by which to assess success. One line of emerging research focuses on information exchanges between state policy administrators and non-profit organizations as a policy diffuses across the states. Research on the fifty-state implementation of school-to-work programs suggests that non-profit organizations may provide synthesized information about best practices and program evaluation that is particularly valuable to state administrators during the implementation process.^[14] Another emerging approach seeks to model the layers of activity within a network in order to assess the effects and interactions of various network components. A study of the implementation of the end of social promotion within the network that is the Chicago Public School system finds that school-level variations in implementation itself have an effect on the consequent changes in student performance.^[15] These examples of emerging research demonstrate potential avenues for further exploration of policy implementation in the context of complex policy networks and state government.

Building capacity for successful policy implementation is also predicated upon understanding how to define success. In the networked environment, state governments must also consider more specifically the goals they expect to accomplish, as states rely on entities outside the state itself for the design and delivery of public services and public policy. State governments will be challenged to move beyond a "one-size-fits-all" orientation. The complex policy network context argues in favor of multiple, flexible, and multimethod performance evaluation methods that capture the

diversity of actors and uniqueness of programs and missions.^[16]

CONCLUSION

The study of complex policy networks has advanced during the past two decades. At the same time, state government continues to face challenges in designing new governing structures for policy implementation that satisfy the demands of democratic accountability and the reality of complex networks in public administration. State government engages in complex policy networks on virtually every area of public policy. In managing policy implementation in these networks, states exercise sovereign authority by implementing the commands of the elected branches of government. In moving beyond traditional hierarchical models of implementation, states must now also seek to connect, promote, and evaluate the activity of multiple actors who are not public employees. Running through each of these challenges is the question of how to best equip public managers and the broader public workforce with the skills and knowledge necessary to perform in the complex policy environment. Future research in these areas holds much promise for state management efforts directed at policy design and implementation, which necessarily involve complex policy networks.

REFERENCES

1. O'Toole, Laurence, J., Jr. Treating networks seriously: Practical and research-based agendas in public administration. *Public Admin. Rev.* **1997**, *51*, 45–52.
2. Provan, K.G.; Milward, G.B. Do networks really work? A framework for evaluating public sector organizational networks. *Public Admin. Rev.* **2001**, *61*, 414–423.
3. Conlan, T. *From New Federalism to Devolution: Twenty Five Years of Intergovernmental Reform*; Brookings Institution Press: Washington, DC, 1998.
4. Salamon, L.M. The resilient sector: the state of non-profit America. In *The State of Nonprofit America*; Salamon, L.M., Ed.; Brookings Institution Press: Washington, DC, 2004; 3–61.
5. Goldsmith, S.; Eggers, W.D. *Governing by Network: The New Shape of the Public Sector*; Brookings Institution Press: Washington, DC, 2004.
6. Rittel, H.J.W.; Webber, M.M. Dilemmas in a general theory of planning. *Policy Sci.* **1973**, *4*, 155–169.
7. Heinrich, C.J.; Hill, C.J.; Lynn, L.E., Jr. Governance as an organizing theme for empirical research. In *The Art of Governance: Analyzing Management and Administration*; Ingraham, P.W., Lynn, L.E., Jr., Eds.; American Governance and Public Policy Series, Rabe, B., Series Editor; Georgetown University Press: Washington, DC, 2004; 3–19.

8. Kickert, W.J.M.; Klijn, E.-H.; Koppenjan, J.F.M. Introduction: a management perspective on networks. In *Managing Complex Networks: Strategies for the Public Sector*; Kickert, W.J.M., Klijn, E.-H., Koppenjan, J.F.M., Eds.; Sage Publications Ltd.: London, 1999 (reprint); 1–13.
9. Kettl, D.F. *The Transformation of Governance: Public Administration for the Twenty-First Century*; Johns Hopkins University Press: Baltimore, MD, 2002.
10. Burgess, P.M. Capacity building and the elements of public administration. *Public Admin. Rev.* **1975**, *35*, 705–716.
11. Gargan, J.J. Consideration of local government capacity. *Public Admin. Rev.* **1981**, *41*, 649–658.
12. Agranoff, R.; McGuire, M. *Collaborative Public Management: New Strategies for Local Governments*; Georgetown Press: Washington, DC, 2003.
13. Rainey, H.G.; Steinbauer, P. Galloping elephants: development of a theory of effective government organizations. *J. Public Admin. Res. Theory* **1999**, *9*, 1–32.
14. Mossberger, K.; Hale, K. Polydiffusion in intergovernmental programs: information diffusion in the school-to-work network. *Am. Rev. Public Admin.* **2002**, *32* (4), 398–422.
15. Roderick, M.; Jacob, B.A.; Bryk, A. Evaluating Chicago's efforts to end social promotion. In *Governance and Performance: New Perspectives*; Heinrich, C.J., Lynn, L.E., Jr., Eds.; Georgetown University Press: Washington, DC, 2000; 34–67.
16. Radin, B.A. *Challenging the Performance Movement: Accountability, Complexity and Democratic Values*; Georgetown University Press: Washington, DC, 2006.

State Government: Promotion of Innovations

Keon S. Chi

*Center for State Governance Transformation, Council of State Governments,
Lexington, Kentucky, U.S.A.*

INTRODUCTION

There are compelling reasons for state officials to innovate with policies and programs. We need alternatives, new ways of managing government agencies and delivering public services, because many traditional and ongoing programs are not working as well as they should. Moreover, the public expect state policy makers and administrators to initiate creative policies and programs. The public also expect state officials to do more with less, especially during times of fiscal austerity and retrenchment.

Innovations are not easy to conceive or replicate, however, because new initiatives usually mean changes or disruptions in an organization. These barriers and obstacles to innovations stem from the nature of government and human resistance to change. Yet, there is hope for state policy makers who desire to overcome difficulties to innovate. Those who want to innovate should focus on problems that are widely recognized as urgent and need to convince others that the current system is not working. They might justify innovation on the basis of cost effectiveness or customer demand. Finally, although innovators may experience some setbacks along the way and go long periods without apparent progress, they should remain tenacious, patient, committed, and optimistic.

DEFINITIONS AND MODELS

As seen in the voluminous literature, research scholars tend to treat policy or program innovations somewhat differently than do practitioners. This is true particularly when defining an innovation in state government. Broadly stated, academic researchers tend to define an innovation as a discovery or the adoption of a policy or program that is new to an adopting agency or a state. In fact, nearly every political scientist and public management specialist who have done extensive research on state innovations have defined an innovation in very much the same way. They defined innovations from the adopter's perspective, not from the inventor's perspective.

For example, in his pioneering study on innovation diffusion in state government, Walker defined an

innovation as “a program or policy which is new to the states adopting it, no matter how old the program may be or how many other states may have adopted it.”^[1] According to him, innovation is not an invention or creation. For Savage, “an innovation is a policy adopted by a state for the first time.”^[2] Downs and Mohr defined innovations as “the earliness or extent of use by a given organization of a given new idea, where ‘new’ means only new to the adopting agent, and not necessarily to the world in general.”^[3,4] Similarly, Nice regarded innovations as policies or programs “new to the state adopting them, even if the policy was first adopted in other states years earlier.”^[5]

On the other hand, practitioners and others tend to define an innovation as an invention or a brand new policy or program that has not been adopted previously by other jurisdictions. They tend to define an innovation as a creation or a novelty, not a mere replication of a policy or program initiated by others. The Council of State Governments' (CSG) innovations awards program, for example, has used this definition since the program's inception in 1975. The national innovations awards program recognizes creative and novel initiatives in place that have proven to be effective in addressing significant policy, program, and management issues but have not been adopted by other states. According to this definition, therefore, replicating a program originated in other states would not be considered truly innovative.

Although innovations in state government take many forms, these forms may be grouped under the two broad types: political or policy innovations and program or management innovations. The first type includes large-scale or “macro” innovations, while the second type includes smaller-scale or “micro” innovations. In his study of innovations in the federal government, Polsby defined a political innovation as “a policy or a set of policies that seem to have altered (or promise to alter) the lives of persons affected by them in substantial and fairly permanent ways.”^[6] And Bingham, in his study of innovation in local government, defined a political innovation as a public policy.^[7] Thus, political or policy innovations in state government, as in the federal and local governments, may include those initiated often, but not exclusively, by enabling legislation. Policy innovations include

enacting statewide education reform laws, restructuring court systems, and providing ongoing welfare-to-work reforms.

On the other hand, program or management innovations are initiated as creative tools or means to implement political or policy goals more effectively and efficiently, with or without legislation measures. Examples of program and management innovations include productivity and management improvement measures as well as technology application. Experiences with the CSG and Ford Foundation–Harvard University innovations awards programs indicate that nearly 9 out of 10 innovations selected as semifinalists or finalists are program and management innovations, not policy innovations. Therefore, public administration and management specialists tend to regard innovations in terms of program and management innovations.

The process of originating innovations in state government has been analyzed in two models: policy planning and “groping along.” Under the policy-planning model, innovations are initiated, for example, with prior planning and well-designed, comprehensive implementation plans. This model is usually found in political or policy innovations. Under the groping-along model, innovations are taking place without much systematic planning. This model de-emphasizes the initial policy-planning idea or legislative requirements, in favor of discretionary or more flexible ways of accomplishing policy objectives through trial and error, often modified by experience. As one researcher observed, “policy-planning model encourages managers to pay careful attention to identifying the right innovative idea, drafting a statute or execute policy that fully reflects it, and planning in advance for problems that might arise in implementation. The groping-along model, by contrast, involves much less attention to the original idea or policy—quite a number of different policies might be effective at starting an organization down a useful path—and much more attention to the organization’s ability to gather and respond to operational experience along the way.”^[8] The policy-planning model may be more appropriate when explaining political-policy innovation processes, and the groping-along model can better explain program or management innovation processes.

A review of innovations in state government by the CSG innovations awards program suggests that most program and management innovations fall under the groping-along model. The same is true with the Ford Foundation and Harvard University innovations award program. According to researchers for the Ford–Harvard innovation award program, “Rather than emerging from systematic research projects or policy analysis, they (innovations) have involved one person or a small group drawing upon past experience and/or informal communications with colleagues in

other settings. Following initial conception the ideas have generally been quickly tried out in practice and have subsequently evolved via trial and error rather than formal experimentation or analysis. Given that the state and local governments almost never invest in research and development, this pattern was perhaps to be expected.”^[9,10]

REASONS FOR INNOVATIONS

The question, “why innovate,” has been discussed in terms of purposes, causes, and motivations of innovations. This question has also been answered by examining political culture, organizational environment, and personal qualities that attribute to initiating or adopting innovations by a state government. Direct and indirect reasons for policy and program innovations in state government include the following:

1. *Alternatives*: New ways of managing government agencies and delivering public services are needed, because many traditional and ongoing programs are not working as well as they should. State policy makers and agency managers need to come up with alternative policies and programs.
2. *Public confidence*: State policy makers and administrators innovate policies and programs, because they want to meet the public’s expectations and ultimately gain voter confidence.
3. *Doing more with less*: Increasingly, American taxpayers have tended to refuse to pay more taxes for government services. As a result, state policy makers and administrators are expected to “do more with less,” especially during periods of fiscal austerity and retrenchment.
4. *Managers*: Unlike in the past, state managers and workers today tend to regard innovations as part of their responsibilities. They also innovate, because they attempt to apply rational approaches to problem solving. State workers in many jurisdictions are also concerned about self-efficacy, self-esteem, and legacy.
5. *Businesses*: State officials innovate because private businesses do. The influence of the private sector in state policy making and administration has been tremendous, particularly when embarking on strategic planning and benchmarking, quality management and productivity improvement, and privatization activities.
6. *Competition*: The states are competing with each other in nearly every policy and program area, and the level of competition has been heightened in recent years. In order to compete with other states, they need to create new

programs or replicate policies and programs implemented by other states.

7. *Technology*: The availability of technology makes it possible to be innovative in old ways of conducting business in state government, especially in telecommunication and management information systems. In fact, many program and management innovations are linked to new technology application in state agencies.
8. *Funding*: For practical purposes, innovations are needed to obtain outside funding. State policy makers and administrators have to be more innovative to experiment with pilot projects or to acquire waivers from federal agencies, for example. They also need to develop new ideas for implementation to raise money from private sources.

INDIVIDUAL INNOVATORS

Individual initiators of new ideas in state government include elected officers, appointed managers, and ordinary workers. Innovators tend to possess certain qualities, however. Based on a review of background information on selected innovations and interview data on innovators, for example, a Ford Foundation innovations specialist characterized the distinguished qualities of state and local innovators as “product champions,” who can communicate their visions and mobilize support for new ideas. The major attributes of typical innovators include energy, persistence, knowledge of the problem, political astuteness, the ability to inspire people, and the capacity to work effectively with diverse groups.

One national survey of innovators in state government conducted by the CSG revealed that innovators are very well educated, with virtually one-half of the innovators possessing an advanced degree from a 4-yr institution.^[11,12] Innovators tend to have a diverse array of academic majors with concentrations in social sciences, business, education, and public administration as the dominant educational backgrounds. Many innovators appear to be mid-career employees who returned to school to advance to managerial opportunities and have prior experience in the private sector, mostly in nonprofit organizations or in private consulting firms dealing with government programs. The innovators in the survey sample were very active professionally. The majority belonged to at least one state and one national professional association. Close to a majority belong to two or more associations at some level. Interestingly, national associations appeared to be more important to innovators than regional associations.

The most common singular role pattern is for innovators to generate innovations as part of their

day-to-day professional responsibilities. The primary groups involved in helping the innovator develop the innovation are those individuals working with the innovator on a day-to-day basis, such as coworkers and supervisors. Innovators usually find their strongest support from those they worked with and from those groups most dependent upon their agencies' services. In more than 80% of the cases, according to the CSG survey, the innovation had a potential effect on the organization.

The individual innovators identified by CSG relied primarily on their immediate coworkers for professional information and secondarily on the professional associations to which they belonged. The innovators appeared to be aware of what other states were doing within their respective policy areas. One-half of the respondents said that they used innovations originated in other states as a source of information. More than 60% of the states mentioned as innovative were in regions other than the innovators. These results depart from the notion that innovators look primarily to regional neighbors when contemplating a new venture for their agencies.

INNOVATIVE STATES

Previous studies on state-level innovations have certain characteristics. First, the political scientists defined an innovation as a state law, not anything else. Second, they used laws in selected policy areas during certain periods to measure the innovativeness of states. And, third, researchers measured and ranked states according to the date or speed of adopting innovations. Thus, states that adopted laws earlier than others were deemed to be more innovative. For example, Walker ranked states according to composite scores of innovations, based on 88 state laws in 11 policy areas enacted by at least 20 states between 1870 and 1966. Gray's innovativeness rankings are based on only 12 laws in three areas: education; welfare; and civil rights.^[13] Savage used 181 laws from 15 policy areas to measure innovativeness of the states: 58 laws in the nineteenth century; 54 laws in the early twentieth century; and 69 laws in the late twentieth century.

As for future research, at least three points may be made regarding methodologies to measure the innovativeness of the states. First, as mentioned earlier, most scholars define innovations as adoptions of innovations implemented elsewhere. The point is that, if this definition is used, however, as a former NSF intergovernmental program coordinator once pointed out, “every state is an innovator,” because all states borrow ideas from each other.^[14] Some would argue that truly innovative states are “pioneering” or “bellwether” states. Therefore, not every state adopting

another state's program should be considered a pioneer or a bellwether state. The question is: Should we continue to define innovations from adopter's perspective?

A second point has to do with the use of state laws to determine the innovativeness of states. The question is, "Can we measure the innovativeness of the states based solely on their adoption of laws?" Laws are certainly a major source of information on how states are doing. Using laws might be a convenient but not necessarily the most comprehensive way of measuring a state's innovativeness. There appear to be several inherent problems, when relying solely on laws, in the study of innovation diffusion in state government, however.

First, the date of adoption of a law does not necessarily correspond with the timing of implementation of that policy. In many states, significant policy or program initiatives are implemented even before relevant laws are enacted. The delay in policy implementation may be attributable to several factors such as the state's budgetary constraints, administrative rules, and regulations. The issue is the elapsed time between the date of adoption of legislation and the time when the policy or program is actually implemented. The gap might prove to be significant in the study of innovation diffusion. In addition, some laws are subsequently changed or repealed. One might ask, therefore, "Can we consider a state innovative even if it failed to implement a new law adopted earlier?"

In addition, when measuring innovativeness based on laws, the content of the legislature measure might need careful examination. The early innovation studies virtually disregarded the nature and extent of an innovation. Perhaps, the researchers assumed that the same or similar types of titles of laws would contain the same or similar provisions. What they needed, it seems, was a list of laws with years of adoption by the states. Although states tend to replicate laws adopted by others, more often, they also tend to revise or sometimes improve such laws to fit their own situations. While the intent of laws might be the same, the implementation process can be substantially different. Should we pay attention only to the intent of laws and not to the methods of implementation? It is also possible to think about a situation where a laggard state can come up with more effective legislation that is more beneficial to the people.

Finally, perhaps more importantly, the previous studies on innovations virtually disregarded program and management innovations that account for most innovations occurring in state agencies. It should also be pointed out that most innovations fall under the groping-along model, not the policy-planning model.

INNOVATIVE ORGANIZATIONS

An innovative agency may be defined as "one in which everyone—from those on the leadership team, to middle managers, to front-line supervisors, to front-line workers—acts on a sense of responsibility for inventing, developing and implementing new ways of achieving the organization's mission."^{15]} In order to make a state agency innovative, leaders, managers, and front-line workers all must have clearly defined goals and proper roles to play.

First, goals are important when creating innovative organizations because goals can set directions organizations want to go, set specific targets to reach, and be measured by quality and quantity. Goals can help define outcomes in meeting the public's expectations and demands. Innovations may not occur without the appropriate organizational environment and opportunities created by leaders who can help articulate goals for managers and workers.

Second, leaders in innovative organizations must have personal qualities, including serious personal commitment and devotion; tolerance and openness to new initiatives, suggestions, and proposals; and a willingness to share power with others, including middle managers and front-line workers. Leaders in innovative organizations need to use realistic strategies developed jointly by managers, workers, union members, and others. Such strategies may be developed through total quality management or similar management tools.

Third, the role of middle managers should be clearly defined. Why are middle managers important in the innovation process? At least three reasons can be offered: Through "buying-in" activities, middle managers can have a sense of ownership of innovations; they can create an environment and allocate the resources necessary to implement innovations; and they can help continue and sustain innovations.

Fourth, to create an innovative agency, state policy makers and managers need to redefine roles of front-line workers. Behn has offered useful "hints" for involving front-line workers in the innovation process: 1) be immediately responsive to requests for improved working conditions (or when they ask for a new photocopier, produce it); 2) support mistakes (or sit next to the first honest innovator who is called before a legislative committee); 3) create an explicit mission and related performance measure (or give people a real reason to be innovative); 4) broaden job categories (or do not let each individual do only one narrow task); 5) move people around (or do not let workers think they need learn only one job for life); 6) reward teams, not individuals (or find ways to beat the formal performance-appraisal and promotion system); 7) make the hierarchy as unimportant

as possible (or at least walk around without an entourage); 8) break down functional units (or do not let the procurement guys tell everyone “no”); and 9) give everyone all the information they need to do the job (or do not let the overhead units hoard the critical data).^[15] Tell everyone what innovations are working (or have front-line workers report their successes to their colleagues).^[16]

DIFFUSION OF INNOVATIONS

Most studies on innovation diffusion have focused on horizontal diffusion among the states. Innovation researchers have attempted to explain why some states tend to be more innovative than others. One key question has been, “What causes a state government to adopt a new policy or program?” Despite many case studies on the question, findings of such studies may be classified into two models: “internal determinants” and “regional diffusion.”^[17]

The internal determinants model posits that causes of adopting innovations may be answered by examining internal characteristics—political, economic, and social—of a state. The classic answer to the question is that the larger, wealthier, more industrialized states adopt new programs somewhat more rapidly than their smaller, less well-developed neighbors. Other scholars found, however, that states that are innovators in one policy area are not necessarily innovators in other areas, and that “innovativeness is not a pervasive factor; rather, it is issue- and time-specific.” Still, some researchers discounted interaction effects in the innovation diffusion process and claimed that “only the policy itself can be assumed to be invariant over time.”

The regional diffusion model, to the contrary, focuses more on the influence of “geographically-proximate states” in the innovation adoption process. It has been widely thought that state policy makers and managers look primarily to regional neighbors when contemplating a new venture. There is no question that regionalism plays a major role in innovation diffusion. As mentioned in the previous section, however, innovators in state governments tend to replicate innovations from any states, not necessarily from neighboring states. Today, thanks to communications technology and other factors, state innovators can learn about creative and novel solutions to problems from any jurisdiction without regard to its location or characteristics.

How long did it take for all states to replicate an innovation in a given policy area? This question has been investigated by several political scientists. For example, based on 88 state laws enacted between 1870 and 1966, Walker found that the average elapsed

time of innovation diffusion decreased from 52.3 yr for all adoptions (or 22.9 yr for the first 20 states) in 1870–1899 to 25.6 yr for all adoptions (or 18.4 yr for the first 20 states) in 1930–1966. And, Welch and Thompson, based on 57 laws, including 52 from the Walker data, found the average diffusion time to be nearly 30 yr.^[18] It is interesting to see if the average diffusion time has further decreased in recent years. It can be assumed that the diffusion time may have drastically shortened in part because of the roles played by third parties, including federal agencies with a variety of financial incentives.

On the other hand, researchers on innovations in state government have paid little attention to vertical diffusion of state innovations, transferring innovations from states to the federal government. Vertical diffusion occurs because of “a search for ideas from within and pressures from below and outside to adopt policy innovations.”^[19] Vertical, as well as horizontal, innovations may be diffused through third parties—interest groups, profit and non-profit organizations, civic organizations, and national and regional organizations of state officials. Third parties can act as information brokers and catalysts by conducting various activities such as clearinghouses, publications, conferences and workshops; by offering policy options and recommendations; by providing technical assistance; by giving interstate consulting services; and with recognition and award programs to encourage continuous adoption of innovations.

The major national organizations representing state governments in vertical diffusion of innovations include the National Governors’ Association, the CSG, and the National Conference of State Legislatures. Other national organizations of elected state officials are the National Conference of Lieutenant Governors, National Association of State Treasurers, National Association of Attorneys General, and National Association of Secretaries of States. In addition, there exist a number of organizations of appointed state officials that can help states diffuse innovations from one state to another, and these include national associations of state budget officers, personnel executives, purchasing officials, state administration and general services, emergency management, parole and probation, and telecommunication and information resources.

CONCLUSION

Innovations are not easy to conceive or replicate, because new initiatives usually mean changes and disruptions in an organization. There are institutional barriers as well as employee resistance to innovations,

especially from those affected by implementation new policies or programs. The absence of pressure or incentives is another obstacle in adopting innovations.

REFERENCES

1. Walker, J.L. The diffusion of innovations among the American states. *Am. Polit. Sci. Rev.* **1969**, 63, 880–899.
2. Savage, R. Policy innovativeness as a trait of American states. *J. Polit.* **1978**, 40, 212–224.
3. Downs, G.W., Jr. Toward a theory of innovation. *Admin. Soc.* **1979**, 10 (4), 379–408.
4. Mohr, L.B. Determinants of innovations in organizations. *Am. Polit. Sci. Rev.* **1969**, 67, 1174–1185.
5. Nice, D.C. *Policy Innovation in State Government*; Iowa State University: Ames, 1984; 7 pp.
6. Polsby, N.W. *Political Innovation in America: The Politics of Policy Initiation*; Yale University Press: New Haven, 1984; 8 pp.
7. Bingham, R.D. *The Adoption of Innovation by Local Government*; Lexington Books: Lexington, MA, 1976; 217 pp.
8. Golden, O. Innovation in public sector human services programs: the implications of innovation by groping along. *J. Policy Anal. Manage.* **1990**, 9 (2), 219.
9. Altshuler, A.A.; Zegans, M. Innovations and creativity: comparison between public management and private enterprise. *Cities* **1990**, 7 (1), 16.
10. Altshuler, A.A.; Altshuler, A.A.; Behn, R.D.; Eds. *Bureaucratic Innovation, Democratic Accountability, and Political Incentives Innovation in American Government: Challenges, Opportunities, and Dilemmas*; The Brookings Institution: Washington, DC, 1997; 38 pp.
11. Grady, D.O.; Chi, K.S. Innovators in state governments: their organizational and professional environment. In *The Book of the States, 1990–91*; The Council of State Governments: Lexington, KY, 1990; 382–404.
12. Chi, K.S. Innovations in state government. In *The Book of the States, 1996–97*; The Council of State Governments: Lexington, KY, 1997; 537–546.
13. Gray, V. Innovations in the states: a diffusion study. *Am. Polit. Sci. Rev.* **1973**, 67, 1174–1185.
14. Cox, R.W. *Organizing for Innovation*, Paper Presented at the Annual Meeting of the American Society for Public Administration, 1985; 2 pp.
15. Behn, R.D. *Creating Innovative Public Agencies: A Challenge for State and Local Government*; Terry Sanford Institute of Public Policy: Durham, 1995.
16. Behn, R.D. Creating an innovative organization: ten hints for involving front-line workers. *State Local Gov. Rev.* **1995**, 27 (3), 221–234.
17. Berry, F.S.; Berry, W.D. *Explaining Innovations by American States: The Case of Taxes*, Paper Presented at the Annual Meeting of the American Political Science Association, Washington, DC, 1988.
18. Welch, S.; Thompson, K. The impact of federal incentives on state policy innovations. *J. Polit.* **1980**, 24 (4), 715–729.
19. Spart, A. *State-Federal Relations: The Vertical Diffusion of State Policy Innovations*, Paper Delivered at the Annual Meeting of the American Political Science Association, Washington, DC, 1997; 2 pp.

State Legislative Fiscal Offices and State Government Administration

Anita Chadha

Department of Social Sciences, University of Houston - Downtown, Houston, Texas, U.S.A.

INTRODUCTION

The Legislative Fiscal Office (LFO) serves as an important support agency for legislatures in all 50 states. Nonetheless, relatively little research has been done on their operations, authority base, or duties and functions. It is generally recognized that LFOs provide financial data to legislators enabling them to make more informed decisions. The responsibilities of LFO may include development of legislative fiscal notes (LFNs), short assessments of revenue and expenditure implications of proposed legislation or analysis of executive agency estimates of such implications. Some LFOs assist legislators in developing alternatives to the executive budget; others have responsibility for research reports, and program and management evaluations.

There is interstate variation in the structural arrangements of LFOs. Some states have two separate offices (dual LFOs)^a serving the upper and lower legislative chambers, respectively. Other states have a single office (joint LFOs) which serves both chambers. Dual LFOs tend to perform more functions and utilize more resources than joint LFOs,^[1] suggesting that competition between chambers and their fiscal agencies leads states with dual offices to expand their scope and to increase their support. Regardless of form, the principal goal of the Legislative Fiscal Office is to provide factual, objective information to legislators in a timely fashion.

BACKGROUND

Research on the LFO specifically and legislative staffing generally has been limited,^[2,3] dated, and primarily focused on single-state case studies done in the 1970s. There have been few comparative studies of legislative staff,^[4,5] which is unfortunate. In recent decades, legislative staffs have grown rapidly in size and functions, making them important forces in the development of state policy.

To provide comparative data on legislative fiscal offices, several questions regarding their performance are addressed in this article: What are the duties assigned to the LFO by state law or legislative rules? Which duties receive priority? What is the LFO workload? What are their staffing levels? Are there functions that LFOs have assumed on their own?

Data on duties, functions, and workloads of LFOs were collected in a two-phase survey conducted in 1997 and 2000.^[1,6] The surveys included questions on the amount of time spent on functions, numbers of support staff employed, estimates of the numbers of fiscal notes, research reports, program evaluations, and other analyses performed. Additional questions focused on those duties and functions dealt with in the past years.

DUTIES PERFORMED

Survey results indicate that state law and/or legislative rules can assign LFOs authority to perform a variety of budget-related tasks, including writing fiscal notes for bills, forecasting revenues, evaluating the executive budget(s), developing alternative budgets, and assisting in the development of legislative budgets. While state law or legislative rules may assign LFOs authority to undertake these varied tasks, not all states perform all authorized tasks. Conversely, in a few states where no legal authority or legislative rule exists, the right to perform a task has developed over time. To assess the relative importance of each duty or task, an estimate was made on the amount of time the LFO spent on budget development, fiscal notes, research reports, program and management evaluations, and staff support.

Budget Development

Not surprisingly, budget development received the most attention among the duties performed during regular sessions. Three-quarters of the responding LFOs reported that one-third or more of their time is devoted to this task. However, the number and type of budgets developed vary greatly. In some states one unified budget is developed; in others, two or more separate

^aThese states are Florida, Massachusetts, Michigan, Minnesota, New York, Oklahoma, Pennsylvania, and Rhode Island.

budgets deal with specific aspects of state operations. Over 91% of LFOs performed some budgetary function. There are differences in time spent on budget development, when budgets were framed (whether during regular, special, or between sessions), and when the different kinds of budgets were considered.

There are many reasons for this variation. For instance, differing constitutional, statutory, and judicial constraints govern the number of budgets passed for each state. Also, state legislatures differ in the number of budget bills that are introduced and considered. States may have separate committees in each chamber with some exercising the responsibility for both revenue and spending matters and others having the sole responsibility for one or the other.

While 20 state legislatures entertained only a single budget bill, legislatures in the other 30 states considered from 2 to 350 separate budget bills. By way of illustration, the Oregon LFO assisted in developing 126 different versions of unified budgets, while the Oklahoma upper house LFO assisted in developing 156 general fund budgets. The LFO in Alabama (one of six states that do separate education budgets) assisted in 10 versions of that budget. The Alabama LFO is also one of eight that assist in add-on or special budget development. On the other hand, several LFOs, such as those in Colorado, Georgia, Maine, and Virginia, are not required to assist in developing unified, general, special, or add-on budgets.

Fiscal Notes

Legislative fiscal note drafting is an aspect of budgetary activity historically identified with LFOs. It is second in importance only to budget development as a duty performed during regular sessions. The fiscal office generally prepares a written fiscal note for each general bill that will have an impact on state or local government finances. The fiscal note is designed to provide information as to the fund or agency budget that will be affected, in what manner, and by how much. The aim of the fiscal note is to give the best estimate of a bill's fiscal impact to legislative committees and the full legislature prior to consideration of that bill.

Legislative Fiscal Offices vary in the number of fiscal notes written (ranging from 0 to 7361) depending on formal requirements. Some LFOs, such as those in Alaska, Arkansas, Georgia, Kansas, Minnesota, Virginia, Washington, and West Virginia do not write fiscal notes at all. The number of fiscal notes is affected by the number of bills introduced in a session and the specific details on how fiscal notes are treated and counted. For example, a bill can have more than one fiscal note depending upon the number of times the bill is amended. In some states, as a bill moves from

chamber to chamber, new fiscal notes are written. But, in other states, no fiscal notes are required of a House bill that does not affect state revenues or expenditures, though Senate bills are required to have fiscal notes.

The number of notes reported varies for several reasons: states allow or require fiscal notes on all bills, others require them for all bills reported from committee, and still others require them when requested from a named official. In some states without statutory or rule authority fiscal notes may be produced according to the logic that, if law or rule neither authorizes nor denies such authority, its use is discretionary.

Duties are assigned to LFOs by state laws and legislative rules. The fact that 52% of LFOs responded to the 1997 survey and 62% responded to the 2000 survey indicated that state law and/or legislative rules authorized legislative fiscal notes. Overall, fiscal notes are produced by more fiscal offices than reported here, as some LFOs write fiscal notes only for bills reported out of committee, or when requested.

Research Reports

Another general responsibility of legislative fiscal offices is the writing of research reports. The relative importance of reports varies; some LFOs do considerable research (the Oklahoma lower house LFO produced more research reports than any other office in all categories), while others do none at all. Reports are done in response to requests from several sources. Survey responses in 1997 and 2000 show that 82–88% of the reports were for individual legislators and committee chairs. In 1997, 79% of the LFOs, and, in 2000, 84% of the LFOs state that they have authority to initiate their own research reports. Offices in two states, Georgia and West Virginia, can produce reports for the governor. In both 1997 and 2000 committee chairs were the group most likely to have the authority to ask for LFO research reports.

Evaluations

Monitoring programs to ascertain their compliance with legislative intent and statutory authority may be another function of legislative fiscal offices. The monitoring frequently involves efficiency and effectiveness measurements to determine if programs are accomplishing specified goals and objectives. Evaluations can be program or management based. In 1997, the Nevada LFO did the most program evaluations reaching 450.

In some 10 states, LFOs have done program evaluations that were requested by legislators, legislative officers, committee chairs, or the self-initiated. According to survey results, authority to do program evaluations for individual legislators was done by 54% of LFOs in

1997 and 51% in 2000. Authority for evaluations for legislative officers was available to 49% of LFOs in 1997 and 58% in 2000; for committee chairs, 44% in 1997 and 66% in 2000. None of the legislative fiscal offices did program evaluations that were requested by the governor.

Staff

Adequate professional and support staff is an important determinant of workload and performance by legislative fiscal offices. With 113 personnel, Texas had the largest full-time staff, followed by the New York Assembly with 40, Colorado with 39, and Virginia with 35. The smallest staffs were in Vermont and West Virginia (four members). The Texas LFO also had the largest full-time secretarial staff with eight; LFOs in Alaska, Vermont, Washington, West Virginia, and Wisconsin had none. As for clerical staff, the Texas LFO again led with eight and again several LFOs had none.

CONCLUSIONS

Assessments of the work of Legislative Fiscal Offices is important for advancing basic knowledge of state legislative performance and interstate variations in institutional capacity. In this article, the workloads

and functions performed by LFOs have been examined. Explanations have been offered for differences in workload and functions over time structure of LFOs (joint vs. dual), difference in responsibilities during and between legislative sessions, and legal authority to perform specific tasks. It is now possible to study systematically the implications of differences between and among LFOs in support to state legislatures.

REFERENCES

1. Chadha, A.; Permaloff, A.; Bernstein, R. The consequences of independence: functions and resources of state legislative fiscal offices. *State Local Govern. Rev.* **2001**, 33 (3).
2. Balutis, A.P. Legislative staffing: a review of current trends. In *Legislative Staffing: A Comparative Perspective*; Heaphey, J., Balutis, A.P., Eds.; John Wiley & Sons: New York, 1975; 23–28.
3. Oppenheimer, B.I. How legislatures shape policy and budgets. *Legislat. Stud. Q.* **1983**, 8, 551–597.
4. Mezey, M.L. *Comparing Legislatures*; Duke University Press: Durham, NC, 1979.
5. Hammond, S.W. Legislative staffs. In *Handbook of Legislative Research*; Loewenberg, G., Samuel, C.P., Malcolm, E.J., Eds.; Harvard University Press: Cambridge, 1985.
6. Chadha, A.; Permaloff, A.; Gentle, P. What do state legislative offices do? *J. Pub. Affairs Issues* **2000**, 4 (Fall), 63–86.

Strategic Management and Productivity

Arie Halachmi

*Institute of Government, Tennessee State University, Nashville, Tennessee, U.S.A.
and Zhongshan University, Guangzhou, P.R. China*

INTRODUCTION

Strategic management is the effort to capitalize on the strengths of an organization by taking advantage of favorable conditions inside or outside the organization. It involves minimization of cost by making operations consistent and predictable, and by avoiding challenges to structure and the organizational culture. Thus, it is the process for implementing a strategic plan that integrates the organization's goals, policies, and action sequences into a cohesive whole. That plan, as an expression of strategic choices, is the product of a carefully elaborated and orchestrated decision-making process. According to Quinn,^[1] strategic decisions:

- Intimately shape the true goals of the enterprise
- Help to delineate broad limits of operations
- Dictate both the resources of the enterprise that will be accessible for its tasks and the principal pattern in which these resources will be allocated
- Determine the effectiveness of the enterprise—whether its major thrusts are in the right directions given its resource potential—rather than whether individual tasks are performed efficiently

Strategic management has the potential to affect organizational performance, but it is not a guarantee of success or productivity. It can easily go wrong, leading organizations astray. Regardless of its results, strategic planning always involves the consumption of a lot of precious organizational resources that could have been used to underwrite other important organizational activities. For example, the time, creativity, and budgets used for developing strategic plans could have been used to provide employees with cross-training or to build team spirit. Such activities have the potential for improving organizational performance just as strategic planning does. Yet, when strategic planning is done right and the organization is capable of taking advantage of it, it has much to offer. The complexity and the rapidly changing circumstances under which public administrators must operate make strategic management a necessity. Recent advances in information technology make the development and use of strategic plans easier than it used to

be. Further, legislation resulting from efforts to increase accountability and performance, such as the Government Performance and Results Act of 1993 in the United States^[2] or Best Value in the United Kingdom,^[3,4] made the development of strategic plans, as well as annual reporting on the effort to implement them, a mandate.

The conceptual framework for strategic planning and management can be a potent tool for administrators because it facilitates systematic thinking. The strategic planning phase helps managers to relate assumptions (about the environment, desired results, availability of resources, and capacity) to a blueprint for action. The enhanced capacity to manage the organization strategically results from several things.

First, the strategic plan provides all members of the organization with a common frame of reference for taking care of their daily choices among competing options for action. As such, a good strategic plan can serve as a great coordinating mechanism for all management levels. Asking all employees to rank-order such options according to how they can contribute to the ultimate organizational objective is important it helps to reduce the odds that the wrong action will occur because on its face it seemed to serve best the interests of a subunit or a short-time need. In other words, the availability of a strategic plan can reduce the odds that the organization will miss its target due to suboptimization.

Second, the strategic plan defines necessary feedback and feedback loops as a function of programmatic (i.e., the need to know) rather than ceremonial (i.e., the right to know) considerations. The strategic planning establishes precontrols over the discretion of lower-level managers by defining important values, desired results, and relevant considerations. This, in turn, contributes to consistency in the handling of similar decisions by different individuals and the various organizational units, and it increases the predictability of operations.

This article describes the main characteristics of strategic planning and management, and points out how they can influence and contribute to productivity. The article starts with a discussion of two basic ways of approaching strategic planning. It goes on to describe

the various steps of a simple model of strategic planning and some of the necessary conditions for getting the most out of it.

MISSION STATEMENTS: TWO APPROACHES TO STRATEGIC PLANNING

A mission is the ultimate goal, philosophy, or purpose of an organization. In the case of a private organization, the mission provides the answer to such questions as what is our business? or what beliefs or values should guide us? A company's mission is defined as the fundamental, unique purpose that sets a business apart from other firms of its type, and identifies the scope of its operations in product and market terms.^[5] The missions of public agencies, however, justify their right to exist and use public resources. Sometimes the mission of a public agency is written into the enabling legislation that created the agency, but often legislation such as GPRA at the federal level or an executive order by a governor at the state level can mandate the development of a mission statement or articulate it in connection with an effort to review performance. Governor George H. Ryan Executive Order no. 7 (March 1999) "Creating the Office of Statewide Performance Review" for the state of Illinois is one possible example that illustrates this point.^[6]

According to Byars, without a concrete statement of purpose, it is virtually impossible for any organization to develop clear objectives.^[7] If this assertion of Byars is true, one must wonder whether a public manager can be successful in an effort to develop a strategic plan and to implement it. After all, in the public sector, developing a mission statement and deriving clear objectives from it are much more difficult and complicated than they are in the private sector.

The missions of agencies are sometimes set by various pieces of laws and court decisions, or within a specific law or a court decision. In such cases, managers have little discretion and agency employees may have little, if any, input, even though they may be the experts whose advice the agency should seek. In theory, when a mission is not articulated for an agency by outsiders (i.e., the legislature or the courts), managers will have more of a say about how to develop it. In the absence of a mission statement created by legislation or court decision, newly elected officials and appointed administrators have some flexibility to chart their preferred courses of action and to break away from the priorities and practices of those who used to be in office. In practice, however, this discretion may be taxing. When public managers need to derive a mission statement from multiple sources, they are likely to find the cards are stacked against them. The guidelines they may be able to derive from various

sources may be either too broad, too vague, or inconsistent with other regulations that govern operations (e.g., the directives of central staff agencies such as the Office of Management and Budget or the Government Accounting Office). Under these circumstances, they do not have the convenience of saying "this mission was written into the law (or court order)" while having to come up with excuses for all the things they wanted to include in the mission statement but could not.

The purpose of the mission statement in public agencies is to allow the formation of a winning coalition to support acceptance of the strategic plan by stakeholders both within and outside the agency. The general language, which allows members of the coalition to give it different interpretations, is often the only way of making the mission acceptable to various stakeholders. A mission that is too concrete may deprive the coalition from the support of some important stakeholders. Thus, one reason for the difficulty public agencies have in enunciating clear objectives is that they need to be derived from a mission statement that was articulated to serve a political function—to mobilize support for the strategic plan—rather than to facilitate the actual planning process. Thus, the mission statement that can bring unity and conformity in the case of a private organization may bring diversity and create ambiguity in the case of government agencies.

In the public sector, participation in the process of developing a strategic plan and the document that results from it can reduce the haze and cultivate a common understanding of the mission across the board. Although private sector organizations can secretly assign the development of strategic plans to ad hoc teams or specialized units, strategic planning in the public sector must be open and involve broad-base participation. Such a planning process is needed in the public sector for two reasons. First, it may accord the strategic plan its legitimacy. Second, such an open process is needed for fostering a sense of ownership of the resulting plan by those whose support and cooperation are critical for its implementation.

It is possible to identify two major schools of thinking about strategic planning and the mission of the organization.^[8] One school views the strategic planning process as an analytical exercise that results in the definition of a mission for the organization. Byars, for example, claimed that an organization's purpose must be defined at its inception, and then be redefined regularly during both difficult and successful times.^[7]

The second school of thought assumes the pre-existence of a mission. According to this second approach, the purpose of strategic planning is to identify the preferred way of carrying out a given mission, optimizing the use of resources by adopting an outward perspective.

Subscribers to the first approach assume that those involved in the strategic exercise are empowered to define and redefine important parameters of the system. The analysis results from concerns about past, present, or future performances of the organization. The only guideline for developing the plan is that it should lead to a better future. In 1787, the Founding Fathers in Philadelphia approached their task in this manner. They redefined the mission of the central government and articulated some important values for the purpose of forming a better union. The values and the definition of the mission were then used as parameters for devising a mechanism for carrying out the mission and for preserving the union as they envisioned it.

A SIMPLE MODEL OF STRATEGIC PLANNING

Developing the Metaplan

The first step in the planning process calls for a general investigation of the internal and external environments of the agency. The purpose of the investigation is to establish the boundaries of the relevant environments. What issues, considerations, organizations, constraints, groups, or individuals should be included in the analysis, and what should be left out? What is an adequate timeframe for making and using projections in the plan? What is within the domain of the agency, and what should be considered outside it? What forces exist in the agency's immediate (i.e., operating) environment, and what forces shape this environment by influencing the general environment?

Dror stressed the need to establish a clear framework for action that is commensurate with the authority and resources of the agency. He used the term cutoff points, or horizon, to delineate time, territory, and spheres of social activity beyond which the effect of a policy can be ignored.^[9]

One way of arriving at the cutoff horizon is by "stakeholder analysis."^[10] A preliminary stakeholder analysis means the identification of coalitions of groups and individuals within and outside the organization who have a stake in the status quo. The analysis next requires the identification of issues that are important in the present and those that will become important within 5 to 10 years. Then the potential influence and interest of each stakeholder in issues of importance, in the present or in the future, can be estimated. Those coalitions, groups, or individuals who do not seem to have a direct or long-term stake in what the agency is doing, or in issues that are important to the agency, may be excluded from the analysis. The same is true for issues that are not salient on the agenda of any important group.

A cutoff horizon can reduce the strategic planning effort to a manageable task, but this benefit is not without risk of ignoring some important issues, groups, or individuals. Some of the risk results from the fact that groups and individuals may relate to a given issue in more than one capacity.^[10] It is not easy to estimate whether a group or an individual is consistent in reactions to a given issue as the capacity of involvement changes. For example, employees of the state department of education may have different reactions to a new tax to finance education in their capacities as taxpayers, parents of school-age children, or department employees.

To reduce the possible consequences of such an oversight, the organization should reexamine the assumptions and considerations that were used to frame the boundary of the analysis. Such examinations should be woven into the planning effort itself as part of the planned iterations. The strategic plan should include set intervals for reevaluation of said assumptions during the implementation of the strategic plan. Such examinations may coincide with the development or review of proposed budgets.

The purpose of the investigation is to establish the backdrop and the framework for the strategic planning process. Even though it is usually limited in scope, it should give the individuals in charge of designing the planning process some insight into the issues that should be dealt with during the planning effort. In other words, the metapanning phase is the time when the organization sorts the wheat from the chaff by determining what constitutes a strategic issue and what does not. The importance of the investigation for strategic issue diagnosis is that it delimits and constrains subsequent information collection, evaluation, and choice.^[11]

Bryson suggested that "identifying strategic issues is the heart of the strategic planning process." He pointed out that the purpose of this step is to identify the fundamental policy choices facing the organization.^[12] The model presented in this article uses a different approach to avoid some of the difficulties that were identified by Bryson. The effort to identify strategic issues is an attempt to define the forces that will shape an agency's environment and influence its freedom of action. The model assumes that it is only after a force field analysis has been completed that the agency may be in a position to identify "fundamental policy choices." Using this approach, which allows for redefinition of the force field or any of its elements, the model creates an environment that is conducive to an orderly development of the strategic planning process. The model makes two important assumptions. First, it assumes that any disagreement about the forces that is going to affect the agency is going to be less trying for the participants if it is divorced from

the effort to select a course of action for dealing with them. Second, the model assumes that participants in the planning process are more likely to cooperate if the initial working paper, which contains a tentative, preliminary inventory of critical issues, is produced by the coordinator of the planning process rather than by one of them. Knowing that the inventory can and is likely to be revised as the planning process progresses, and not having to defend a position that was taken before the beginning of the process, can induce cooperation. In other words, by developing a working paper before the beginning of the planning process, the coordinator of the process allows the participants to question and research the issues without attacking or putting each other on the defensive.

The investigation is usually part of the metaplaning phase because its results should influence the scope, timetable, and organization of the planning process. To use an analogy, the investigation during the metaphase of the planning process corresponds to the work of a Grand Jury. In each case, the products establish the case for subsequent action, providing the justification and the framework for a more comprehensive effort of data collection and analysis.

The metaplaning effort serves two purposes. First, it develops a model of the proposed planning process. The model is a logical presentation to help participants to understand and anticipate the various stages of the process, and to see how each step of the process will lead to the other, and how and why each step forward requires a review of implications for earlier work. The second purpose of the metaphase is to estimate the necessary time and resources for completion of each step, control, and coordination mechanism for facilitating the process (e.g., PERT (Project Evaluation and Review Technique) diagram), and the apparatus for evaluating the quality of the strategic plan and the process for its development. In essence, the purpose of the metaphase is to provide the individual in charge of strategic planning the necessary information to carry out the planning process and to evaluate it.

Taking Stock

The second step of the strategic planning process is a comprehensive, indepth collection of data. With advances in information technology, most of the effort may involve retrieval of information from sources inside and outside the organization. The convenience and efficiency of this option should always be evaluated against the possible error, noise, or built-in bias that are likely to result from the use of data that were initially collected for other uses.

The purpose of this second step is to take a snapshot of the relevant “reality”—the proper context for

assessing the existing situation and alternative courses of action. To facilitate this step, the leader of the planning effort should consult with prospective users and participants to define the set of questions that should guide the collection of data. The questions may be grouped into distinct categories such as:

- What are the legal requirements, mandates, or constraints that must be observed by the agency? What authority does it have? What is the scope of its discretion in exercising this authority? What other bureaus can influence or be influenced by what the agency is doing? What is the relationship between all of these and the evolving economic, political, and social trends?
- What is the relationship between past and present performance, and the interest of various stakeholders? What is the common interest of coalitions, groups, or individual stakeholders in each policy arena? How strong is the interest of each member of such a coalition in what the agency is doing (or, in how it does it)? What is the likelihood that any of them may change their stated position and join another coalition? What may cause such a change?
- What are the indicators of a popular desire to see a change in the formal definition of the agency’s mission? What indicates a desire for a change in the agency’s formal structure or administrative practices as established by the enabling legislation? What forces are seeking to affect interpretation of the law, informal practices, and the use of discretion?
- Using professional standards, what can be said about subcomponents of the agency or elements of the program(s) under its jurisdiction? What are the horizontal and vertical interdependencies among subunits? What is the functional nature of these interdependencies,^[13] and which ones are critical? Do top administrators, rank-and-file employees, preliminary service recipients, and other stakeholders recognize and agree on the existence of interdependencies?
- What is the reputation of the agency among legislators? the media? central staff units that regulate and determine allocation of resources like budgets or personnel? groups other than the primary stakeholders?
- What are the expectations and preferences of top agency officials about issues such as substantive results, use of resources, organizational structure, administrative practices, relationships with various stakeholders, and new initiatives?
- What are the possible implications of changes in office and information technology given current political, social, and economic trends?

- When it comes to the services the agency provides different publics, what are (and what should be) the relationships among variables such as capacity, quality, queuing, equality (equity), automation, staffing levels, client group awareness, multisite operation, use of vendors, and user fees?

Strategic Analysis

The third step of the analysis involves a critical review of the answers to all the questions about the agency, its operations, structure, the way it relates to the external environment in the present, and the way it should relate in the future. The product of the third step of the strategic planning process is a list of alleged strengths and weaknesses. A strength is a virtue (e.g., well-trained personnel or relevant past experience). A strength indicates a capacity to operate without the presence or the prospect of interference. A weakness is a deficiency. It indicates a threat to organizational capacity or the value of the goods or services it provides, unless there is a change inside or outside the agency. Low morale, burnout, poor image, inadequate facilities, lack of experience, or insufficient resources are common reasons for weakness in public agencies.

The disclosed strengths of the agency establish the desired direction for developing alternatives for action. The weaknesses reveal the constraints on action and areas where change is necessary. The exact definitions of what constitutes a strength or a weakness must be developed during this stage and requires direct input from top administrators. The reason is that what may constitute a weakness for one may be a source of strength for another. For example, one chief of police may consider a public image of a trigger-happy police force as a helping factor due to its contribution to deterrence. Another chief may consider such an image as an obstacle to effective action because it reduces trust and communication with minority groups.

Alternative Generation

The fourth step of the strategic planning process is identification (or generation) and comparative analysis of alternative courses of action. Using the strengths and weaknesses that were identified in the previous phase of the planning process as guidelines, the organization seeks ideas about what can be done and what should not be done, and why. To identify as many ideas and alternatives as possible, the metaplan for the strategic planning process may call for the simultaneous use of several independent efforts. Some of the techniques for generating ideas through mobilization of insights and different kinds of knowledge include the use of expert systems,^[14] small (focus)

groups,^[15,16] brainstorming sessions,^[17] use of Delphi or simulations, consultants, advisory groups, and electronic bulletin boards.

There are four reasons for the simultaneous use of several of these techniques. The first reason is to assure that no viable option is ignored. Different techniques may be more suitable for helping various individuals and groups articulate insights and desires about the agency and its management or programs. Second, the effort to establish an inventory of options should facilitate the collection of information and opinions so that the virtues and implications of each option are not over- or underestimated. Use of different techniques reduces the odds that an important consideration will be overlooked. The third reason is that simultaneous use of different techniques can speed up the process and prevent a stalemate. The fourth reason is that by use of alternative techniques and by involving different groups, the organization increases the odds for overcoming cognitive deficiencies, which can mar the initial identification of strategy and the development of the metaplan.^[18,19]

The various alternatives for interpreting the enabling legislation (the mission), the options for organizing the agency (the grand strategy), and possible administrative practices (functional strategies) should then be ranked on the basis of two considerations: 1) the extent to which they take advantage of agency strengths and avoid its weakness; and 2) the extent that they are in line with the preferences of important stakeholders. Ignoring either consideration may defeat the whole purpose of developing a strategic plan.

Strategic Choices

In the fifth step of the planning process, the organization chooses the options for implementation. This includes commitments to achieve specific annual objectives and to follow particular functional strategies and policies, institutionalizing the strategy and establishing of control.^[5] In selecting an option for implementation, the organization determines the following:

- What is the agency out to achieve?
- What will be done to achieve it?
- Who is responsible for each element of the program?
- What is the timetable for carrying out the plan?

Organizational Leadership and Support

The implementation of any plan requires preparation. The purpose of this sixth step of the strategic planning process is to tie the necessary preparation to the planning process itself. Lack of resources or any other

constraints on making the necessary preparations should be taken into consideration and influence the development of the plan.

Sometimes the preparation for implementing a strategic plan may call for special training,^[20] purchasing of new services or equipment, modification of facilities, or rewriting of manuals. Some of the preparations may require a lead time. The gap between selection of a strategic plan and the beginning of the implementation process provides a fertile ground for rumors and other activities that may jeopardize the success of the plan.

Therefore, even before preparing to implement the plan, the organization must embark on a campaign to educate all employees and other stakeholders about the plan and how the organization is going to prepare for it. Managers should treat the implementation as an introduction of a planned change. They must be ready to provide leadership and willing to get involved in mobilizing support. According to Holden, if strategic planning is to be a productive management function, support for it must exceed the resistance against it.^[21] It follows that in the absence of strategic leadership, the creation of an overall sense of purpose and direction that guide integrated strategy formulation and implementation,^[22,23] sufficient level of support within the organization is unlikely and the wisdom of engaging in strategic plan should be questioned.

Review and Evaluation

During the seventh step, the organization establishes milestones, and a framework for monitoring and evaluating the implementation process. Establishing this framework in advance of implementation prevents the temptation to rationalize (or excuse) deviations from the plan. It helps the manager to identify such deviations and scrutinize assumptions that were used for developing the plan.

The evaluation framework defines the quantitative and qualitative measures of input, process, and output of each subsystem of the agency. Classifying an administrative subunit as a revenue, cost, or responsibility center^[24] allows the manager to compare expected and actual performance. It also allows the organization to trace the relative contribution of each subunit to the use or generation of resources.

RESPONSIBILITIES AND ROLES IN STRATEGIC PLANNING AND STRATEGIC MANAGEMENT

The strategic planning process can be a bottom-up or top-down planning process. As a bottom-up process, the effort begins with investigation of the

environment by each subunit. The data from all the units reaches a central location that uses it to develop a tentative design for the planning process. This design—the metapanning phase of the strategic planning effort—is then circulated for comments and suggestions before it is finalized. The final design provides each unit of the organization with a timetable, the division of labor, and the various stages of the process to develop the strategic plan.

When the process originates at the top, the initial study is carried out by a central unit that also develops the first draft of the design for the planning process. The draft is circulated for comments before it is finalized.

Selection of a top-down or a bottom-up approach depends on the distribution of strategic planning experience throughout the organization and the availability of resources. Organizations that have gone through the process of strategic planning once may benefit from an action research approach where the planner and those who will be asked to implement it collaborate.^[25] Action research assumes participation by all employees and thus is less biased by the perspective from the top. Such an approach can lead to better definitions of issues and considerations than an approach that limits participation. However, when prior experience of strategic planning is missing or when the organization cannot spare the resources to train employees before the beginning of the strategic planning effort, subunits may spend too much time pondering and wondering what is expected of them or how to do it. The organization may save time and unnecessary aggravation by using professionals to conduct the initial survey that establishes the parameters or boundaries of the planning process.

The person in charge of the strategic planning process is a broker of knowledge. This person needs to know something about the agency and its special circumstances. However, the person's primary expertise should be in strategic planning—the how and why of going through the various steps of the process.

The person in charge of the strategic planning effort is also a change agent. As such, the person needs to bring to the organization extraordinary leadership and interpersonal skills. The strategic planning process involves scrutiny of past achievements, reordering of priorities, possible shifts and changes of structure, and personnel and administrative practices. All these may threaten employees' careers, job security, seniority, authority, or power, and they can generate tension, particularly in labor relations. As a change agent, the coordinator of the strategic planning effort should be able to smooth ruffled feathers and reduce the uncertainty.

To be successful as a change agent, as a broker of knowledge, and as coordinator of the planning process,

that individual must have the support of top officials and easy access to the ranking administrator. All agency employees should know that the individual's authority to request cooperation or data is derived from the authority of the chief administrator. The administrator should spare no effort to communicate to the rank and file the importance of the strategic planning effort and the commitment of the organization to see it through. Without clear and strong commitment from the top, employees may not take the strategic planning effort seriously. This, in turn, may lead to delays or may compromise the quality of the data being collected by employees for the planning effort.

The importance of organizational commitment cannot be emphasized enough: without it, the whole effort does not stand much of a chance. After all, for most agency employees, participating in and contributing to the planning process is a duty that is added to their regular responsibilities. Most employees like to be involved, but resent the extra work they need to do just for the planning effort. Thus, special attention should be given during the design of the strategic planning effort to building good will.

In a similar way, the organization should spare no effort when familiarizing employees and stakeholders with the strategic plan, the rationale behind it, and how its various parts are related. Such efforts should reduce ambiguity, and counteract the influence of rumors and allegations that may interfere with the implementation of the plan. The symbolic value of educating employees and the public about the plan will help the organization to legitimize and secure cooperation for implementing the plan.

CONCLUSION

Halachmi^[8] questioned whether strategic planning and management are likely to be as useful or feasible in the case of all public agencies. In one publication,^[26] he asserted that well-performing organizations are likely to be able to take advantage of strategic planning and management, whereas weaker agencies may not. In the early 1990s, Mintzberg, who wrote extensively on the subject of strategic planning and management, suggested^[27] that the inherent characteristics of planning and strategic management may undermine one at the expense of the other. According to Mintzberg, planning involves analytical and systematic analysis and thinking, whereas strategic management involves a synthesis, learning and thinking outside the envelope. In his words: "Strategic planning is not strategic thinking. Strategic planning often spoils strategic thinking... this confusion lies at the heart of the issue: the most successful strategies are visions, not plans."^[27]

The implications of the observations offered by Halachmi and Mintzberg are clear: strategic planning and management may be a useful tool for all managers, but they may not be suitable for all organizations.

REFERENCES

1. Quinn, B.J. Strategies for change. In *The Strategy Process*; Quinn, B.J., et al., Eds.; Prentice-Hall: Englewoods Cliffs, NJ, 1988; 3 pp.
2. Halachmi, A. Promises and possible pitfalls on the way to SEA. In *Organizational Performance and Measurement in the Public Sector*; Halachmi, A., Bouckaert, G., Eds.; Quorum Books: Westport, CT, 1996; 77–100.
3. Halachmi, A.; Montgomery, V. Best value and accountability: issues and observations. *Int. Rev. Adm. Sci.* **2000**, *66*, 393–414 (September).
4. Bovaird, A. Best value in the united kingdom: using benchmarking and competition to achieve value for money. *Int. Rev. Adm. Sci.* **2000**, *66* (3), 415–432, September.
5. Pearce, J.A.; Robinson, R.B. *Strategic Management*; Irwin: Homewood, IL, 1988.
6. *Ryan Starts Statewide Performance Review and Establishes Office of Strategic Planning*; March 10, 1999; <http://www.state.il.us/gov/press/99/Mar/exec7&8.htm>.
7. Byars, L.L. The strategic management process: a model and terminology. *Manag. Plann.* **1984**, *32* (6), 38–44, May.
8. Halachmi, A. Strategic planning and management? not necessarily. *Public Prod. Rev.* **1986**, (vol. 10 no. 1, issue no. 40), 35–50 (Winter).
9. Dror, Y. *Public Policy Re-examined*; Chandler: San Francisco, 1968.
10. Freeman, R.E. *Strategic Management: A Stakeholder Approach*; Pittman: Boston, 1984.
11. Dutton, J.A.; Walton, E.J.; Abrahamson, E. Important dimensions of strategic issues: separating the wheat from the chaff. *J. Manag. Stud.* **1989**, *26* (4), 379–397, July.
12. Bryson, J.M. *Strategic Planning for Public Nonprofit Organizations*; Jossey-Bass: San Francisco, 1988; 139.
13. Wintrobe, R.; Berton, A. Organizational structure and productivity. *Am. Econ. Rev.* **1986**, *76* (3), 530–538, June.
14. Ashmore, G.M. Applying expert systems to business strategy. *J. Bus. Strategy* **1989**, *10* (5), 4650, Sept./Oct.
15. Morgan, D.L. *Focus Groups as Qualitative Research*; Sage: San Francisco, 1988.
16. Stewart, D.W.; Shamdasani, P.M. *Focus Groups: Theory and Practice*; Sage: San Francisco, 1990.
17. Moore, C.M. *Group Techniques for Idea Building*; Sage: San Francisco, 1987.
18. Stubart, C.I. Managerial condition: a missing link in strategic management research. *J. Manag. Stud.* **1989**, *26* (4), 325–328, July.
19. Haley, U.C.V.; Stumpf, S.A. Cognitive trails in strategic decision making: linking theories of personalities and cognitions. *J. Manag. Stud.* **1989**, *26* (5), 477–498, September.

20. Gainer, L. Strategy: a primer. *Train. Dev. J.* **1989**, 43 (9), S5–S30.
21. Holden, A.L. Strategic planning in healthcare systems raises new management concerns. *Healthc. Financ. Manage.* **1985**, 39 (10), 50–60, October.
22. Shrivatava, P.; Nachman, S.A. Strategic leadership patterns. *Strategic. Manage. J.* **1989**, 10, 51–61.
23. Westley, F.; Mintzberg, H. Visionary leadership and strategic management. *Strateg. Manage. J.* **1989**, 10, 1732.
24. Anthony, R.N.; Young, D.W. *Management Control in Non-profit Organizations*, 4th Ed.; Irwin: Homewood, IL, 1988.
25. Halachmi, A. On the job training of social workers. *Int. J. Soc. Work* **1980**, 12 (1), 33–44.
26. Halachmi, A. Strategic management for productivity, or is it the other way around? In *Handbook of Strategic Management*, 2nd Ed.; Rabin, J., Miller, G.J., Hildreth, W.B., Eds.; Marcel Dekker: NY, 2000; 343–370. Ch. 15.
27. Mintzberg, H. The fall and rise of strategic planning. *Harvard Bus. Rev.* **1994**, 72 (2), 107–114, January–February.

BIBLIOGRAPHY

1. Chandler, A.D. *Strategy and Structure: Chapters in the History of the American Industrial Enterprise*; MIT Press: Cambridge, MA, 1962.
2. Fredrickson, J.W. The strategic decision process and organizational structure. *Acad. Manage. Rev.* **1986**, 11 (2), 280–287.
3. Lenz, R.T. Environment, strategy, organization structure and performance: patterns in one industry. *Strateg. Manage. J.* **1980**, 1, 209–226.
4. Miller, D. Strategy making and structure: analysis and implications for performance. *Acad. Manage. J.* **1987**, 30 (1), 7–32, March.
5. Mintzberg, H. The structuring of organizations. In *The Strategy Process*; Quinn, B.J., et al., Eds.; Prentice-Hall: Englewoods Cliffs, NJ, 1979.
6. Morone, J. Strategic use of technology. *California Summer* **1989**, 31 (4), 91–101.
7. Sherman, J.D.; Smith, H.L. The influence of organizational structure on intrinsic versus extrinsic motivation. *Acad. Manage. J.* **1984**, 27 (4), 877–885, December.

Subnational Counter-Cyclical Fiscal Policy in the United States

Yilin Hou

*Department of Public Administration and Policy, University of Georgia,
Athens, Georgia, U.S.A.*

INTRODUCTION

This article is a review of the counter-cyclical fiscal policy (CCFP) at the subnational level in the United States. The term “subnational” is used in contrast to “national,” referring exclusively to the 50 states. Although the literature and data used are restricted to the United States, the academic and practical implications go beyond this country because budgetary stabilization at the subnational level has been a universal topic, bearing significance in many other countries as well.

This article is organized as follows: “Evolution of Theory” examines the origin and development of CCFP in the United States, which, in a period of about 70 years, presents many lessons. “The Practice of CCFP by State Governments” looks into the practice of CCFP by state governments, covering the choice and shift of policy instruments and responses to states’ use of CCFP from the capital market, professional organizations, and the United States Congress. “The Future of CCFP” shifts focus onto the future of CCFP. It will consider the ongoing argument about the economic stabilization function of CCFP as well as the fundamental public good of reducing revenue shocks from economic fluctuations. It ends with the proposition that CCFP is not a panacea; politics and basic budgetary discipline are the ultimate baseline.

EVOLUTION OF THEORY

The study of business cycles has a history almost as long as the market economy in the western world. Prior to the 1930s, the dominant economic philosophy was the classical economic theory. In the 1930s, with the work of John Maynard Keynes emerged the theory and practice of CCFP. The Great Depression led to a thorough examination of the monetary policy as well as fiscal policy at the federal and subnational levels. Scholars found that the federal, state, and local sectors behaved perversely during the Depression because both levels were bound by the requirement to balance the annual budget.^[1]

In 1932, Leland^[2] studied how governments could best meet financial crisis and proposed that “a wise fiscal policy requires . . . a longer term financial program [than the fiscal year] that should take account of the fluctuations of the business cycle and should control the policy of the annual budget.” Leland’s proposal was discussed in 1933 by a University of Chicago roundtable as a method that the federal government could use to balance the budget over economic cycles.^[3]

In 1941, Hansen further expanded this idea. Hansen observed that to promote security for unemployed workers and economic stability during the Great Depression, the federal government at first relied primarily on monetary policy, but the huge number of the unemployed compelled enormous expenditures. Fiscal policy, more by accident than by design, came onto the front stage.^[4] Although an important weapon, monetary policy “has severe limitations and must be supplemented with [fiscal policy, which] includes first a [counter-] cyclically adjusted public spending program and second a [counter-] cyclically administered tax policy.”^[4]

In 1944, Hansen and Perloff extended the argument for CCFP to the state and local sectors: In periods of depression, tax rates should be reduced, and public expenditure should be high by drawing on accumulated reserves and incurring public debt. The policy, in general, shall be directed toward increasing consumer purchasing power and stimulating investment and business activity. For this purpose, tax rates during economic booms should be high to drain off excess consumer purchasing power and to accumulate reserves for revenue shortfalls in the future; public expenditure, in contrast, should be kept low—at a level just adequate to maintain essential social services. Borrowing for further expansion in such periods should be eliminated, and accelerated repayment of public debt should be pursued if the boom is strong.^[1]

Intuitively, the counter-cyclical features of such a fiscal policy make sense, but they are difficult to implement, especially during economic prosperity. Major obstacles may come from three sides. First, states are limited by their constitutions or statutes from

accumulating high levels of surpluses. Even in the absence of legal restraints, individual and corporate taxpayers tend to exert spending pressure on elected officials during booms to expand public expenditures and/or reduce tax rates; and elected officials either share the general optimism or simply follow the voters' will to facilitate their own reelections. Finally, the goals of economic development lead to interstate competition for businesses, which induces, or even dictates, reduction of tax rates.

A subnational government can resort to one or more of the following three means in a recession to implement a CCFP, each with its pros and cons. First, because the federal government has monetary policy and controls more and wider taxing power, federal grants are a valuable financial source. However, obtaining such grants is often unreliable. As experience goes, federal grants may not come as early or at the amount states may need and expect. This lesson was confirmed in the 2001 recession: When the national recession coincided with a catastrophic disaster, federal aid dwindled, and the federal government mandated expenditures that added to states' financial burdens. Besides, where a grant is based on a matching ratio, poor states are hit hard. When a state is in fiscal distress, it often has to cut its assistance to local governments.

The capital market is another means of addressing recession pressures because incurring/increasing public debt in downturns is technically sound. However, recession years are the time when states adopt drastic fiscal measures that adversely affect their credit rating, thus eroding their borrowing capacity on the market. Besides, strict debt limitations as prescribed in state constitutions or statutes do not allow state governments indefinite room for incurring debts.

Then, the only means that is consistently reliable, independent of outside control, is own-source reserves. The availability of reserves depends on two internal factors: first, the legal framework-statutory balanced budget requirements on a (bi)annual basis, which offer little room for flexibility in this context; and, second, the human factor; chief elected officials must possess not only the managerial foresight for the necessity of reserves but also the political courage and skill to resist and manipulate spending pressures during prosperity. Once accumulated, reserves must be guarded against any possible raids in boom years. The best means for counteracting these two internal factors is an institutionalized device that provides statutory protection of the reserves.

The July 1949 "Budget Theory Symposium" dealt with legal framework. Conferees succinctly pointed out that "there is no magic" about the 1-year period of the annual budget; that the most important issue is "stabilization of tax expectance [for businesses];" and that "longer projections and more frequent reviews by the

legislative body are both desirable"^[5] This statement partially softens the theoretical foundation of the annual budget.

Another round of theoretical exploration came in 1987 when Gramlich reexamined macroeconomic evidence and concluded that conventional wisdom may have been inaccurate (states should not only pursue CCFP but, in fact, adopt such practices to stabilize their budgets) and that we can reasonably assume that subnational fiscal policy can exert at least short-term influence.^[6] Gramlich, in 1991, further distinguishes three types of state fiscal policies: neutral, perverse, and stable. The neutral type discourages counter-cyclical policy and advocates automatic fluctuation of taxes and budget surpluses with changes in income. With no counter-cyclical action and no changes to the tax structures over the business cycle, revenues move procyclically. If following a perverse policy, governments would initiate discretionary tax increases and/or expenditure reductions during recessions, cut taxes, and increase spending in boom years. This policy amplifies business cycles fluctuations.^[7]

Gramlich advocates a fiscal policy for stability: Governments smoothen the cycle by raising taxes and/or cutting spending during expansion to accumulate reserves, then by lowering taxes and/or increasing spending during recession by decumulating reserves. The budget would then be balanced over the cycle. This is a stronger version of CCFP, involving accumulating reserves and tax rate changes. No states have adopted this version—although saving reserves are common, tax rate changes, especially permanent ones, are much more difficult and thereby less common because tax increases are politically unpopular. A weaker version without tax changes is what many states have practiced.

THE PRACTICE OF CCFP BY STATE GOVERNMENTS

In the world of administration, New York was the first state to respond to the 1944 Hansen and Perloff proposal by establishing a budget stabilization fund (BSF) to stabilize own-source revenues, with funding from annual surpluses,^[8] but the concept and adoption of BSF spread very slowly among the states until the 1970s. The earliest state BSFs were designed following the weaker version of CCFP—to save annual surpluses in prosperity but not to raise tax rates in boom and then lower the rates in a recession.

Nevertheless, the practice of CCFP had spread out through a wise use of general fund surplus (GFS) accumulated from boom years, evolving the perverse policies of the 1930s into the rational fiscal policies since

World War II. Independent academic research identified this orientation shift in the mid-1960s: "...[T]he patterns of deviations from the trend for state and local receipts were found to have been more and more stabilizing with each succeeding [business] cycle."^[9] A report by the United States Advisory Commission on Intergovernmental Relations (ACIR) in the late 1970s confirms this policy change: "...[D]uring each economic downswing since World War II, state and local fiscal behavior was 'correct' [meaning 'counter-cyclical'] because [spending more from reserves] added to aggregate demand [during downturns]."^[10]

The severe economic and budgetary difficulties of Michigan from 1974 to 1975 and earlier recessions forced the state to seek a permanent solution. Such efforts led to the creation, in 1977, of the Michigan "counter-cyclical budget and economic stabilization fund," which, in comparison with its earlier counterparts in other states, has two new features. First, it places an economic element in the purpose of the fund—to assist in stabilizing not only governmental revenue during periods of recession but also employment when the rate of unemployment is high. Second, the source and use approval of the fund are by a preset formula—the percentage above the 2% annual growth rate benchmark determines the amount to be transferred into the fund in the next fiscal year, and the percentage of negative annual growth rate determines the amount to be used from the fund in the current fiscal year.^[11] The Michigan legislation became a model for many other states. It was "copied" in Ohio, Indiana, and Washington, and it influenced the BSF legislation in many more states. It is not an exaggeration to say that widespread adoption of BSFs among the states started with the 1977 Michigan model legislation accompanied by other social economic factors.

Policy Instruments

Once a CCFP is shown to be rational, state governments face the task of choosing policy instruments that are technically convenient and politically operable while guaranteeing the reserves necessary to handle revenue shocks. The two major ones are the BSF and GFS, although, conceptually, there can be many more possible instruments (some are variations of BSF and GFS in different forms). To practitioners, all forms of reserves, including various contingency funds, can be counted in the BSF category—those that are fine are for one time use and will be returned as soon as the economy recovers. (In fact, the Government Finance Officers' Association (GFOA) lists contingency funds as the same as BSF; see *Recommended Budget Practices—A Framework for Improved State*

and Local Government Budgeting, Section 4.1, Ref.^[12]).

Although we can assume that, in general, economic differences between BSF, GFS, and other forms of reserves do not exist—because total cash position at times of a budget shortfall is the issue^[13,14]—it is important to remember that because budgeting is always closely related to politics, the real test for the above assumption is whether elected officials, and taxpayers collectively, can "administratively and politically" exercise adequate fiscal self-restraint on a routine basis in times of expansion^[15] (i.e., to protect fiscal reserves against current spending pressure for use in a subsequent downturn).

From 1946 to the mid-1980s, the adoption of BSFs among the states was very slow; most states relied on GFS as the policy tool, and even with those states with a BSF, GFS was an indispensable supplement. In the late 1970s, the "tax and expenditure limitation movements," starting with Proposition 13 in California, rapidly spread to other states. A direct consequence of these movements was a strict limit on general fund balance levels. GFS levels plummeted in states with such a limit. In the course of these movements came the severe national recessions of 1980 and 1982, which repeated in state legislatures and state governments, in general, the historical lesson of the chaos that may ensue when inadequate reserves are on hand to ameliorate sudden revenue shocks. Thus the tax and expenditure limitation movements and the recent recessions became the driving force and catalyst in the wider adoption of BSF, triggering a chain reaction among the states. Soon, over a dozen states passed BSF legislation.

The lesson can be best summarized with a quote from the 1981 State of Washington legislation for their "emergency reserve fund:" "The current budgetary system of Washington lacks stability. It encourages crisis budgeting and results in cutbacks during lean years and overspending during surplus years." To safeguard the fund against raids, a very strict procedure was put into place: Appropriations from the fund can be made "only if approved by two-thirds of each house and by a vote of the people at [the] next general election." (Revised Code of Washington 43.135.045)^a Historical data show the shift of state fiscal reserves from GFS to BSF—although the historical average balance of BSFs has increased from the early to the more recent ones, the historical average level of GFS has tended to decrease.

^aRevised Code of Washington, Title 43: State Government—Executive, Chapter 43.135: State Expenditures Limitations (formerly Tax Revenue Limitations), Section 045: Emergency Reserve Fund.

Response from the Congress, Market, and Professional Organizations

Adoption and implementation of CCFP by the states have been welcomed and encouraged by legislators, the debt market, and professional organizations. In 1985, the United States Congress concluded that “the states believe there is a role for [state] government in countering recession, irrespective of Federal action.” Three design issues are prominent: first, state reserve funds have to be large enough to buffer revenue shocks in a recession. There is no uniform level of the reserve proper to all states. The level varies by each state’s cyclical economic fluctuation; second, states need a policy of maintaining reserves of a given size to avoid year-to-year debates about the proper size; and, third, the accumulation and release of reserves should be gradual.^[16]

In 1990, research by the GFOA argues that establishing a fund balance policy and the BSF can alleviate sudden short revenue shocks or deficits.^[13] GFOA advocates CCFP and lists it as a best practice.^[12] Credit rating agencies claim that “maintaining an operating reserve is the most effective practice that can enhance an issuer’s credit rating” and lists it as the first of 12 best practices that have “significant rating value.”^[17] Thus CCFP at the subnational level in the United States has been widely accepted and adopted.

THE FUTURE OF CCFP

In the field of macroeconomic theory, the debate about the stabilization policy is yet to conclude. Some economists hold that variations in governmental revenue flows are caused by a combination of multifactors. Fluctuations in the economy (recession and expansion) are but one of them. Others are changes in the tax structure, additions or discontinuations of taxes, and raising or reducing of tax rates, which are results of not only the economy but also, and more often, election politics.^[18] Because monetary and fiscal policies can influence these fluctuations and offset shocks, policy makers should act to stabilize the economy to keep output and employment close to their natural rates.

However, some other economists doubt whether the government is able to stabilize the economy. On one hand, between the recognition of a recession for state governments and the start of counter-cyclical fiscal programs, there are long and variable lags that are inherent in economic policy making. On the other hand, our understanding of the economy is still very limited and economic forecasting has repeatedly failed our expectations. Therefore it may not be a bad idea for policy makers simply to passively follow a fixed policy rule.

A more seemingly convincing argument about CCFP goes that: assuming stabilization could be

successfully achieved, by the natural rate hypothesis, such a policy can only reduce the magnitude of fluctuations around the natural rate by eliminating the peaks of booms as well as the troughs of recessions so that the average benefit would be small.^[19] Finally, the real-business-cycle theory takes fluctuations as the optimal response of the economy to changing technology, and believes that even if stabilization were possible, policy makers should not pursue this goal.

A direct answer to such doubts is straightforward: Drastic fluctuations of state budgets in history have caused, and will again cause, huge damage and interruption to public life. Revenue shocks from business cycles have repeatedly shown that the subnational sector is prey to economic fluctuations. Public expenditures should not be as subject to cyclical fluctuations as private investments.^[20] Therefore, stabilization of state budgets is a big public good, which is worth every effort to achieve. Scholars and practitioners have also been calling for CCFP for decades. The tide has been rising with recessions and ebbing with expansions—the lesson is still not thoroughly learned.

CONCLUSION

However, a rational fiscal policy and proper policy tools are not a panacea to governmental revenue shocks. The policy and policy tools are but the means that must go through the human hands of implementation to be effective, which in turn lies ultimately in the domain of politics. Specifically, a wise fiscal policy may be brushed aside; appropriate policy instruments can run distorted; and fiscal reserves that are adequate by the most sophisticated formula might as well be used off the right target. Therefore, the solution returns to the very basics of fiscal restraint and budgetary discipline. It would be a misconception that a rational policy can solve the problem once and for all.

REFERENCES

1. Hansen, A.H.; Perloff, H.S. *State and Local Finance in the National Economy*; W. W. Norton & Company, Inc.: New York, 1944; 40–55.
2. Leland, S.E. How governments can best meet the financial crisis. In *Conference Paper: Convention of the International City Managers' Association*; Cincinnati, OH, October 24, 1932.
3. Gideonse, H.D., Ed.; *Balancing the Budget—Federal Fiscal Policy During Depression*; The University of Chicago Press: Chicago, 1933.
4. Hansen, A.H. *Fiscal Policy and Business Cycles*; W. W. Norton & Company, Inc.: New York, 1941; 116, 261 pp.
5. Symposium on budget theory. *Public Adm. Rev.* **1950**, *1* (10), 20–31.

6. Gramlich, E.M. Subnational fiscal policy. In *Perspectives on Local Public Finance and Public Policy*; Quigley, J.M., Ed.; JAI Press Inc.: Greenwich, CT, 1987; Vol. 3, 3–27.
7. Gramlich, E.M. The 1991 state and local fiscal crises. *Brookings Pap. Econ. Act.* **1991**, (2), 249–287. This tripartite theory traces back to A.H. Hansen and H.S. Perloff, 1944.
8. New York State Consolidated Laws: Chapter 56. State Finance Law: Article VI. Funds of the State. Section 92.
9. Rafuse, R.W. Cyclical behavior of state-local finances. In *Essays in Fiscal Federalism*; Musgrave, R.A., Ed.; The Brookings Institution: Washington, DC, 1965; 118 pp.
10. *Counter-Cyclical Aid and Economic Stabilization*; The United States Advisory Commission on Intergovernmental Relations (ACIR): GPO: Washington, DC, 1978; 6 pp.
11. *Act 76 of 1977 (Repealed), and Act 431 of 1984; also Michigan Compiled Laws: Chapter 18. Department of Management and Budget. Sections 351–359.*
12. *Develop Policy on Stabilization Funds*; GFOA: Chicago, IL, 1999.
13. Allan, I.J. Unreserved fund balance and local government finance. In *Research Bulletin—Research and Analysis on Current Issues*; Government Finance Officers' Association: 1990; 4 pp.
14. Sobel, R.S.; Holcombe, R.G. The impact of state rainy day funds in easing state fiscal crises during the 1990–1991 recession. *Public Budg. Finance Fall* **1996**, 16, 30.
15. *States' Use of Surplus Funds*; Congressional Budget Office: Washington, DC, November, 1998; 15–16.
16. *Federal and State Roles in Economic Stabilization*; United States House of Representatives. December 31, 1985; 3–14. 99th Congress, 1st Session, Report 99-460.
17. Richard, P. Larkin Impact of management practices on municipal credit. In *Fitch IBCA Special Report in Its Public Finance Newsletter*; May 4, 2000; <http://www.fitchibca.com> (accessed June 2000).
18. Stonecash, J.M. The revenue problem: revenue fluctuations and forecasting, New York State, 1950–1990. In *Case Studies in Public Budgeting and Financial Management*; Khan, A., Hildreth, W.B., Eds.; Kendall/Hunt Publishing Company: Dubuque, IA, 1994; 153–163.
19. Mankiw, N.G. *Macroeconomics*; Worth Publishers: New York, 1992; 486 pp.
20. Musgrave, R. *The Future of Fiscal Policy—A Reassessment*; Leuven University Press, 1978; 31 pp.

Supplemental Appropriation

Jerry McCaffery

*Department of Systems Management, Naval Postgraduate School,
Monterey, California, U.S.A.*

INTRODUCTION

Supplemental appropriation bills provide funding supplements for the current fiscal year already in progress, over and above funding provided in the regular appropriation bills. Supplementals normally fund disaster relief efforts, emergent military needs unforeseeable in the normal budget process, and occasional emergency economic relief. At the federal level, supplementals usually provide budget authority, which is then committed to programs by obligations. The actual outlay or spending of that money may occur over a longer period than the year for which the appropriations were made. Supplementals are used at all levels of government in the United States. At state and local levels, supplementals usually are passed in the current year and spent out in the current year. In some instances, a supplemental appropriation may become part of the ongoing funding for an agency; usually, however, supplemental appropriations are for specific one-time purposes of an emergency nature.

Before the passage of the Congressional Budget and Impoundment Reform Act of 1974, supplementals were “regularly used to provide funding for day-to-day agency operations,” such as pay raises for federal employees; but, in the recent years, supplementals “have been mainly used to provide funding for unanticipated expenses—although there is sometimes an argument about the whether the requirements should have been anticipated or not.”^[1,2] Supplementals mirror the complexity of the normal budget; they may be spent for agency programs or go directly to individuals, domestically or abroad. They may support direct actions for disaster relief, such as road and bridge building and other infrastructure repair, or they may fund a variety of disaster relief credit and loan programs administered by various federal agencies. States and local governments have come to count on federal government leadership and financial participation in disaster relief and may underbudget from their own funds for these purposes. At the federal level, recent attempts to offset some supplemental costs with rescissions has meant that, while supplemental bills increased funding for some programs, others were decreased, not supplemented. Moreover, recent attempts to hold to spending caps and balanced budget

agreements may have resulted in slight but purposeful underfunding of some of the normal appropriation bills with the difference made up later in a supplemental bill, deemed either a “dire emergency supplemental” and thus beyond any spending cap or offset by a rescission in someone else’s program. The federal government has a long history of using supplementals, and the latter argument over use of supplementals to escape fiscal discipline in the normal appropriation process is not new.

HISTORY

Supplemental spending bills are not new. Supplementals began with the second session of the First Congress in 1790^[3] and were commonly used in the early 1800s. They frequently included additional appropriations for agencies that had overspent their appropriations and, as such, became known as deficiency appropriations. This practice became so routine that the House Appropriations Committee divided them into “general deficiency” bills and “urgent deficiency” bills. In the 1870s when antideficiency legislation was passed, critics of this process accused Congress of underfunding the regular appropriations bills so they would appear to be frugal stewards of public funds, and then, “after elections were over, make up the necessary amounts by deficiency bills,” according to James Garfield, chairman of the House Appropriations Committee in 1879.^[3] The Antideficiency Act of 1905 attempted to control deficiency spending by giving the Treasury Department the authority to apportion funds to agencies to reduce the need for supplementals. This power was further refined by the Budget and Accounting Act of 1921 and the Antideficiency Act of 1950. These revisions limited supplemental appropriations to legislation enacted after the president’s budget had been submitted and for emergencies relating to the preservation of human life and property.

The issues surrounding supplementals have remained contentious. Aaron Wildavsky identified the strategy of “coercive deficiency,” in which an agency would overspend its appropriation secure in the knowledge that Congress would provide supplementals. This strategy was primarily effective for popular agencies

and programs; in the late 1800s, examples of this coercive deficiency can be found in the practices of agencies helping to open up the West through mapping and transportation infrastructure. The strategy was also based on good relationships between Appropriation Committee chairmen and the agency heads; agencies without such relationships or such popular programs could not count on this strategy.^[4] In 1966, the Joint Committee on the Organization of Congress again raised the issue of lawmakers projecting an image of economy by underfunding the regular appropriation bills with the tacit understanding that they would later pass supplemental appropriation bills.^[3] The Congressional Budget and Impoundment Control Act of 1974 also attempted to control the use of supplementals, by providing that expected supplementals be included in the president's budget and that an allowance be set aside in the budget resolution for anticipated supplementals. While these tactics have not been particularly helpful, Congress and the president have tried to offset supplementals with rescissions or cancellations of budget authority provided in earlier appropriations bills that has not been obligated or has been deemed unnecessary. An examination of the 1990s shows that rescissions have made an impact on supplementals, but only in 1995 were rescissions large enough to offset supplementals. In fact, in that year, rescissions were more than the amount appropriated for supplementals.

SUPPLEMENTALS AND BUDGET FORMULATION

Normal budgetary procedures require that budget or appropriation bills be passed before the beginning of the fiscal year, although, in practice, debate over what and how much to fund often results in the passage of these bills after the fiscal year has begun. In contrast to regular appropriation bills, supplementals are quite small as a percentage of the total budget and often result from emergency situations unforeseen in the normal budgetary process that require immediate funding. Customarily, the federal budget process takes 18 to 21 months to complete from the time when agencies begin drafting their budgets (e.g., March 2002) to the time when the president signs the resulting appropriations bills (e.g., legally by October 1, 2003, but usually in November). No such lengthy process could possibly cope with unexpected and unanticipated events. Supplemental appropriation bills provide funding to meet such suddenly emergent needs in the United States and abroad. These bills adjust current year funding levels. They may provide for natural disaster aid to states, regions, and directly to individuals for blizzards, floods, drought, fires, and hurricanes. Supplementals

helped out in the northeast blizzard of 1978; the Mount St. Helens eruption in 1980; Hurricane Hugo and the Loma Prieta earthquake in 1989; Hurricanes Andrew and Iniki, the Chicago flood, and the Los Angeles riots in 1992; the Northridge earthquake in 1994; and flooding in the Dakotas in 1998. Supplementals met emergent military missions in the 1990s beginning with the Gulf War in 1991 and Bosnia in 1996 and thereafter.^a Funds from supplementals have been spent both inside and outside the United States as policy makers have reacted to events suddenly thrust upon them.

SUPPLEMENTALS AND AGENCY BUDGET EXECUTION

Agencies have a certain amount of flexibility in the way they administer their budgets. Congress officially recognized the need for agency discretion in budget execution since the growth of federal expenditures after World War II. Jones and Bixler suggest that the 1956 Defense Appropriations Bill expressed congressional policy on reprogramming that still remains in force.^[5] The report attached to the bill stated that there would always be some unforeseen changes in operating conditions and circumstances in the interval between building and presenting the budget and executing it. These might include changes in operating conditions, revisions in price estimates, wage rate adjustments, and so on. Consequently, rigid adherence to the original budget might prevent the effective accomplishment of the planned program. The report cautions that it was not the intention of Congress to give the military departments unrestrained freedom to reprogram or shift funds from one category or purpose to another without prior notification and consent from Congress. Since 1959, Congress has emphasized prior approval before implementation of certain transfers, certain threshold requirements, and periodic reporting on issues of special interest to Congress. In 1974, Congress added a restriction that prohibited the Department of Defense from transferring funds to restore budget items specifically denied by Congress. However, supplementals are aimed at a different class of actions, namely emergencies arising within the fiscal year where expenses are large enough to warrant the appropriation of additional funds and where the legislature may need to authorize that the mission is something

^aFor the definitive analysis of Supplemental Appropriations, see the series of reports issued by the Congressional Budget Office, Budget Analysis Division, including Supplemental Operations in the 1970s (July 1981); Supplemental Operations in the 1980s (Feb. 1991); and Supplemental Appropriations in the 1990s (March 2001). The most recent version may be found on the Congressional Budget Office web page: www.cbo.gov

that should legitimately be carried out by agents of the U.S. government. Thus, supplemental needs may be too large to accomplish with agency fund transfers or reprogramming. Moreover, they may cross agency and jurisdiction lines and respond to new needs outside the mission of any one agency or, indeed, outside the United States. If an agency were to reprogram or transfer funds to meet this new need, it would not have the money to fulfill the missions it had been given in its appropriation bill. In that sense, it would be breaking a promise to Congress; thus, in addition to provision of funds, a supplemental bill also legally sets the new programs or tasks as a legitimate mission of the government and the agencies affected. Thus, a supplemental is about both dollars and policy.

Nowhere is this more apparent than in the defense policy area. The failure to fund certain military assistance needs in a supplemental in April 1975 came at a time when Saigon was falling and was a signal to the world that the United States would pursue no further military aims in Vietnam. Conversely, the three supplementals for the Gulf War in 1991 and for peacekeeping operations in Bosnia beginning in 1996 were clear signals of U.S. interests and intentions. Within the last decade, supplemental appropriations have funded wars, supported peacekeeping operations, responded to terrorist attacks, and helped various disaster victims both in the United States and abroad. Supplemental appropriations allow for both concrete help and a symbolic expression of policy. A supplemental passed to aid flood victims in Central America or for peacekeeping reminds everyone of the purposes to which the U.S. government is willing to commit both its treasure and its people.

STATE PATTERNS

States also have to meet emergencies. In testimony before Congress, Theresa Gullo of the Congressional Budget Office reported on state practices.^b In general, states plan to meet two types of fiscal emergencies: economic downturns and natural disasters. Gullo reported that most states have created budget stabilization funds (usually referred to as Rainy Day funds) to meet a weak economy. These may be used for a variety of purposes but only in times when revenues fall short of expectations. Gullo observed that when faced with an unexpected decline in revenues or an increase in spending needs, 44 states used their budget stabilization

fund to tide them over. Normally a vote of the legislature precedes spending the funds. Rainy Day funds are rarely used to pay for natural disasters or other types of emergency needs. Randall Holcombe notes that Rainy Day funds have been successful in mitigating some of the aspects of fiscal stress, but a Rainy Day fund that would completely eliminate any negative impact from a recession like that in 1990–1991 would have had to have been about one-third the size of a state's annual budget; Holcombe observes "No state has ever had a rainy day fund anywhere near this large and many cap their funds at much lower levels."^[6] Holcombe concludes that Rainy Day funds are not large enough to avert depressions but do provide states with a cushion going into a recession so that they have some additional time to adjust their spending patterns. For natural and man-made disasters and other non-fiscal emergency needs, states establish a variety of contingency and emergency accounts. Gullo noted that Maryland had both a Contingent Fund and a Catastrophic Event Fund. The Contingent Fund was established by the state constitution and was used when an agency's annual appropriation was insufficient to cover salaries and operating expenses. The Catastrophic Event Fund, established by statute, held funds to meet natural disasters and other emergency situations.^b

Gullo found that state legislatures typically appropriated small sums ranging from a few hundred thousand to several hundred million dollars to emergency-response accounts. Generally when a disaster occurs, the governor declares a state of emergency and requests a supplemental appropriation from the legislature. This appropriation is drawn from the disaster fund. This procedure allows the state legislature to maintain control over the total amount spent for disasters, while the governor is given the discretion to decide how and when that money is spent. Occasionally, a governor has the emergency authority to bypass the legislature entirely and borrow from almost any other state budget account. Gullo found that Idaho, Florida, and Virginia empowered their governors to transfer all unobligated resources of all state agencies to the emergency account to address a disaster. Usually, governors must go to the legislature for supplemental sums of money. Gullo noted that in 36 states, the governor authorized the use of emergency funds, often after declaring a state of emergency (or after a federal emergency was declared). She observed that South Carolina and Ohio had established emergency councils that made decisions about how to allocate appropriated funds for emergencies and that Oregon was almost alone in placing authority for spending for emergencies solely with the legislature and in an emergency board of legislators when the legislature was not in session.^b Other states used some combination of

^bTestimony of Theresa A. Gullo, Chief, State and Local Government Cost Estimates Unit on "How States Budget and Plan for Emergencies" before the House Budget Committee on June 23, 1998. Discussion of state practices is largely drawn from this testimony.

executive and legislative authority. Gullo concluded that states count on the federal government to step in during a large scale emergency and their goal is to have sufficient funding in their emergency account to provide any required state match for federal disaster assistance. States also use their emergency money to get disaster relief and recovery planning started as soon as an emergency occurs, so that disaster relief and recovery efforts can be carried out expeditiously when federal money becomes available.

CONCLUSION

Supplementals in the 1970s and 1980s were biased toward the mandatory expenditures but have been dominated by defense and disaster expenses in the 1990s. Some attempt has been made to offer up offsetting rescissions in the 1990s, but a full offsetting of supplementals has been rare. Nonetheless, the use of rescissions means that appropriations for an agency may actually be decreased in a supplemental bill if it has a program that is offered up in a rescission to support supplemental funding in another agency or program. Moreover, the necessity to find offsetting funds or designate funding as “emergency” has led to an increase in the use of that designation; as a result, not only do emergency designations appear in supplemental bills, but the emergency designation has slipped into regular appropriations bills, so that now some regular appropriation bills have emergency funding within them as do some supplementals. Moreover, sometimes the emergency funding lasts for more than one year, thus the supplemental with emergency funding could affect not only this year but the next year.

This tends to confuse an already complex budget process and critics suggest the second year of emergency funding should be included in the regular appropriations bills. Others observe that when spending caps are set unrealistically low, spending is deemed “emergency spending” simply to get around the caps and that more realistic caps would lead to less emergency spending. Supplementals have been an issue since 1790, but they are not a big issue. However, their symbolic importance is dramatized by the size and speed of the passage of the supplemental in response to the attack on the World Trade Center of September 11, 2001. Whatever their limitations and complexities, supplemental appropriations must be seen as part of the full panoply of budgetary tools any modern government must deploy.

REFERENCES

1. Tyszkiewicz, M.; Daggett, S. *A Defense Budget Primer*; Congressional Research Service, Library of Congress: Washington, DC, 1998; 42–43.
2. Godek, P. *Emergency Supplemental Appropriations: A Department of Defense Perspective*; Naval Postgraduate School: Monterey, CA, 2000.
3. CBO. *Supplemental Appropriations in the 1990s*. Congressional Budget Office: Washington, DC, 2001; 1, 2.
4. Wildavsky, A. *The Politics of the Budgetary Process*; Little, Brown and Co.: Boston, 1964; 31–32.
5. Jones, L.R.; Bixler, G. *Mission Financing to Realign National Defense*; JAI Press: Greenwich, CT, 1992; 58–60.
6. Holcombe, R. Rainy day fund. In *International Encyclopedia of Public Policy and Administration*; Shafritz, J., Ed.; Westview Press: Boulder, CO, 1998; 1903 pp.

Sustainable Development

Richard C. Feiock

*Askew School of Public Administration and Policy, Florida State University,
Tallahassee, Florida, U.S.A.*

Stephen C. Adams

*Office of Strategic Projects and Planning, Florida Department of Environmental Protection,
Tallahassee, Florida, U.S.A.*

INTRODUCTION

Sustainable development is the term used to refer to the idea of simultaneously improving societal economic well-being, protecting the resilience of the natural environment, and addressing inequities in the distribution of power and wealth. A consensus on a precise definition of sustainable development has eluded the international policy community despite a decade and a half of intense debate. But the working definition with the greatest currency is the one advanced by the 1987 final report of the World Commission on Environment and Development^[1] commonly referred to as the Brundtland Commission after its chair, Norwegian Prime Minister Gro Harlem Brundtland:

Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs. It contains within it two key concepts: the concept of 'needs,' in particular the essential needs of the world's poor, to which the overriding priority should be given; and the idea of limitations imposed by the state of technology and social organization on the environment's ability to meet present and future needs.

As Rao^[2] notes, most references to the Brundtland definition refer only to the first sentence and so are "less than fair to the spirit of the original contributors." He considers this a critical omission in that the second sentence qualifies the first in an important way, as the first sentence in isolation tends to emphasize intergenerational over intragenerational equity. Says Rao, "It is rather naïve to believe that [sustainable development] can be viewed as a meaningful approach when it is formulated totally independent of income distribution policies."

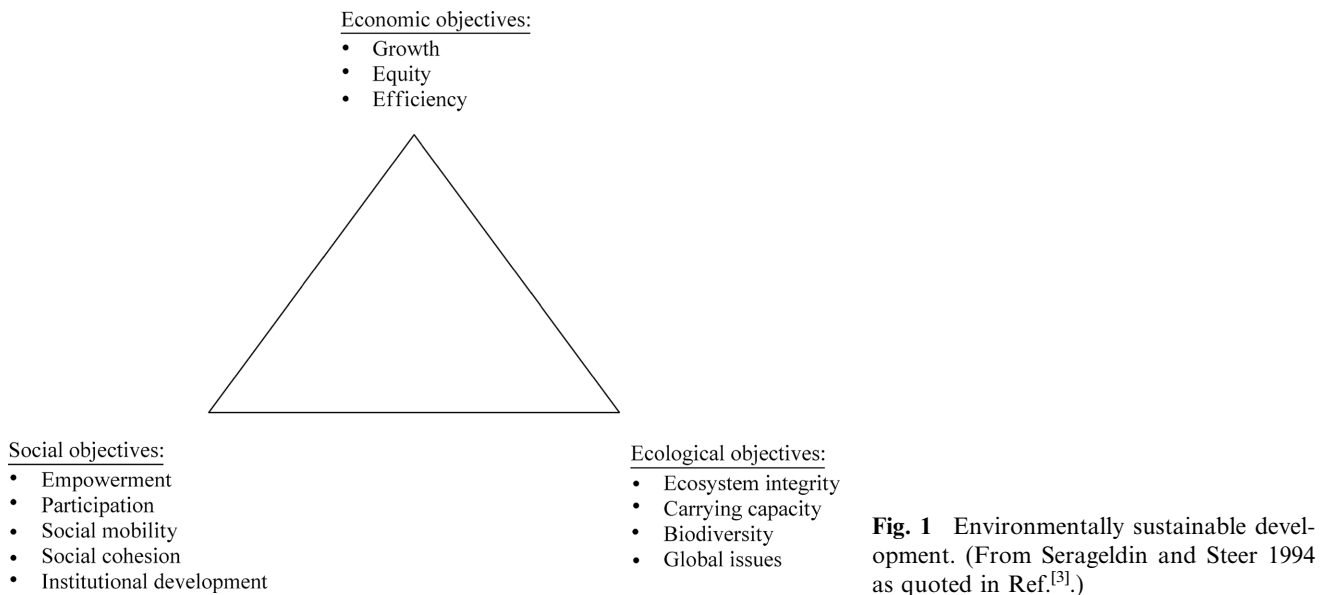
This analysis of the Brundtland Commission's definition represents the concept of sustainable development in its broadest sense and, as such, is an appropriate base from which to explore the debated meaning of the term. The scholarship of Desmond McNeill provides the foundation for first understanding

sustainable development as a phenomenon of conflict and then in sketching a "structural analysis" of the debate surrounding its meaning. We will consider some dissenting views, and we will conclude with an assessment of the sustainability of the concept of sustainable development itself as "an idea in good currency" to use Donald Schon's phrase.

A STRUCTURAL ANALYSIS OF THE DEBATE

McNeill's^[3] view of sustainable development focuses on the "potential conflict" with which the term is pregnant. Of the theaters of conflict described by McNeill, we focus on three: between present and future generations, between humanity and the rest of nature, and between the rich and poor of the current generation. At issue are three sets of objectives, best illustrated by the Environmentally Sustainable Development triangle (Fig. 1).

McNeill argues that the Brundtland process was a "remarkably successful agenda-setting exercise" that effectively solved "a perceived dilemma between growth and the environment." In effect, the Brundtland Commission partially succeeded in superceding two long-standing conflicts related to development and the environment by redrawing the lines of "cleavage" between the camps. This strategic alteration of the scope of conflict is consistent with E. E. Schattschneider's description of the mobilization of bias.^[4] Interpretation of the process helps us understand the lingering debate about the meaning of sustainable development: The lines are no longer as clearly drawn among the warring camps as they were prior to the introduction of the sustainable development concept. As McNeill notes, the Brundtland Commission's relative success in fusing the dynamics of the two debates into the "sustainable development debate" was hamstrung by the use of less specific language necessary to achieve a consensus among the contributing authors. McNeill offers a highly useful "structural analysis" of the sustainable development debate and the origins of the specific issues from the two debates it replaced (Fig. 2).



McNeill may be overly modest in fretting about whether the “lines of fission” of the two debates “map” cleanly to the sustainable development debate. As he insightfully notes, the coherence of the debate of the third column as a viable alternative to the two separate debates is central to the dispute over the meaning of sustainable development. The future viability of the sustainable development debate depends on the veracity of the claims that the three competing sets of objectives can be simultaneously achieved.

DISSENTING VIEWS

Ecology

Scholars and policy partisans each have brought a wealth of critiques to the debate surrounding sustainable development. Many scholars have focused on the contributions of the various disciplines. Rao, for instance, emphasizes the specific contributions of the science of ecology to the Brundtland definition.

Specifically, Rao argues that the reference to environmental limits within the context of prevailing technology or institutional arrangements is a factual characteristic of the non-linear nature of ecosystems rather than a neo-Malthusian condition of resource constraints. On this point, Drummond and Marsden^[5] writing from the perspective of academic urban planners, are quite critical. They argue that the debate must “move beyond the somewhat naïve conceptions of a ‘resource crises’ engendered by such publications as ‘Limits to Growth’^[6]...The requirement for such limits may appear to be intuitively obvious and a practical necessity, but the utility of any approach to sustainable development...is highly questionable.”

Economics

The discipline of economics can also be credited with vital contributions to the sustainable development debate. An environmental studies generalist, Paehlke^[7] provides great value to the enterprise of defining sustainable development by tackling the imprecision of



Fig. 2 Structural analysis of the sustainable development debate. (From Ref.^[3].)

the language involved in the debate. He differentiates sustainable development from sustainability, defines several applications of sustainability as an idea, and then argues that facilitating greater precision in the language used should be high on the agenda for researchers and policy partisans alike.

Paehlke, in differentiating sustainability from sustainable development, does so in full recognition that “sustainability as a concept travels with somewhat less political baggage” as many “environmentalists see sustainable development as an oxymoron, little more than political cover for otherwise unacceptable corporate environmental practices.” This insight defines quite precisely much of the tenor of the dissent voiced by partisans in the sustainable development debate. To some extent, this relates to the relative success of the sustainable development debate to supplant the development and the environmental debates. Dobson^[8] gives voice to this skepticism by concluding that “for those whom the defense of the non-human natural world is a prime consideration need to pick and choose their political allies with care.” But environmental activists aren’t alone in their skepticism. Poverty advocates, antiglobalism activists, and social justice organizations have demonstrated ample skepticism of sustainable development. As McNeill notes, “Global warming is certainly not a priority issue for many of the poorest people of the world, but it does in many ways exemplify the challenge—practical, ethical, and political—posed by the concept of sustainable development.”

IMPLICATIONS FOR PUBLIC ADMINISTRATION

The challenge for public administration scholars is to subject the claims of sustainable development to rigorous testing. Some commentators have pronounced certain approaches to environmental management as capable of achieving sustainable development, but they do not back the claims with credible evidence. For example a much-cited report by the Institute of Southern Studies^[9] draws this conclusion based on a positive correlation between state environmental regulation efforts and personal income. Most studies do not provide systematic explanations for *how* trade-offs between environmental quality and economic growth are overcome, nor do they subject their hypotheses to rigorous empirical tests. Recently Feiock and Stream^[10] applied a transaction cost approach to critically examine claims of sustainable development. The challenge for public administration practitioners is to promote sustainable

development but not apply sustainability as a label or symbol to legitimize certain policy agendas.

The Sustainability of SD as an Idea

Ultimately the “sustainability” of sustainable development as an idea in good standing will depend upon its successful implementation. Many observers have pointed to a political agreement for the global control of greenhouse gas emissions as the first test case for implementing a sustainable development approach. Given that the Kyoto Protocol called for graver restrictions on the developing countries has been seen as evidence of a growing world consensus that the developed countries must assume a greater share of the costs of global environmental protection initiatives. Such projects will have to demonstrate that economic development can create more just human societies without negatively impacting the environment. Perhaps implementing sustainable development will prove less contentious than defining it.

REFERENCES

1. World Commission on Environment and Development *Our Common Future*; Oxford University Press: Oxford, England, 1987.
2. Rao, P. *Sustainable Development: Economics and Policy*; Blackwell Publishers: Malden, MA, 2000.
3. McNeill, D. The concept of sustainable development. In *Global Sustainable Development in the 21st Century*; Lee, K., Holland, A., McNeill, D., Eds.; Edinburgh University Press: Edinburgh, 2000.
4. Schattschneider, E.E. *The Semisovereign People*; Harcourt Brace Jovanovich: Fort Worth, TX, 1960.
5. Drummond, I.; Marsden, T. *The Condition of Sustainability*; Routledge: London, 1999.
6. Meadows, D.H.; Meadows, D.L.; Randers, J.; Behrens, W. *The Limits to Growth*; Universe Books: New York, 1972.
7. Paehlke, R. Towards defining, measuring and achieving sustainability: tools and strategies for environmental valuation. In *Sustainability and the Social Sciences*; Becker, E., Jahn, T., Eds.; Zed Books: London, 1999.
8. Dobson, A. Sustainable development and defense of the natural world. In *Global Sustainable Development in the 21st Century*; Lee, K., Holland, A., McNeill, D., Eds.; Edinburgh University Press: Edinburgh, 2000.
9. Hall, B. *Gold and Green*; Institute for Southern Studies: Durham, NC, 1994.
10. Feiock, R.; Stream, C. Environmental protection vs. economic development: a false trade-off? *Public Adm. Rev.* **2001**, *61*, 313–321.

Tax Expenditures

Iryna Illiash

*Graduate Department of Public Administration, Rutgers University,
Newark, New Jersey, U.S.A.*

INTRODUCTION

Tax expenditures are reductions in tax liabilities that result from preferential provisions in the federal tax code, such as special exclusions and exemptions from taxation, deductions, credits, deferrals, or preferential tax rates. Tax expenditures include revenue losses for the income tax, the estate and gift taxes, as well as payroll and excise taxes.^[1] Table 1 provides a list of the 20 largest tax expenditures in federal income taxes.

As special provisions, tax expenditures are treated as deviations from the normal tax structure. Theoretically, a provision is structural if it exists to accurately estimate income. In practice, however, there has been continual disagreement among tax experts about what to consider a “normal” structural provision and what not. So far only personal exemptions, standard deductions, and guaranteed individual rates are universally recognized as part of the normal structure.^[2]

The underlying assumption behind the concept of tax expenditures is that the tax system is not just the means of raising revenue, it is an important tool used to accomplish a variety of objectives of public policy. The term “tax expenditure” is meant to express the idea of government subsidy through the tax system, parallel to direct government spending, with a view to influence particular kinds of economic and social behavior of both individuals and corporations and to encourage certain investment (by increasing the after-tax returns) and consumption (by reducing the after-tax cost of particular goods and services) activities in order to provide economic relief to selected groups of taxpayers and consumers. The criterion for such subsidizing is a choice of activities that further economic, cultural, and social objectives of the entire society, not just confer benefits on those individuals or corporations who are directly engaged in the activity. Basic research, experimentation investment, health care consumption, or charitable giving are the examples of such activities. The rationale behind such subsidizing is that individual taxpayers or businesses alike, if they do not capture all benefits themselves, would undertake too little of the activity in the absence of a subsidy.^[1]

PURPOSE

Whatever the purpose, all tax expenditures invariably favor certain selected groups and activities. Some tax expenditures provide for adjustment for differences in individual’s ability to pay taxes. For example, if two taxpayers have the same income, but one has a serious illness and costly medical bills, the other taxpayer is considered better able to pay taxes on his income. Some tax expenditures, such as the Low-Income Housing Tax Credit, are designed to mimic the features of a direct spending program, others, such as deductions for charitable contributions or home mortgage interest, are attributed to the American tradition and the way of life.^[3] There are also tax expenditures believed to compensate for the deficiencies of the tax system. For example, the special tax treatment of capital gains is justified on the grounds that they are not indexed for inflation.^[1]

One of the most thorough classifications of tax expenditures belongs to Witte. He believes that tax expenditures should be classified in such a way as to help shape the theoretical issues central to tax debates. Therefore, his classification is based on the distinctions between need, equity and economic efficiency of tax expenditures.^[4] It includes the following six categories summarized in Table 2: need-based provisions, tax equity provisions, special group benefits, general economic incentives, specific economic incentives, and miscellaneous provisions.

Tax expenditures are considered inefficient when they subsidize an activity that could be undertaken in the absence of a subsidy, or when they cannot be traced down to those entities that are meant to receive the benefits.

HISTORIC OVERVIEW

Although tax expenditures have entered the political debate quite recently, largely due to the attention to the problem of the federal budget deficit, they have existed almost as long as taxes. The home mortgage interest deduction, introduced into the tax code in 1913, became the first tax expenditure. It was followed

Table 1 The twenty largest tax expenditures in federal income taxes fiscal years 1998 and 1998–2002 (billions of dollars)

Tax expenditure item	1998	1998–2002
Exclusion of employer contributions for medical insurance premiums and medical care	75.8	435.7
Net exclusion of pension contributions and earnings	56.2	285.4
Deduction of mortgage interest on owner-occupied homes	52.1	284.8
Exclusion of capital gains at death	32.0	173.0
Deduction of state and local income and personal property taxes	40.0	169.0
Accelerated depreciation of equipment	29.3	176.0
Deduction of charitable contributions	22.3	123.4
Exclusion of Social Security benefits	18.5	100.3
Deduction of property taxes on owner-occupied homes	17.4	95.3
Exclusion of interest on public purpose state and local debt	15.7	77.2
Deferral of capital gains on home sales	15.3	81.2
Exclusion of interest on life insurance savings	11.9	68.0
Exclusion of interest on state and local debt for various private purposes	8.9	42.4
Net exclusion of Individual Retirement Account contributions and earnings	8.6	45.5
28% maximum rate on capital gains (other than agriculture, timber, iron ore, and coal)	8.5	45.0
Earned income credit ^a	5.8	31.7
Exclusion of worker's compensation benefits	5.3	29.6
Exclusion of capital gains on home sales for persons age 55 and over	5.1	27.2
Graduated corporation income tax rate	4.9	27.2
Accelerated depreciation of buildings other than rental housing.	4.7	13.2
Deduction of medical and long-term care expenses	4.3	26.5

^aThe figures in the table indicate the effect of the earned income credit on receipts. The effect on outlays (in billions of dollars) is 22.0 in 1998 and 118.8 for 1998–2002.

Note: These estimates do not reflect the effects of the Taxpayer Relief Act of 1997.

Sources: From Refs.^[3,17].

by the adoption of tax expenditures for employer pensions (1914–1926), employer-provided health benefits, and for child-care expenses (1954).^[5] In the 1960s, Stanley S. Surrey, assistant secretary of the Treasury for tax policy in the Kennedy and Johnson Administrations, developed the concept of “tax expenditures” and a related concept of the tax expenditure “budget” to raise awareness about the existence of tax subsidies and to facilitate their comparison with direct government outlays.^[1] The first Tax Expenditure Budget was released in 1969. Since then, it has been published annually as a special appendix to the federal budget and organized according to functional policy areas, such as energy, natural resources, income security, etc. The first “official” definition of the term “tax expenditures” appeared in the Congressional Budget and Impoundment Control Act of 1974.

A significant attempt to curb the growth of tax expenditures was made in 1986. The major thrust of the 1986 Tax Reform Act was to offset tax rate reductions by eliminating hundreds of billions of dollars of tax expenditures. The act repealed 14 tax expenditures and scaled back 16 others. In 1990, the Budget Enforcement Act reduced the historical advantages of

tax expenditures even more through placing absolute caps on discretionary spending.

The Government Performance and Results Act of 1993 raised the issue of assessing the relationship between spending programs and related tax expenditures, thus focusing attention of the political community on a possible trade-off between these policy tools. Finally, the 1997 Taxpayer Relief Act signified the rebirth of tax expenditures by creating new, and expanding already existing, tax subsidies.^[3]

PRESENT-DAY ROLE

The proliferation of tax expenditures is believed to be closely linked to the current transformation of the public sector driven by government downsizing, privatization, and contracting out. These processes have turned the government from the provider of public services to the purchaser of such services. They have also changed the structure of government finance. Tax expenditures result from the fact that nowadays the spending mostly occurs not within but outside the government.^[6]

Table 2 Witte's classification of tax expenditures

Category	Provisions	Description	Who benefits
Need-based provisions	Exclusions of Social Security (OASI, survivor's benefit and disability payment), workmen's compensation, unemployment benefits, veterans' disability payments, public assistance, capital gains on home sales of the elderly, railroad retirement, military disability, disability pay, and of coal miners' disability benefit; deduction of medical expenses, casualty loss, and of adoption expenses; additional exemption for the elderly and for the blind; earned income and retirement tax credit; exemption of students over 18; and excess of percentage over minimum standard deduction	Provisions in this category are designed to take into account presumed economic needs, such as conditions imposing restraints on income-earning potential, extraordinary expenses, situations of economic hardships, and conditions affecting the elderly.	Mostly low-income but also middle-income groups
Tax equity provisions	Deductions of state and local taxes, real estate taxes, state gasoline tax, and deductions for two-earner couples; exclusions of state and local bond interest and of employer-furnished meals and lodging; child-care credit; and deferral of taxes on foreign corporate income	These provisions are designed to either eliminate taxes on taxes or take into account expenses incurred while earning income.	Middle- and higher-income groups
Special group benefits	Exclusions of military benefits, veterans' pensions, reinvested utility dividends, certain agricultural cost sharing; capital gains for timber income and for certain farm income, as well as capital gains on coal and iron ore royalties; expensing of farm capital outlays; excess bad debt for financial institutions; deferral of tax on shipping companies, exemption of credit union income.	Provisions of this category must meet the following criteria: 1) The provision affects a reasonably permanent and identifiable demographic or occupational group (This criterion applies a restrictive definition to the concept of a group delimited by three tests: whether the "group" has formal organizations or associations, whether its members have common identifications, and whether the membership in the group reflects a relatively permanent condition); 2) The economic conditions addressed by the provision cannot be assumed to be exclusive to that group (This criterion gives concrete meaning to the term "special"); 3) There is no indication that the benefit is being granted primarily on other grounds.	Low-income as well as some higher-income groups

(Continued)

Table 2 Witte's classification of tax expenditures (*Continued*)

Category	Provisions	Description	Who benefits
General economic incentives	Investment credit, capital gains, asset depreciation range, research and development benefits, dividend exclusion, excess first-year depreciation, all savers certificates, etc.	The provisions in this category are meant to induce investment and to enhance economic growth.	Almost exclusively higher-income groups
Specific economic incentives	Exclusion of interest on life insurance, on industrial development bonds, housing bonds, pollution control bonds, student loan bonds and on savings bonds; deduction for mortgage interest, deduction of charitable contributions, such as to health organizations or for education; exclusions of self-employed contributions to pensions, exclusions of scholarship and fellowship income; deferral of capital gains on home; credits for residential energy expenses, for new energy technology and alternative fuels, for political contributions; tax credit for corporations in U.S. possessions, investment credit for employee stock ownership plans, general and targeted jobs credit, investment credit for housing rehabilitation; and benefits for preserving historic structures	The purpose of these provisions is to encourage specific forms of desirable behavior other than general economic investment. Distinguishing between category V and category III involves the nature of organized interests and the consideration whether the provision is designed to aid a group or to induce a specific activity.	Middle- and upper-middle-income groups
Miscellaneous provisions	Deduction of interest on consumer credit, safe harbors leasing, exclusions and deductions for income earned abroad, and maximum tax on personal services "earned" income	This category contains those provisions that cannot be easily placed into any of the other categories.	Middle- and higher-income groups

Source: From Ref.^[4].

CHARACTERISTIC DYNAMICS AND PROBLEMS

The concept of tax expenditures is controversial as it involves some perceptually difficult issues. One of the conceptual difficulties, mentioned earlier, is caused by the attribute "special" used in the definition of tax expenditures. It leaves room for arbitrary judgment as to what should be considered the tax expenditure.

Another difficulty arises from the semantic ambiguity of the term "tax expenditures" itself leading to different interpretations of its meaning. Thus, according to the conservative view, the term implies that all income inherently belongs to the government, and only government's generosity allows people to keep income

they produce.^[4] On the other hand, the progressive view indicates the recognition that tax expenditures are, effectively, part of the overall budget process^[7] and as such, are analogous to programs on the expenditure side of the budget.

However, contrary to direct spending programs, tax expenditures have low visibility, as they do not require the allocation of actual funds. They are, in fact, the funds not collected by the government. For this reason, tax expenditures do not receive the same degree of congressional scrutiny as direct spending.^[8] Tax expenditures are immune from the annual appropriations process and not subject to any mechanisms of audit, oversight, or systematic review.^[5] At the same time,

they reduce the government's resources in the same way as regular spending does. Tax expenditures often match and sometimes exceed direct expenditures in size. For example, in 1998 the federal government financed more programs through the tax code than through the regular appropriations process.^[2] The low profile of tax expenditures allows policy makers to achieve a variety of goals by discretely allocating tangible benefits outside of the budget process.^[9]

The main cause of "the structural protection" of low visibility, according to Meyers, is the inherent difficulty of measuring tax expenditures.^[8] Unlike most direct expenditure programs, tax expenditures are open-ended as to the amounts involved.^[10] Also, as Perry indicates, the estimate of revenue forgone due to tax expenditure provisions cannot be directly translated into tax revenue gained as a result of the elimination of those provisions. First of all, tax expenditures are the incentives targeting taxpayers' behavior, therefore this behavior will change in the absence of the benefit. For example, employees might become reluctant to accept employer-sponsored health insurance if its cost is included in taxable income. Secondly, a tax expenditure and a direct spending program with the same objective may allocate benefits to different target populations. For instance, old age security pensions provide assistance to senior citizens with little or no other income, while the pension tax credit benefits those with taxable income. Finally, a tax benefit can be more valuable to an individual in terms of real money than the same amount of direct (and taxable) payment.^[11]

Because tax expenditures in general are not subject to congressional review, their growth is uncontrollable. The tax expenditures system has proliferated at a much more rapid rate than the gross national product, federal outlays, or income tax receipts. From the mid-1960s to the early 1980s, the number of tax expenditures increased almost fivefold, which allowed Meyers to conclude that they had become the preferred tactics of the period.^[8] In 1993, there were 124 different income tax exemptions, exclusions, deductions, credits, deferrals, etc. As concerns the federal excise tax system, 9 different types of excise taxes produced 244 tax expenditure provisions.^[12] According to some estimates, tax expenditures now total somewhere between \$700 billion and \$800 billion a year.^[13]

Moreover, since tax expenditures function like budgetary entitlements, they can grow even without advocacy.^[5] For this reason, they are hard to eliminate. Once enacted, tax expenditures virtually never get removed from the tax code. Out of more than 100 tax expenditures enacted in the period between 1913 and the passage of the 1986 Tax Reform Act, only 13 had been permanently eliminated. Some say that the design of tax expenditure policies, with their

built-in growth, holds the key to the "indestructibility" of these provisions. It essentially commits policy makers to enforce decisions made long ago, while denying them the ability to reallocate resources toward the newly emerged needs.^[1]

Also, under tax expenditures' tendency to survive and expand lie strong economic and political motivations. As tax expenditures reduce the amount of taxes, they also force overall statutory tax rates to rise in order to obtain the same amount of revenue. It sets out a chain reaction: Higher rates prompt interest groups to increase their efforts in obtaining new, or increasing existing, tax benefits. Because stakes are high, these groups are strongly motivated to organize support of a given provision in such a way as to make it difficult to later eliminate or change it. In contrast, taxpayers in general, who would benefit from smaller tax rates, are too diffuse and disorganized to perceive any advantages from the elimination of the tax expenditure provision. Thus, the lobby for the tax expenditure is much stronger than the opposition to it.^[1]

Tax expenditures diminish the tax base, have a negative effect on revenue capacity, and decrease the resources available for funding other programs or reducing federal deficit. Interestingly, the largest revenue losses have been caused by those tax expenditures that either accompanied the enactment of the income tax or were introduced soon after. Almost 85% of 1993 tax expenditure revenue losses were attributed to the tax expenditures listed before 1950, and among them, almost 50% were caused by the tax expenditures enacted before 1920. In contrast, only 6% of the 1993 revenue losses stemmed from the tax expenditures originated since 1970.^[1]

Tax expenditures increase administrative complexity of the tax code. Witte explains, "Fine-tuning to achieve refined economic or political effects often adds amendment on amendment to provisions such as capital gain, the expensing and depletion of mineral resources, the minimum tax, and a range of real estate and agricultural provisions designed to contain tax shelters. The result is often that, within the scope of a single provision, small policy empires are created that may be thoroughly understood by only small groups of government and private-sector experts."^[4] Major side effects of this complexity are decreased compliance with the tax code and the inefficient use of private as well as government resources.

Contrary to their counterpart, direct expenditures, tax expenditures are not designed to affect income redistribution. Their role is to subsidize population in order to create various economic incentives and to encourage certain investment and consumption activities. However, while providing some benefits for nearly every income class or occupational grouping (see [Table 2](#)), tax expenditures disproportionately benefit

the well-off, as well as private corporations. For instance, the top 10% of earners receive 24% of employer-sponsored health insurance benefits. When applied to families with different annual income, this means that those earning \$100,000–\$200,000 receive from the government \$1,700 for health insurance a year, while those earning \$10,000–\$20,000 get only \$150.^[13] This effect is known as “upside-down” subsidy.^[1]

Witte, in this regard, argues that we cannot apply to tax expenditures the standard of an equal division among people, usually used in discussions of equality of income and wealth, because the assumption that tax expenditures should be equally distributed among the filers of tax returns is unrealistic and misrepresents the political intent of these provisions. Tax expenditures are distributed in direct proportion to the amount of taxes paid. They tend to provide relatively greater benefits to higher income taxpayers because the latter pay higher taxes.^[4]

At the same time, Howard, in his innovative research, provides a new outlook on tax expenditures as indirect tools of social policy. According to Howard, tax expenditures have had a significant impact on the development of the American welfare state, of which they are an integral although less visible part—“a hidden welfare state,” as he^[5] puts it.

As tools of social policy, tax expenditures possess certain qualities that give them a considerable tactical advantage over direct spending. These qualities are low visibility of tax expenditure costs, less stringent budgetary reviews, ease of enactment (can be appended to a bill), and their defensibility on multiple grounds.

CONCLUSION

The concept of tax expenditures is not universally accepted. But, like it or not, with the demise of “the era of big government” their role as policy tools will only increase. They’ve already become the “vehicle of choice” in distributing government funds.^[2] And in many respects, tax expenditures are the efficient, effective, and equitable way to provide government assistance to citizens. At the same time, the flaws inherent in them call for increased scrutiny over them. To address this need, several new frameworks have emerged, such as to bring tax expenditures within the scope of the budgetary process,^[14] to incorporate tax expenditures into the performance review process,^[15] and to evaluate them under the same constitutional criteria as direct spending programs.^[16] The ultimate goal of these and many other efforts is to increase the usefulness and visibility of tax expenditures and force the comparisons between conventional and non-conventional spending.

REFERENCES

1. United States General Accounting Office *Tax Policy. Tax Expenditures Deserve More Scrutiny*; GAO: Washington, DC, 1994; pp. 2, 3, 14–16, 27, 28, 31, 32.
2. Sugin, L. Tax expenditure analysis and constitutional decisions. *Hastings Law J.* **1998–1999**, 50, 407–474.
3. Davie, B.F. Tax expenditures: The basics. In *Evaluating Tax Expenditures: Tools and Techniques for Assessing Outcomes*; Datta, L.E., Grasso, P.G., Eds.; New Directions for Evaluation; Jossey-Bass Publishers: San Francisco, CA, 1998; Vol. 79, 9–23.
4. Witte, J.F. *The Politics and Development of the Federal Income Tax*; The University of Wisconsin Press: Madison, WI, 1985; pp. 269, 273–283, 291, 299.
5. Howard, C. *The Hidden Welfare State: Tax Expenditures and Social Policy in the United States*; Princeton University Press: Princeton, NJ, 1997; pp. 3, 6, 44, 45, 183.
6. Schick, A. Off-budget expenditure: An economic and political framework. In *Paper Prepared for the Organization for Economic Cooperation and Development, Paris*; 1981. Quoted in *Budgeting: A Comparative Theory of Budgetary Processes*, 3rd Ed., Wildavsky, A.; Transaction Publishers: New Brunswick, NJ, **1997**, 349–350.
7. Snow, D.R. Do legislative procedures affect tax expenditures? *J. Pub. Budg. Account. Financ. Manag.* **1999**, 11 (3), 357–385.
8. Meyers, R.T. *Strategic Budgeting*; The University of Michigan Press: Ann Arbor, MI, 1994; 79–81.
9. LeLoup, L.T. From microbudgeting to macrobudgeting: Evolution in theory and practice. In *New Directions in Budget Theory*; Rubin, I.S., Ed.; State University of New York Press: Albany, NY, 1997; 19–42.
10. Datta, L.E., Grasso, P.G., Eds.; *Evaluating Tax Expenditures: Tools and Techniques for Assessing Outcomes*; New Directions for Evaluation; Jossey-Bass Publishers: San Francisco, CA, 1998; Vol. 79, 1–7.
11. Perry, D.B. Comparison of direct spending and tax expenditures. *Can. Tax J.* **1995**, 43 (3), 793–803.
12. Davie, B.F. Tax expenditures in the federal excise tax system. *Natl. Tax J.* **1994**, 47 (1), 39–62.
13. Madrick, J. Economic scene: The richest get the biggest tax benefits. So what else is new?. *N.Y. Times* June 7, **2001**, C2.
14. Schick, A. Controlling nonconventional expenditure: Tax expenditures and loans. *Public Budg. Finance* **1986**, 6 (1), 3–20.
15. Heen, M.L. Reinventing tax expenditure reform: Improving program oversight under the government performance and results act. *Wake For. Law Rev.* **2000**, 35 (4), 751–826.
16. Adler, D.D. The internal revenue code, the constitution, and the courts: The use of tax expenditure analysis in judicial decision making. *Wake For. Law Rev.* **1993**, 28 (4), 855–917.
17. Office of Management and Budget *Analytical Perspectives, Budget of the United States Government, Fiscal Year 1998*; Government Printing Office: Washington, DC, 1997; 79 pp.

Tax Incidence

Daphne Greenwood

*Department of Economics, University of Colorado at Colorado Springs,
Colorado Springs, Colorado, U.S.A.*

INTRODUCTION

Who really pays the taxes? Many people worry about having to pay more than their “fair share” of taxes. It is important to understand where the true economic burdens of a tax lie, for purposes of ensuring whatever concept of fairness has been agreed upon and for understanding the effect of the tax on efficiency (its influence on economic decisions). Of course, because equity is often in the eye of the beholder, there are many different views of fairness.

The economic effect of a tax is often not as clear as it initially appears to the observer. This is because the actual burden of a tax can be shifted from the party who is legally required to pay the tax. The tax burden may show up in higher prices to buyers. If it is difficult to pass the tax on to buyers, it may be possible for a company to shift the burden to workers. While it is unusual for workers to see an actual decrease in nominal wages or salaries, they may fail to receive wage adjustments that fully cover inflation, or they may be expected to pay a higher share of medical benefits. Both of these are also ways to shift the tax burden from the producer to the workers. Depending on how much of the tax can be shifted to workers or consumers, the owners of capital may have to bear all or part of the burden in lower profits.

In other words, the economic incidence of the tax may differ greatly from the statutory incidence. This is often not understood by proponents or opponents of a tax, but it is important for a policy maker to consider when weighing the options of potential tax reform or new taxes. It may be worthwhile to try to educate the citizenry about some of the simple incidence issues. This is much easier if the policy makers understand incidence issues.

Tax incidence studies analyze the shifting of taxes from one party to another. The focus may be on whether the tax is disproportionately borne by people in one region or one industry. Opponents of tobacco taxes often raise issues of regional incidence, because only a few states are major suppliers of tobacco. More often, incidence analysis focuses on how people of different income levels are affected by taxes. The economic burden will fall on those who actually pay the tax through higher prices, lower profits, or lower wages.

Often, this burden is shared between two or three groups, as only partial shifting occurs.

An incidence analysis may or may not take into account the effects of other taxes and of government expenditures. The full incidence of a tax depends not only on how the money is raised, but also on how it is spent. If revenues raised by tax are targeted or legally restricted to a particular use, the analysis should ask how the benefits of the expenditures are distributed across the population.

Often, tax revenues are not targeted to any particular use. In that case, the simplest kind of incidence analysis is appropriate. It is based on the assumption that other taxes as well as the makeup of government expenditures stay the same. Therefore, the only necessary effects to consider are those of the new tax. Alternatively, a tax incidence study could examine the effects of substituting one tax for another, with the level and composition of government spending staying the same. If property taxes are reduced, while sales taxes are raised to make up the lost revenue, this would be an appropriate question to study. The most complex variant of incidence analysis examines the spending effects as well as the taxing effects on various population groups.

TAX INCIDENCE IN COMPETITIVE MARKETS

It is commonplace to allege that profit-maximizing businesses will pass the tax on to customers.^a However, in competitive markets, where there are so many businesses that no one business can set the price on their own, this is unlikely to be the case, at least in the short run. Raising prices by the full amount of a new tax would reduce sales and hurt profitability. In Fig. 1, a tax of \$1 per item is levied on the seller, causing the supply curve to shift to the left. While producers were willing to sell 400 units at a price of a \$1.50 each, they now have higher “costs” due to the \$1 tax. Ideally,

^aRecall that firms in a competitive market produce the quantity at which marginal revenue (or price) equals marginal cost. An excise or sales tax raises the marginal cost of production, leading the market determined price to rise also.

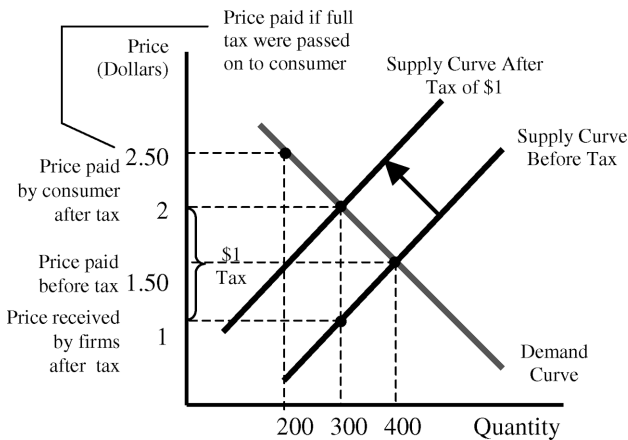


Fig. 1 Half of new tax borne by buyers.

producers would like to continue to sell 400 units and charge \$2.50 per unit, recouping the entire burden of the new tax. Buyers, however, will not purchase 400 units if they have to pay \$2.50. In fact, the only way to get them to buy 300 units is to share the cost of the tax with them, charging a market price of \$2.

The share of the tax actually passed on to buyers is determined by the price elasticity of consumers' demand for the good or service (their responsiveness to a price change) and the price elasticity of supply of the sellers. In Fig. 1, half of the new tax will be paid by buyers in higher prices and half will be borne by sellers. The new equilibrium price (after tax) is \$1.50 per unit. Total sales have fallen from 400 units to 300, but not as sharply as they would have if the sellers had insisted on imposing the entire tax on buyers. The desire of sellers to maximize profits leads them to accept part of the tax burden rather than see sales plummet. This principle holds true whether the tax is levied as a percentage of sales price (an *ad valorem* tax) or per unit sold, as in the tax on gallons of gasoline or packs of cigarettes.

In Fig. 2, a much larger share of the upward shift in supply due to new or increased taxes is reflected in the price paid by consumers than was in Fig. 1. A \$1 tax results in a price increase of .90, with only .10 of the dollar being borne by sellers. The short-term market for gasoline would fit this model well. Gasoline purchases are known to be quite unresponsive to price changes (in economic terms, inelastic) in the short run. This is because it takes time to change the fuel efficiency mix of vehicles on the road and to alter established driving patterns between home and place of work. Even at a higher price, gasoline expenditures are a small part of the budget relative to automobile and home prices. Therefore, a large portion of a gasoline tax will be passed on to consumers, because it will impact the quantities they buy slightly. The supply of gasoline is more responsive to price changes (elastic)

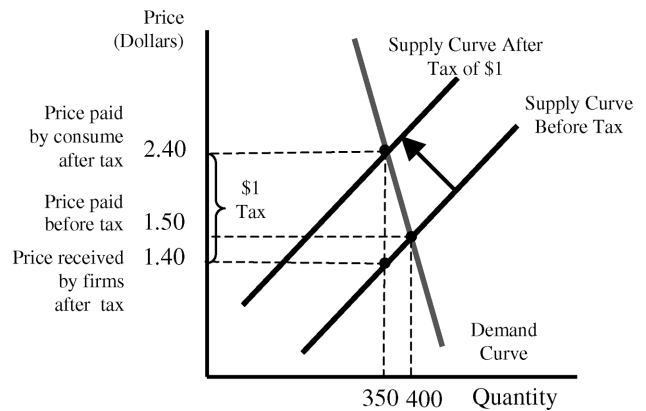


Fig. 2 Ninety percent of new tax borne by buyers.

than is the demand for gasoline, making it possible to shift a large part of the tax burden forward to consumers.

Over time, consumers find ways to economize on more expensive gasoline. They purchase vehicles with better mileage performance and alter their living patterns or place of work so as to reduce the need for gasoline. As a result, sellers of gasoline will find themselves bearing more of the cost of a tax as time goes by and the elasticity of demand increases. The long-run demand for gasoline is more elastic in Fig. 3 than was the short-run demand in Fig. 2. Although the supply of crude oil is fixed in the long run, higher prices can encourage more exploration or make high-cost drilling situations profitable, causing some increase in supply, as in Fig. 3. Consumers do not bear as high a share of cost as they did in Fig. 2.

TAX INCIDENCE IN MARKETS WHERE THERE IS SOME MONOPOLY POWER

The way in which tax burdens are shifted away from the party with statutory responsibility is less clear

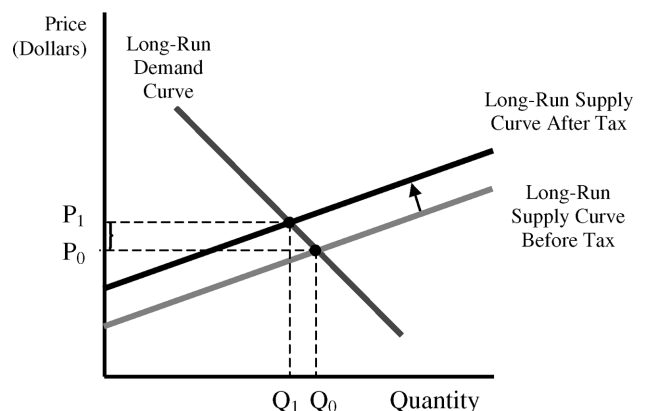


Fig. 3 Supply and demand for gasoline before and after tax.

when sellers of goods or services have a substantial degree of monopoly power. A new tax raises unit costs of production for these firms also. But although these sellers are able to maintain a price greater than their marginal (or incremental) cost of production, they may find it impossible to pass on the full amount of a new tax to buyers. Because sellers can make a profit, even in the long run, they also have some ability to forego profit.

The relative responsiveness to price changes of demand and supply also matter for these businesses. As in a competitive market, if buyers are highly responsive to price changes, they will cut back their purchases when faced with the prospect of paying the entire tax. In response, the business owner with some price setting power is likely to choose to increase price little and to accept a lower level of profit. In other cases, where consumer demand is inelastic (does not respond much to price changes), the price charged could actually increase by *more* than the amount of the tax. The Rosen and Stiglitz works (referenced below) contain examples of these cases.

WHO REALLY PAYS THE TAXES?

If we consider the total amount of taxes (federal, state, and local) paid by all sectors of the economy, the tax burden is 30 to 33% of income for individuals in all but the very highest and very lowest income groups. The tax system as a whole is roughly proportional, meaning that people of low, middle, and high incomes pay approximately the same percentage of their income in taxes. This result is a product of the combined progressivity of income and estate taxes, where higher-income individuals generally pay a higher percentage of income in taxes, and the regressivity of sales and excise taxes which bear more heavily on lower-income individuals.

Some taxes, such as the social security payroll tax and the property tax, are largely proportional with elements of progressivity or regressivity in them. The corporate income tax, discussed further below, is like all business taxes, in that it must be paid by some human entity. It may be paid by corporate shareholders through lower profits, by workers through lower wages or benefits, by suppliers of other inputs through lower prices, or by consumers of the final product through higher prices. Once the incidence of all of these taxes is taken together, the result is a remarkably similar tax burden in terms of percentage of income across income groups. Our overall tax system in the modern-day United States has little income redistribution built into it. However, increasing or decreasing any one tax can have a differential effect on different income groups.

The Individual Income Tax

State governments rely more heavily on the income tax today than in earlier times. Sometimes, there is a progressive structure of rates, rising as income rises. Others apply one flat rate to all taxable income. When examining the incidence on different income groups, it is important to look at effective tax rates rather than statutory rates. If the calculation of taxable income to the state is based on current federal standards for calculating income taxes, an income tax with a single rate for all payers still has a mild degree of progressivity. This is because a smaller fraction of gross income is collected at lower income levels than at higher income levels. On the other hand, if adjustments and deductions allowable from gross income favor high-income individuals, statutory rates which appear progressive might be closer to a proportional system.

The Corporate Income Tax

As mentioned above, taxes on business are often shifted away from the owners of the business to workers, suppliers, or consumers. The increasing globalization of capital markets means that the rate of return to capital is established in highly competitive world markets. Because financial capital flows easily from one region, or country, to another, its supply to any one state, locality, or even nation is highly elastic. This makes it difficult to shift a tax to the owners of capital in the form of lower profits. As a result, the tax is likely to be borne by consumers, workers, or in some cases, suppliers of inputs. More of the tax will be shifted to those who are unresponsive to price or wage changes in their demand or in the supply they offer to the corporation. This pattern of tax shifting is likely to change with the passage of time, as the ability of the various parties to change purchasing patterns or other behavior increases. Once again, those who are most flexible are likely to be able to avoid the tax.

Payroll Taxes

The social security, or payroll, tax falls equally on employer and employee from a statutory perspective. Employers match the contribution of their employees to the social security system. However, in this case, the supply of labor is less elastic than the demand for labor.^b As a result, most of the employer tax is shifted

^bMales 25–55 show almost no responsiveness in their labor force participation to changes in overall wage levels. Single females have a similarly low wage elasticity. It is only the increasing participation of married women that brings much elasticity to the supply of labor.

backward to workers in the form of lower real wages. Given the elasticity of demand for the particular product or service, some of the tax may also be shifted to buyers in the form of higher prices. This is an excellent example of how the statutory division of the tax between employer and employee has little effect on who bears the actual tax burden.

Rapid increases in the social security tax rate in the 1970s and early 1980s are one culprit when we seek to explain why low-wage workers received pay increases less than the rate of inflation during this period. The additional payroll taxes that employers paid for each hour of work absorbed the productivity gains that would normally be used for increases in worker pay.

Sales or Excise Taxes

In assessing the incidence of an excise or sales tax, there are four key questions to ask:

1. Are the sellers of the affected products responsive to price changes in terms of the quantity they offer on the market? Are they easily able to shift into other lines of production or service which will not be taxed? Fig. 1 shows a greater elasticity of supply than does Fig. 4, in which the supply of land is completely fixed and, therefore, unresponsive to price changes.
2. Are the buyers of the products responsive to price changes in terms of the quantity they will buy? Is this answer different for some buyers than for others?
 - a. Are they able to easily substitute other goods or services that will not be taxed and thus resist the seller's attempt to pass on the tax? The demand curve for gasoline

in Fig. 2 shows far less responsiveness to price than the demand curve in Fig. 1. As a result, more of the tax is shifted to the buyers in Fig. 2 than in Fig. 1.

- b. Is the item a substantial one in their budget or a fairly small part of their total expenditures, even with the tax? The more important the item in their budget, the more responsive buyers will be to a price change, making it harder to shift the tax to them.
3. How will each of the above answers change over time? How much time will it take for the effects to be noticeably different? Generally, the elasticities on the demand and supply sides will be greater as time passes than they are in the near term (Fig. 3 versus Fig. 2).
4. Is the market a reasonably competitive one in which many buyers and sellers jointly determine price, with no large buyer or seller exerting market power? Or, is it a less than competitive market, where one entity, or a group acting in collaboration, dominate pricing and production strategy?

Property Taxes: Unimproved Land Versus Taxes on Improvements and Structures

Property taxes are levied by local governments on the owners of property, regardless of whether another individual or business rents the property for its use. The component of property value due to raw land and its location will pay most of the tax burden in the long run, because it is immobile and cannot move to another location to avoid taxes, as in Fig. 4. There is a fixed amount of land in supply in any one region or country that will not be affected by the level of taxes associated with that land. As a result, a tax on unimproved land, in contrast to taxes on improvements or structures, cannot be shifted to other parties. This result led the eighteenth century economist Henry George to call for a "single tax" on land. Pittsburgh is the only U.S. city that taxes unimproved land more heavily than property improvements, with some interesting results.^[1]

On the other hand, property improvements can be moved from one area to another and can more easily resist efforts to shift a tax to them. Taxes on business personal property are often a source of concern in local economic development efforts, because they discourage investment in plant and equipment, which is generally highly mobile given a long enough time horizon. Only if a locality has some specific advantages of location, natural resources, or cultural amenities will it be able to charge these taxes without losing business investment over the long term. It is the net incidence of

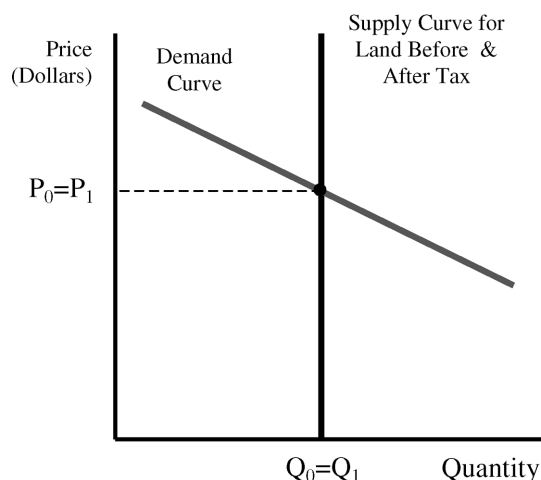


Fig. 4 Supply and demand for land before and after tax.

taxes as well as of beneficial public expenditures that influence the firm or individual.

LOCAL TAXES: WHAT INCIDENCE ISSUES ARE MOST IMPORTANT?

It is important to consider the sources of tax revenue and the uses to which those expenditures are directed when evaluating the incidence of taxes. On the sources side of the ledger, the increasing mobility of skilled labor and of capital in the modern economy mean that the major impact of local taxes is borne in the long run by the owners of land, along with unskilled labor. Skilled labor and capital can migrate to other areas with lower taxes, as long as those areas provide enough in public services and amenities to meet their needs. Land is obviously immobile. Unskilled labor has much lower mobility than skilled labor, because the costs of relocation must be borne upfront and may outweigh the benefits.

Debt-financing, as well as taxes, lowers the value of land or property through a process called capitalization. Potential future owners factor in the additional cost of tax payments to the price they are willing to pay for land. On the other side of the ledger, local expenditures on education, roads, park, and police and fire protection raise the value of property, because buyers will pay more for property associated with these amenities. In the end, the net effect of taxes or debt-financing on property values depends on the efficiency with which government provides services and the values and preferences of current and future residents. This is an important point to remember when proponents or opponents are arguing only the negative effects of taxes or only the positive effects of expenditures.

Arguments for and against various local taxes often ignore the effects of tax shifting and the evidence regarding likely incidence. For example, areas with high tourism favor sales taxes as a way of getting tourists to pay local costs. But if tourists are highly mobile, able to choose between various destinations offering sun and sand (or shopping and culture), sales or lodging taxes may simply depress the prices charged for housing, food, and other amenities. The tax would then be shifted to owners of establishments catering to tourists, and possibly to their workers in the form of lower wages.

On the other hand, if an area has a unique attraction that tourists want to see (the Grand Canyon or the Statue of Liberty), spending is influenced less by price comparisons. More of the tax burden can be shifted to tourists and away from the proprietors of restaurants, lodging, and tourist attractions. Applying the principle that taxing items that are large in the

budget causes more consumer response than taxing items that are small portions of the budget, for example, exempting ski tickets from the sales tax may have little effect on the economics of tourism.^c The ski lift ticket is a relatively small part of the total expenditure for tourists who must spend on airfare and lodging.

TAX INCIDENCE: A SUMMARY

As policy makers and managers consider the role of different forms of taxation, it is important to remember the following.

1. The economic burden of a tax does not necessarily lie with the person legally responsible for paying the tax. As a result, debates about business paying its fair share have more political than economic appeal.
2. Taxes on goods or services are often shifted forward to consumers or backward to workers and suppliers. The way in which the tax is divided depends on the relative responsiveness to price change (elasticities) of the parties involved.
3. If the demand for a good or service is fairly elastic (responsive to price changes) and the supply is relatively inelastic, suppliers will bear more of the tax. Conversely, if the demand is fairly inelastic and supply is more elastic, consumers will bear more of the tax. Where the elasticities of demand and supply are equal, as in Fig. 1, the tax is equally shared.
4. Over a long period of time, individuals and businesses are likely to be able to adjust their behavior much more than is possible during a short time period. This ability to adjust makes them more responsive to the additional costs of a tax and makes it more difficult to shift the tax to them. The relative elasticities of demand and supply may change over time, and with them, the pattern of tax incidence (see Fig. 3 versus Fig. 2).
5. The immediate effect of changes in the expected future stream of taxes, debt, or expenditures on the price of land is called capitalization.
6. An increase in taxes or bonded indebtedness has the immediate effect of lowering the value of land, even though the tax or bond payments will be made over time.
7. If expenditures associated with the tax or debt are viewed as desirable, land values may remain stable, or even rise, rather than fall due to the tax.

^cDespite this, states like Colorado who are dependent on the ski industry often exempt lift tickets.

REFERENCE

1. Oates, W.E.; Schwab, R.M. The impact of urban land taxation: the pittsburgh experience. *Natl. Tax J.* **1997**, *50*, 1–21.

BIBLIOGRAPHY

1. Fullerton, D.; Roger, D.L. *Who Bears the Lifetime Tax Burden?* Brookings Institution: Washington, DC, 1993.
2. Kotlikoff, L.; Summers, L.H. Tax incidence. In *Handbook of Public Economics*; Auerbach, A.J., Feldstein, M., Eds.; North Holland: New York, 1987; Vol. 2, 1043–1092.
3. Metcalf, G.E. The lifetime incidence of state and local taxes. In *Tax Progressivity and Income Inequality*; Slemrod, J.B., Ed.; Cambridge University Press: New York, 1994; 59–88.
4. Musgrave, R.A. *Progressive Taxation, Equity, and Tax Design*; Slemrod, J.B., Ed.; Cambridge University Press: New York, 1994; 341–356.
5. Rosen, H.S. *Public Finance*, 5th Ed.; Irwin McGraw-Hill: Boston, 1998.
6. Sheffrin, S. Perceptions of fairness in the crucible of tax policy. In *Tax Progressivity and Income Inequality*; Slemrod, J.B., Ed.; Cambridge University Press: New York, 1994; 309–314.
7. Stiglitz, J.E. *Economics of the Public Sector*, 3rd Ed.; W.W. Norton & Company: New York, 2000.

Team Building

R. Wayne Boss

Leeds School of Business, University of Colorado at Boulder, Boulder, Colorado, U.S.A.

Mark L. McConkie

Graduate School of Public Affairs, University of Colorado at Colorado Springs, Colorado Springs, Colorado, U.S.A.

INTRODUCTION

According to Katzenbach and Smith, “A team is a small number of people with complementary skills who are committed to a common purpose, performance goals, and approach for which they hold themselves mutually accountable.”^[1] While some of these characteristics are obvious, others need more explanation.

Small number of people: Although 50 or more people can theoretically constitute a team, a group of this size breaks into subgroups to do their work. For all practical purposes, most teams range from 2 to 25, and the majority are less than 10.^[1]

Complimentary skills: Effective teams develop the right mix of skills, which include technical or functional expertise, problem-solving and decision-making skills, and interpersonal skills (i.e., those that relate to such things as effective communication, conflict management, listening, and supportiveness).^[1]

Committed to a common purpose: The best teams have a clearly defined common purpose that is clearly understood and that motivates members. In short, it provides a fundamental reason for the team’s existence as a group.^[1]

Specific performance goals: Since specific performance goals are an integral part of the purpose, effective teams consciously take steps to define their team product, create opportunities for clear communications and constructive conflict, identify specific objectives that are attainable and measurable, share power and influence with one another, allow members to experience small wins, and challenge team members.^[1]

Common approach: Teams members need to define how they will work together to accomplish their purpose. Specifically, they need to define who will do what jobs, how they will schedule their work, how decisions will be made, what skills they will need, and how that will happen. They also need to pay attention to the social, economic, and administrative aspects of their work.

Mutual accountability: No group ever becomes a team until it can hold itself accountable as a team; and accountability is impossible without commitment and trust. Accountability is central to team effectiveness.

This includes more than merely having the team leader hold members accountable; it means that people are willing to hold one another accountable to keep the promises they make to themselves, which builds both commitment and trust—two essential elements to any effective team.^[1]

WHAT IS TEAM BUILDING?

Team building activities are interventions designed to build groups of people into teams that have the above characteristics. Although specific designs of team-building interventions vary widely, according to Golembiewski, they typically have the following characteristics:

1. involve as participants individuals who are hierarchically and/or functionally involved in some common flow of work.
2. deal with organizational entities whose members have real and unresolved issues with one another.
3. involve the mutual development of images as a basis for attempting to highlight unresolved issues.
4. involve the sharing of those images or perceptions as a first step toward mutually working through any relational problems.
5. assume that significant organizational problems often are caused by blockages in communication. Confrontations “free up” people to “level” in communicating and thus set the stage for authentic interaction and effective problem solving. Some objective dilemmas such as a critical lack of money cannot be resolved by confrontation designs. In such cases, the ideal outcomes are greater clarity in communicating about the dilemmas and greater willingness to collaborate in doing as much as conditions permit.
6. are short-cycle affairs (lasting one or two days), compared to a typical stranger experience in sensitivity training, which lasts two weeks.
7. are often seen as a springboard for organizational action.^[2]

Successful team-building designs typically result in greater skill development, increased commitment to a common purpose, development of specific performance goals, increased collaboration and agreement on a common approach, and higher levels of mutual accountability. Additional results include more effective communication, a better understanding of one's colleagues and the nature of the problems that affect them, an action plan for dealing with those problems, and a higher level of confidence and interpersonal trust among the participants.

THE DIFFERENCES BETWEEN TEAM BUILDING AND TEAM DEVELOPMENT

Historically, team-building interventions have been conducted almost exclusively by an outside person who is a professional consultant in the field of human resource development or organization development. An outside person could mean either someone from outside the organization or an internal person who is employed by the organization but is outside the work group planning the team building activity. Dyer argues that, "The use of a consultant is generally advisable if a manager is aware of problems, feels he or she may be one of the problems facing the work unit, and is not sure exactly what to do or how to do it, but feels strongly that some positive action is necessary to pull the work group together for more effective operation."^[3]

Team building activities can be intense, in terms of the feelings generated, depending on the degree to which participants confront problems and address issues, particularly those that are interpersonal in nature. These interventions require an extensive knowledge of group dynamics and a high level of skill in facilitating group processes and in dealing with conflict.

Team development interventions have the same objectives as team-building designs, in terms of enhancing problem-solving abilities by dealing with issues that stand in the way of increased effectiveness. However, "The team development facilitator is not a professional, but the team's own formal leader or a designated member of the team. The activities that can be conducted by a team leader or member, who is generally not...[a] professional, are, of necessity, different from those designed to be facilitated by a professional: they are less intense, produce less affect, and demand less of the person who is facilitating and the team members as participants. After a team's members have been introduced to team building by [an outside consultant], that professional can suggest that they follow up with team-development activities to enhance their ability to resolve their own issues."^[4]

THE PHASES OF TEAM BUILDING

The four phases of a team-building intervention include: data gathering, diagnosis, implementation, and follow-up.

Data Gathering

Prior to the team-building meeting, the consultant interviews the participants to identify issues standing in the way of effective teamwork and goals for the meeting. Typical questions include those that help the participants identify group strengths, challenges and problems, leader effectiveness, interpersonal issues, and things people would like to accomplish. Participants can also complete questionnaires designed to measure various aspects of effective teamwork. Additional data can also be gathered by direct observation of group behavior in meetings.

Diagnosis

The consultant thoroughly examines the information from the interviews, questionnaires, and observations, and identifies problems that prevent team members from working together effectively. The importance of this diagnostic phase cannot be overstated. Indeed, the success of the team-building intervention depends on its accuracy. Consultants should always do a diagnosis before prescribing treatment.

Once the diagnosis is completed, the consultant identifies the intervention that will help participants resolve the problems and achieve the goals of the meeting. The general diagnosis and recommended design are then shared with the team leader and members to ensure ownership of the diagnosis and commitment to the meeting. The consultant must exercise caution to avoid opening up problems that cannot be immediately resolved. This is particularly important if the team-building meeting is scheduled more than 24 hours after this feedback meeting.

Implementation of the Team-Building Design

Participants meet together in a room that is comfortable, private, and free from interruptions. The consultant shares the detailed diagnosis and then helps the group build an agenda of what they want to talk about and goals for the meeting. The agenda items are then prioritized, and the group goes to work on the agenda. The consultant's responsibilities include group facilitation, clarifying issues, providing theory input, and documenting decisions made during the meeting, including what will be done, who will do it, when it will

be done, and how it will be evaluated. Mechanisms should also be put in place for how team members will evaluate the group's effectiveness once they get back home.

Follow-Up

Follow-up activities ensure that the team-building action gets implemented back home. This can take place during group meetings and/or individual meetings between the participants and the leader to determine the extent to which they have kept their team-building commitments and to identify new issues that impact the group.

A TYPICAL TEAM-BUILDING DESIGN

A widely used team building intervention^a called a "perception sharing" or "Three-Dimensional Image" (3-D Image) design,^[6] is intended to deal with the dysfunctional aspects of conflict as they affect the perceptions that members of a group have of one another. The group members share perceptions and then confront and resolve whatever problems they indicate. Each 3-D Image confrontation team-building session is conducted off-site and lasts three days.

The participants from each team are divided into two groups: top management personnel and subordinates. The subordinates were asked to respond independently to three questions:

1. How do we see ourselves as members of this work group?
2. How do we think [the leaders] see us as members of this work group?
3. How do we see [the leaders] as members of this work group?

The leaders respond to three similar questions:

1. How do we see ourselves as members of this work group?
2. How do we think they [e.g., our subordinates] see us as members of this work group?
3. How do we see them [our subordinates] as members of this work group?

Both groups respond to their questions in separate rooms, without the aid of the consultants. When the number of subordinates is relatively large (eight or more people), the subordinates are divided into smaller groups of four or five people each.

All participants are asked to write their responses to each question on separate sheets of newsprint.

They are asked to be as explicit as possible, using adjectives and short phrases, and to keep in mind specific examples they might later use to illustrate each descriptive phrase. They are further instructed that consensus is not required for the phrase or descriptive statement to be posted. Rather, if different people have diametrically opposite perceptions, both perceptions are to be posted, since the purpose of the exercise is to share as much accurate information as possible.

Once the responses to each question have been written on the newsprint, the various sheets are taped to the wall, side by side, and silently reviewed by all participants. After everyone has read all the responses, one person from each group reviews the group's responses to the various questions and gives examples of specific behaviors that might have contributed to those perceptions.

The basic ground rule for this portion of the session is that questions are to be directed toward clarifying and/or illustrating specific perceptions. At this stage, agreement is neither expected nor necessary. It is important, however, that the participants understand what is meant by each descriptive phrase and why group members perceive the situation the way they do. Those items about which there is disagreement or which seem to be problems are systematically flagged for later discussion. The review of the various responses posted on the newsprint typically generates a great deal of energy, and much interest is manifested in both the perceptions posted and the examples of specific behavior that precipitated those perceptions.

The 3-D Image procedure helps to clarify perceptions, generate valid information, and identify current problems. Specifically, the approach is effective for a number of reasons:

1. Basically, participants need such information, disconcerting or even initially hurtful though it may be.
2. Participants usually understand that the best—indeed, perhaps the only—way to raise the probability of receiving such needed information in the future is to be accepting of the 3-D Images in the present. It should be noted that acceptance does not necessarily mean agreement. "I can understand how you see it that way," such a case of acceptance without agreement might go, "but I hope you recognize there are many things about which reasonable people can and do differ."
3. Confrontation with the 3-D Image procedure is a shared experience that can build mutual identification and understanding, which is what many participants are seeking. The design is accepted and valued, consequently, and, in a sense, made to work.

^aMuch of this discussion in this section is adapted from Ref.^[5].

4. Most individuals are uncomfortable if their verbal or non-verbal behavior is at some substantial variance with what they really know, believe, or feel. That lack of comfort increases sharply if the persons suspect that relevant others are being similarly incongruent. The analysis of 3-D Images usually helps to reduce such variance by encouraging a mutual escalation toward openness and owning.
5. Substantial agreement typically exists between pairs of 3-D Images.... This statement almost always increases the participants' sense of mutual competence and acceptance by confirming that one person or group shares perceptions with another, as well as by signaling that a real process of exchange has been begun.^[7]

Once the responses on the newsprint have been sufficiently explained, the flagged items are then addressed, beginning with the issues the subordinates have with the leaders, then moving to those that the leaders have with the subordinates.

Central to the success of the design is the contracting stage in which participants negotiate decisions on issues associated with the various perceptions. As an essential element of the design, the consultant documents the solutions and individual commitments made during this stage. These solutions and individual commitments (often numbering six to ten pages of single-spaced typewritten material) are reviewed periodically throughout the meeting to ensure that what is recorded is exactly what the participants intend. The list then constitutes an action agenda and is made available to all participants and used as a basis for follow-up at a specified time.

Additional activities during the three-day meeting can include the completion of questionnaires relating to interpersonal communication and participants' needs in the work environment. In addition, the consultants give short lectures that provide a theoretical perspective on the processes taking place at a given moment and the difficulties team members are experiencing with those processes.

THE PROBLEM OF REGRESSION

A real limit exists on what can be accomplished in a short team-building session.^[8] Without systematic follow-up, the positive results achieved by the team-building design may be negated. Indeed, a basic problem with off-site team-building sessions is the regression or fade-out that often occurs after the participants return to the home organization.^[9] Typically, the high level of interpersonal trust and cooperation may last for several weeks, at best. The organization's

climate then begins to deteriorate toward its previous conditions, requiring an additional off-site session to work through problems and resolve interpersonal issues that either remain unresolved or have arisen since the last meeting. A number of interrelated variables contribute to this tendency toward regression, including unfulfilled expectations, low tolerance for error, lack of process skills among members, lack of accountability, consultant dependency, turnover among key staff members, changes in policies and procedures, a lack of cultural preparedness back home, a lack of political support, a lack of a clear reward system to reinforce new behaviors, and a lack of consultant availability.

HOW TO PREVENT REGRESSION

In order to prevent regression, a team-building session must be seen as part of a larger and ongoing change effort. Research has shown that one-shot team-building efforts do not produce the desired long-range results and that few, if any, changes take place without effective follow-up.^[3]

One way to prevent regression following team-building meetings is for the leader to begin holding a regular personal management interview (PMI) with his or her direct reports. A PMI is a regular, private meeting between the leader and each of the people who report to him or her. Detailed notes are kept on all action items, and both people receive a copy of the notes at the conclusion of the meeting. The first few minutes of each meeting are spent in follow-up of the action items of the last meeting.

The purpose of the PMI is to meet the needs of the individual team members. Hence, appropriate areas for discussion include administrative and organizational problems, training in administrative and management skills, interpersonal problems that may exist between the two participants, information sharing to bring people up-to-date, individual and organizational needs, successes and achievements, and personal problems. The last few minutes are spent in a review of the action items generated during that meeting.

Even though the PMI appears to be a very simple intervention, it has a powerful impact on both the individual team and the organization. Research indicates that when leaders hold regular PMIs with their direct reports, regression can be prevented for as many as 20 years with no additional interventions.

REFERENCES

1. Katzenbach, J.R.; Smith, D.K. *The Wisdom of Teams*; HarperCollins: New York, 1993; pp. 45, 47-54, 60-61.
2. Golembiewski, R.T. *Approaches to Planned Change. Part 2*; Marcel Dekker: New York, 1979; 139-141.

3. Dyer, W.G. *Team Building: Issues and Alternatives*, 2nd Ed.; Addison-Wesley: Reading, MA, 1987; 52 pp.
4. Pfeiffer, J.W. *The Encyclopedia of Team-Building Activities*; Jossey-Bass/Pfeifer: San Francisco, 1991; 4–5.
5. Boss, R.W. *Organization Development in Health Care*; Addison-Wesley: Reading, MA, 1989; 90–95.
6. Golembiewski, R.T.; Blumberg, A. Confrontation as a training design in complex organizations: attitudinal changes in a diversified population of managers. *J. Appl. Behav. Sci.* **1967**, *3* (4), 525–547.
7. Golembiewski, R.T.; Kiepper, A. Marta: toward an effective open giant. *Public Adm. Rev.* **1976**, *36* (1), 46–60.
8. Boss, R.W. Team building and the problem of regression: the personal management interview as an intervention. *J. Appl. Behav. Sci.* **1983**, *18* (1), 67–83.
9. Golembiewski, R.T.; Hilles, R.; Kagno, M.S. A longitudinal study of flexi-time effects: some consequences of an OD structural intervention. *J. Appl. Behav. Sci.* **1974**, *10* (4), 503–532.

Telehealth and State Government Policy

Mary Schmeida

*Department of Nursing Research and Innovation, Cleveland Clinic,
Cleveland, Ohio, U.S.A.*

INTRODUCTION

Telehealth represents a new paradigm in health care. Considered both an administrative reform and regulatory policy, it draws on electronic government and electronic commerce. Although no authoritative consensus on the definition of telehealth has been reached, Congress defines it as the use of electronic information and telecommunications technologies to support long-distance clinical health care, patient and professional health-related education, and public health care and administration.

In the era of technology, state policy makers find telehealth important. Advanced communication technology can provide health care services to state residents. To governments, Internet technology can make product and service delivery both cost-efficient and convenient. As a powerful tool for reshaping government, it is expected to help build economic strength, allowing government to be catalytic and enterprising. State policy makers are incrementally passing telehealth-enabling legislation, and a patchwork of laws exists today.

TELEHEALTH EVOLUTION

History documents the beginnings of telehealth throughout the twentieth century. The Nebraska Project of the 1950s, using closed-circuit television, is generally considered the first comprehensive attempt at telehealth, advancing clinical psychiatry and medical instruction^[1] and promoting remote consultation to prisons, courts, and schools. Projects of the 1960s and 1970s involved the U.S. military and the National Space and Aeronautical Administration,^[2] with innovations advancing emergency and cardiology care. Recent evolution is dramatic, combining audiovisual components of traditional forms of media with the interactivity and speed of telephone and e-mail, permitting the exchange of patient information regardless of geographic distance; facilitating interaction between practitioners and consumers; transmitting images via satellite to experts for interpretation; and entering surgery, as seen in experimental robotic coronary artery bypass surgery.

CONDITIONS DRIVING TELEHEALTH POLICY

Several conditions are driving telehealth policy across the states. External conditions driving state telehealth-enabling laws include the lack of access to medical specialists, spiraling costs of health care, and growing consumer sophistication and demand for quality health care. Political conditions driving policy are the bureaucratic issues of a top-down government that no longer works to satisfy citizen demands for quality goods and services. A flexible and agile government can respond to the changing external environment, with quick information technologies to meet citizen demands, breaking down bureaus and reshaping government.

Lack of Access to Medical Specialists

Rural states have a great lack of access to specialists. Across the states, rural regions are facing a growing maldistribution of medical specialists, hospitals, and health care resources. Specialists and advanced diagnostics have been more readily accessible in urbanized areas versus remote rural settings. For example, West Virginia is among the most rural, poorest, and most elderly of the states, and has a lower-than-national-average supply of physicians.^[3]

The original goal was to improve consumer access to health care. Inequitable distribution of medical resources, in relation to the distribution of need in inaccessible state populations, has led to the expansion of the national information infrastructure, making telehealth a feasible solution to minimize disparities. Rural consumers can now seek video teleconsultation at access points or facilities equipped for telecommunications in remote areas. For example, pediatric populations in rural regions can receive early disease intervention that has been previously delayed due to distance and cost. Although the lack of specialty services in rural regions has driven the telehealth idea in the past, professional shortage areas can also include urban regions of a state because metropolitan isolates can also have difficulty accessing services.

Spiraling Costs of Health Care

Spiraling health care costs is a condition underlying telehealth activity. The urban consumer price index shows a rising yearly trend for the cost of health care from 74.9% for 1980, to 162.8% for 1990, to 272.8% for 2001.^[4] The growing elderly population demanding costly services is also driving policy. The projected rise in Medicare spending is a function of the aging population's longevity requiring expensive specialized services, the retirement of baby boomers along with the declining ratios of workers to Medicare beneficiaries, and the overall increased costs of health care such as the increased cost of diagnostic equipment.^[5] Telehealth as a cost containment strategy for entitlement programs can reduce: opportunity costs, costs associated with overtreatment and medical error, out-of-pocket expenses, and costs for insurance companies. State prison systems using telehealth are not only avoiding costs of security guards and transportation, but increasing public safety.

States are seeking creative financial arrangements. Because initial capital costs can be substantial and there are incremental costs with each use, cooperative initiatives between public, private, and non-profit sectors can limit start-up and program sustainment costs. For example, Nebraska created a health care trust fund awarding grants for infrastructure development. Creative arrangements have led to consortiums of neighboring states implementing projects. Regional networking can minimize program costs while sustaining projects, inviting interstate commerce, and bringing new telehealth policy ideas to the table.

Growing Demand for Quality Health Care

Telehealth is expected to improve the quality of U.S. health care. Research links poor U.S. health care to underuse, overuse, and misuse of services. With choice as a proxy for quality, government telehealth websites inform residents on options in choosing a health plan and premium, practitioner(s), and hospital setting for services.^[6]

Many states are improving the quality of care their residents receive by implementing innovative telehealth ideas, with Texas, Hawaii, North Carolina, California, Virginia, and Michigan as leaders in scope of implementation.^[7] For example, telehealth home care can improve quality by: preventing hospital readmissions; predicting and avoiding chronic illness setbacks; and bringing early medical intervention to remote populations, thus improving clinical outcomes.

STATE TELEHEALTH-ENABLING LAWS

There are barriers to expanding the use of telehealth. These include a lack of: uniform statewide physician

licensure statutes; laws regarding disclosure of restricted patient information to unauthorized individuals; interstate reimbursement to practitioners; and strict Internet regulation for the dispensing of pharmaceuticals on-line. Enabling laws create the legal environment that makes telehealth feasible. Although federal legislation has tried to create this environment, the states are taking the lead and leaving a patchwork of enabling laws. However, both state and federal telehealth legislations still lag behind the rapid and dramatic change in technology.

Several states are leaders in passing telehealth-enabling laws. Georgia, Kansas, Louisiana, South Dakota, and Texas are leaders, whereas West Virginia has been relatively inactive in initiatives to overall state telehealth policy. Texas has been revolutionary—assuming a formal position among the states in coordinating overall planning for telehealth development, promoting the idea with public utilities, and setting reasonable telecommunication costs and prison telehealth.^[8]

Telehealth Medicaid Practitioner Reimbursement Laws

States are enacting Medicaid telehealth practitioner reimbursement laws. This administrative reform policy changes previous policy on unfair fee-splitting arrangements for practitioners and the lack of reimbursement for provision of services to Medicaid recipients. Medicare has been on the forefront on this topic, expanding its reimbursement policy to include a variety of originating telehealth sites, shaping policy for state programs, and as a policy example for the private insurance industry. Some state Blue Cross/Blue Shield plans, among other private insurers, are electing to pay for select telehealth services.

Nineteen states have adopted Medicaid telehealth reimbursement laws. As an innovator, Texas requires that reimbursement not be denied due to lack of in-person consultation and also provides parity between telehealth and in-person health care on deductibles and copayments. All states enacting this law reimburse for medical health needs, whereas California, Kansas, Minnesota, Montana, Texas, and Virginia also reimburse for mental health.^[9]

State Health Care Information Privacy Laws

The use of the Internet for patient care is a growing concern. Over the past decade, consumer privacy attitudes have changed, with 60% of Internet users expressing concern that placing medical records online can be violated even if password-protected, and 89% of those seeking on-line health information expressing concern that the website may sell or give

other institutions information about their activities.^[10] Although technology can benefit practitioners, the tradeoffs can jeopardize the integrity of sensitive health information.

States are responding to consumer privacy demands with legislation. Although the Health Insurance Portability and Accountability Act of 1996^[11] provides new safeguards in this area, states must also respond. To date, health privacy statutes vary significantly across the states regarding electronic means for disclosing private health information and storing the information. Twenty states have restrictions on disclosure of patient health information by hospitals and health care entities. For example, Rhode Island prohibits medical information from being disclosed regardless of the institution holding the medical record. Most states have not intended statutes to be comprehensive, thus lagging behind the needs of innovative health care.^[12]

Interstate Physician Licensure Laws

No uniform state licensure laws exist for governing telehealth practitioners. Updated laws would protect consumers and monitor the telehealth activities of providers regulating the practice of physicians across the states. Recognizing the significance of this issue, 17 states have adopted full telemedicine licensure status for physicians. This allows physicians physically located in their home state or foreign country to perform certain patient care services in another state through the use of telecommunications, or to practice medicine in another state. Colorado allows nothing to prohibit consultation between a Colorado physician and a practicing physician in another state. Nine states allow interstate telehealth by certificate, registration, or special licensure, whereas 24 states have no interstate licensure laws for physicians.^[13] Interstate licensure is important to telehealth diffusion because any uncertainty among practitioners about licensure status and associated liability risks will restrict activity.

Internet Pharmacy Regulations

State Internet pharmacy regulations are critical to ensuring quality health care. Internet pharmacy consumers can face injury if prescriptions are electronically issued by pharmacies without knowledge of the consumer's medical history. As an initiative to regulate Internet pharmaceutical activity, President Clinton, in 1999, expanded Food and Drug Administration (FDA) enforcement powers to eliminate unsafe dispensing of drugs on-line. However, the states have traditionally regulated pharmacists and physicians, and are increasingly stepping up to enforce unsafe pharmaceutical electronic commerce activity. Thirteen states are policy

innovators, enacting legislation requiring physician initial examinations before prescribing medication on the Internet—Alabama, Arizona, California, Florida, Iowa, Idaho, Kansas, Maine, Mississippi, Nebraska, New York, Ohio, and Virginia. Although regulations would restrict the pharmaceutical electronic business growth, restrictions would protect consumers and improve the quality of electronic prescription services.^[14]

ECOLOGICAL FACTORS EXPLAINING ADOPTION OF LAWS

Several ecological factors are influencing the adoption of telehealth-enabling laws. Policy diffusion literature shows political, economic, intrastate, and interstate policy networks, and demand factors can affect whether states adopt legislation across issue areas. Using 50 state data with statistical controls, Schmeida^[15] found that ecological determinants matter in explaining the adoption of the four telehealth-enabling laws discussed above. The presence of designated mental health professional shortage areas is found to influence the adoption of all four laws. State legislative professionalism explains the adoption of all enabling laws, except Internet pharmacy regulations. Greater legislative professionalism increases the likelihood that states will adopt privacy laws, but not necessarily telehealth reimbursement and physician licensure laws.

Particular state ecological factors also explain the adoption of specific telehealth laws. Illustrative of these are the following:

- States with higher information technology administration and management capacity, and populations over the age of 64 years are more likely to adopt Internet pharmacy laws.
- States with greater electronic commerce and business regulations are less likely to adopt physician telehealth licensure laws, but are more likely to adopt full licensure status.
- States with a greater percentage of women in the legislature are more likely to pass privacy laws; this is important because women in politics have historically been instrumental in driving health policy.

Such findings provide policy makers with a foundation for understanding this evolving policy area.

CONCLUSION

As an innovative policy, telehealth is important to the state government. Drawing on electronic government and electronic commerce, it has the potential to improve access to health care services, to contain costs

while improving service quality, and to reshape the government. State policy makers are passing enabling legislation, although legislative details vary from state to state. Likely benefits of telehealth policy are cost containment, improved access for underserved populations, and enhanced quality of care.

REFERENCES

1. Smith, H.; Allison, R. *Telemental Health: Delivering Mental Health Care at a Distance*; U.S. Department of Health and Human Services, 1998; 3 pp., <http://www.usdhhs.gov> (accessed November 2003).
2. Office of Health Affairs, U.S. Department of Defense. *Military Health Services System Information Management/Information Technology Strategic Plan*; Defense Technical Information Center, 1996; 1–5. <http://usdod.gov> (accessed November 2003).
3. National Conference of State Legislatures. In *Summary and Analysis of State Initiatives to Promote Telemedicine, Volumes 1 and II*; 2003; 1–91. <http://www.ncsl.org> (accessed September 2003).
4. U.S. Census Bureau Statistical Abstract of the United States: 2002. *Table No. 483*; U.S. Government Printing Office: Washington, DC, 2003; 453 pp.
5. *Centers for Medicare and Medicaid Services*; 1999; 1–9. <http://www.cms.hhs.gov> (accessed November 1999).
6. AARP. *Beyond 50.02: A Report to the Nation on Trends in Health Security*; 2001; 91–117. <http://www.aarp.org/beyond50/> (accessed January 2003).
7. Schmeida, M.; McNeal, R.; Mossberger, K. *Implementing Telehealth in the American States*. Paper to be presented at the 2004 Midwest Political Science Association Conference, 1–30.
8. National Conference of State Legislatures. *Summary and Analysis of State Initiatives to Promote Telemedicine, Volumes 1 and II*; 2003; 1–91. <http://www.ncsl.org> (accessed September 2003).
9. *Centers for Medicare and Medicaid Services*; 2003 ; 1–7 <http://www.cms.hhs.gov> (accessed December 2003).
10. House of Representatives Hearing 107-35. *Opinion Surveys: What Consumers Have to Say About Information Privacy*; U.S. Government Printing Office: Washington, DC, 2001; 1–38. 107th Congress, 1st Session.
11. Public law 104-191. *Health Insurance Portability and Accountability Act of 1996*; US Government Printing Office: Washington, DC, 1996; 1–21. 104th Congress, 2nd Session.
12. National Conference of State Legislatures. *Summary and Analysis of State Initiatives to Promote Telemedicine, Volumes 1 and II*; 2003; 1–91. <http://www.ncsl.org> (accessed September 2003).
13. Office for the Advancement of Telehealth. *2001 Telemedicine Report to Congress*; 2001; 1–26. <http://telehealth.hrsa.gov/> (accessed June 2003).
14. Office for the Advancement of Telehealth. *2001 Telemedicine Report to Congress*; 2001; 1–26. <http://telehealth.hrsa.gov/> (accessed June 2003).
15. Schmeida, M. *Telehealth Enabling Laws: Explaining Adoption and Diffusion Across the American States* Paper to be presented at the 2004 Midwest Political Science Association Conference, 1–35.

Teleology and Utilitarianism

James R. Heichelbech

Graduate School of Public Affairs, University of Colorado at Denver,
Denver, Colorado, U.S.A.

INTRODUCTION

Teleology and utilitarianism describe approaches to ethical thinking that are goal oriented and focused on consequences. Teleology is specifically concerned with purposes or goals, while utilitarianism is the view that an act is ethical to the extent that it satisfies interests or increases happiness. Together, teleology and utilitarianism are contrasted with deontological conceptions of ethics, which are focused on obligations and motives.

TELEOLOGY

Aristotle's Concept of Teleology

From an ethical theorist's standpoint, teleology is a metaphysical concept, especially in the context of Aristotle's ethics. Accordingly, the *Cambridge Dictionary of Philosophy* defines teleology as "the philosophical doctrine that all of nature, or at least intentional agents, are goal-oriented or functionally organized."^[1] For Aristotle, it is a thing's purpose that determines whether it is excellent or virtuous. However, the Aristotelian emphasis on teleology was largely discredited with the emergence of the scientific revolution. Rather than explaining a thing in terms of its purpose, scientists came to believe that things should instead be explained in terms of their formal properties, through principles of nature. Charles Darwin delivered the final blow to teleological science in the nineteenth century, as he argued that nature does not unfold as part of a complex web of purpose, but through natural selection.

The Connection Between Teleology and Utilitarianism

Contemporary philosophers who turn to Aristotle tend to avoid the emphasis on teleology and instead focus more generally on character. Often, the avoidance is an explicit rejection of Aristotle's essentialism, or the view that each thing has an essence. Essentialism is considered problematic in light of its possible implications, such as Aristotle's claim that some people are

natural slaves. The avoidance is also due to lack of theoretical adequacy. However useful it might be in discussions of ethics to say that a person has a purpose, it is not clear that each person's purpose can be determined to the extent necessary for making ethical judgments about that person.

While ethical theorists tend to avoid the teleological element of Aristotle, political philosophers have reintroduced the term in connection with utilitarianism. The basic idea of utilitarianism is that sentient beings have goals and, therefore, have interests that ought to be satisfied. Such interests, then, can be counted as a measure of a being's purpose. Utilitarianism is for this reason considered teleological, not because it is an Aristotelian theory, but because it is concerned with the goals of sentient beings. In other words, the use of the term "teleology" has shifted from the context of character assessment, where goals are an important part of character, to realization of the person's goals.

UTILITARIANISM

Utilitarianism emerged in Great Britain toward the end of the eighteenth century. Consistent with the empiricist tradition prevalent at the time in Great Britain, utilitarians rejected rationalism, arguing that it could not serve as a basis for political or moral justification. Jeremy Bentham and John Stuart Mill argued from common sense that what matters is not the dictates of reason but the capacity for pleasure and pain.

Jeremy Bentham's Utilitarianism

In *The Principles of Morals and Legislation*, Jeremy Bentham (1748–1832) described the utilitarian principle as that according to which we can evaluate actions depending on their tendency to increase or decrease the happiness of those whose interests are in view.^[2] Ethics, therefore, concerns the happiness of individuals and nothing else.

The focus on the capacity for suffering leads to a shift from focus on the moral agent to a focus on the moral "patient" and widens the scope of moral considerability to include all beings with interests, such as non-human animals. Because a capacity for

experiencing pleasure and pain is the criteria for identifying interests, the interests in view are the interests of all who are in a position to experience pleasure or pain as a result of one's actions. But Bentham's utilitarianism is not as altruistic as it sounds. Although he acknowledges that we are sometimes motivated by sympathy, benevolence, love of amity, or love of reputation, the only adequate ethical motives are one's own interests.

With the emphasis on individual interests, Bentham's utilitarianism is useful for political justification and for public policy. The public interest, he argues, is merely the sum of individual interests. Government action, then, is justifiable only to the extent that it tends to increase the happiness of the individuals within the community more than it tends to decrease it. In this respect, public policy becomes inherently normative, as all stakeholder interests carry moral weight.

John Stuart Mill's Utilitarianism

In *Utilitarianism*, John Stuart Mill (1806–1873) presented a more theoretically complete version of utilitarianism.^[3] He followed Bentham in claiming that the only legitimate ethical principle is that of utility—an act is ethically right to the extent that it increases happiness, or pleasure, and wrong to the extent that it decreases happiness or causes pain. However, Mill argues for a much stronger principle, one according to which an act is right to the extent that it produces the most happiness for the most people, not just the happiness of those whose interests are “in view.”

Mill's version of utilitarianism differs from Bentham's in a number of other ways. Whereas Bentham merely described the various kinds of pleasures, Mill distinguished higher pleasures from lower ones. Whereas Bentham used utilitarianism to support modest political ends, such as penal reform, Mill envisioned a much broader horizon of social progress. Finally, whereas Bentham argued that all moral motives can be reduced to self-interest, Mill described sentiments of justice.

The distinction between higher and lower pleasures is driven by Mill's belief that “it is better to be a human being dissatisfied than a pig satisfied” (Ref.^[3], p. 10). His argument is that anyone who has experienced lesser pleasures and more cultivated ones would prefer the latter. “No intelligent human,” he says, “would consent to be a fool” (Ref.^[3], p. 9). It is worth noting that this is essentially an empirical argument, not an assumption about human nature included as a supplement to utilitarianism. If it turns out that most intelligent, educated people regret their heightened sense of appreciation, then Mill would have to abandon the claim that “higher” pleasures are better.

Consistent with his focus on higher pleasures, Mill envisions a much more optimistic, progressive political horizon. Although Mill agrees with Bentham that it is reasonable to expect people to focus on the interests of those in view, it is clear that Mill thinks our focus should be as broad as possible. He even goes so far as to suggest that the evils of the world, such as poverty, can be eliminated through the wisdom of society and good sense of individuals.

However, the most significant difference between Bentham and Mill is found in their accounts of moral motivation. Whereas Bentham concluded that the only motive is self-interest, Mill argued that moral motives include “the internal sanction of duty . . . a feeling in our own mind; a pain, more or less intense, attendant on violation of duty” (Ref.^[3], p. 27). Moral motives are more complex and more altruistic. The sentiment of justice, for example, consists of the belief that someone has been harmed and a desire to punish the harm-doer. Furthermore, Mill argued, it is “by virtue of his superior intelligence . . . [that] a human being is capable of apprehending a community of interest between himself and the human society of which he forms a part, such that any conduct which threatens the security of the society generally is threatening to his own” (Ref.^[3], p. 50). This is significantly beyond the motive of self-interest described by Bentham.

Criticisms of Utilitarianism

There are several difficulties with the principle that one should always act to maximize happiness. First, there is the problem of calculating utility. One must always determine what each person's interests are and what makes that person happy, and then weigh those interests against one another. Some offer the distinction between *act*-utilitarianism, which requires that every action maximize happiness, and *rule*-utilitarianism, according to which one need only act according to a rule the actions conformable to which *tend* to increase happiness. Once it is determined that certain rules tend to maximize happiness, then one need only follow the rule. Even so, critics argue, it is not clear that all interests can be adequately measured, much less compared against one another, such that we could identify the aggregate “maximum” of happiness.

A second difficulty is that some ethical choices justifiable by utilitarianism seem counter-intuitive. Imagine a case in which thousands of people suffer from a condition that is excruciatingly painful, though not life-threatening. Imagine as well that their suffering can only be alleviated by extracting a particular chemical from the brains of a few people, resulting in their deaths. Utilitarianism seems to recommend that we sacrifice the few for the sake of the many.

The calculation seems to suggest that the fewer people sacrificed and the more suffering that will be alleviated, all the more reason to sacrifice the few. But, intuitively, there seems to be something wrong with sacrificing even one person's life to alleviate the suffering of others, regardless of how many there are.

A third problem is that utilitarianism has no way to distinguish legitimate interests from illegitimate interests. Even if the people to be sacrificed in the example above *wanted* to sacrifice their own lives, there is no way to determine whether they *should* want to do this. Utilitarianism tells us that interests should be satisfied, whatever they might be.

Contemporary ethical theorists in the utilitarian tradition tend to favor a slightly less rigid conception of ethics, usually some form of *consequentialism*. Consequentialists agree that interests matter, and so the consequences of our actions matter to the extent that interests are either satisfied or not satisfied. But, consequentialists are not necessarily committed to the view that satisfaction of interests is the *only* thing that matters or that *all* interests matter equally. Whereas Mill was burdened with the task of showing that the unsatisfied human is still happier than the satisfied pig, consequentialists are free to evaluate interests on other grounds, such as basic rights.

TELEOLOGY AND UTILITARIANISM IN PRACTICE

Teleology and utilitarianism have served as inspiration and justification. The teleological/utilitarian perspective in ethics is significant for public administrators for two reasons. First, the growing emphasis on "customer service," as found in closer attention to mission statements and organizational values, explicitly asks administrators to be aware of interests. Second, the teleological/utilitarian approach has been key to justifying a free market economy on a macrolevel and the use of cost-benefit analysis on a microlevel. For this

reason, administrators are better able to interpret fiscal policy goals through a teleological/utilitarian perspective, especially with the accompanying emphasis on measurement and calculation.

Although public administration tends not to be teleological or utilitarian in the broad sense, as its goals are externally determined, legislation is frequently motivated by teleological/utilitarian considerations and therefore might require development of public policy accordingly. Spending limitations, for example, reflect efforts to increase the well being of taxpayers, and we benefit from implementation strategies consistent with the utilitarian spirit of the legislation.

The means by which administrators implement legislation are clearly and directly teleological/utilitarian in character. The values of efficiency and effectiveness, for example, insist on limiting waste and providing the greatest possible value. Put another way, to the extent that administration is inefficient and ineffective, the aggregate utility of taxpayers has been diminished. With this in mind, administrators often look to "results-oriented" management practices as an ethically responsible means of administration. This is contrasted with process-oriented or principle-oriented practices (see Ref.^[4]). A good example of results-oriented management is a performance-based pay system.

REFERENCES

1. Audi, R., Ed.; *The Cambridge Dictionary of Philosophy*; Cambridge University Press: Cambridge, MA, 1995.
2. Bentham, J. *The Principles of Morals and Legislation*; Prometheus Books: Amherst, NY, 1988 (Original work published 1781).
3. Mill, J.S. *Utilitarianism*; Hackett Publishing Company: Indianapolis, IN, 1979 (Original work published 1861).
4. Perry, J.L.; Mesch, D.J. Strategic human resource management. In *Public Personnel Management: Current Concerns, Future Challenges*; Ban, C., Riccucci, N.M., Eds.; Longman: New York, 1979; 21-34.

Thailand

Brian Brewer

*Department of Public and Social Administration, City University of Hong Kong,
Kowloon, Hong Kong*

INTRODUCTION

Thailand has experienced frequent changes in its national government over the past 70 years, while the rapid economic growth of the 1970s and 1980s that was halted so dramatically by the 1997 Asian financial crisis has resumed once again. The 1997 constitution, through its reform of the electoral system and the establishment of a range of institutions such as the National Counter Corruption Commission (NCCC), the Constitutional Court, and the Office of the Ombudsman, provides an institutional framework to support principles of good governance. Major public administration challenges include the implementation of public service reforms such as decentralization, a continuous battle against corruption at all levels, ensuring a free media, and facilitating the further development of Thailand's civil society.

GOVERNMENT IN THAILAND

Thailand, with a population of 63.3 million, is generally regarded as a "middle income" country and, though less prosperous than its more industrialized neighbor Malaysia or the city-state of Singapore, its 2002 per capita GDP of US\$6,900 is higher than other countries in Southeast Asia (<http://unstats.un.org/unsd/demographic/social/inc-eco.htm>) (Table 1). Poverty tends to be more widespread in the countryside as industrial development has concentrated mostly in the Bangkok Metropolitan Region.

The only Southeast Asian country never colonized by a European power, Thailand is a constitutional monarchy with executive power vested in a prime minister and cabinet. Under the terms of the 1997 constitution, the bicameral National Assembly consists of a House of Representatives composed of 400 single-member districts and 100 party-list seats, all directly elected for 4-year terms, and a Senate that has 200 members directly elected for 6-year terms.

The current prime minister, Thaksin Shinawatra, is a telecommunications billionaire who has formed a coalition government from the 248 seats won by his Thai Rak Thai (Thais Love Thais, or TRT) in the 2001 general election together with three smaller

parties: two conservative and one centrist. With this comfortable parliamentary majority, Thaksin's government has greater stability than any government in modern Thai history, which between 1932 and 1992 gave rise to 50 different, and often unstable, military and civilian administrations.

There are 76 provinces in Thailand, including metropolitan Bangkok. The governor of Bangkok has been directly elected, whereas all other provincial governors have been career civil servants appointed by the Ministry of the Interior. Under this highly centralized system of public administration, most officials at provincial, district, and subdistrict level have been affiliated with central agencies. Consequently, public policy has traditionally been paternalistically developed within the centrally oriented bureaucracy, while political power has revolved around a narrow set of competing elites in the capital.

1997 CONSTITUTION

Thailand's 1997 constitution is the basis for major structural changes in the Thai political, legal, and administrative systems, although when it was approved overwhelmingly by the National Assembly the vote belied deep divisions between reformers and conservatives. It does, nevertheless, enshrine the principle of constitutional supremacy. Any law, act, or decree contrary to or inconsistent with the constitution is unenforceable, while all branches of the government are bound to ensure that the constitution's integrity is upheld.

A number of constitutional features are designed to undermine corrupt electoral practices and to reduce the fragmentation of political power among too many political parties. All candidates for election either as members of the House of Representatives or the Senate must have at least a Bachelor's degree. This measure is designed to reduce vote buying in rural areas, based on the assumption that many traditional "godfathers" (i.e., wealthy local businessmen who dominate political/economic life within a rural district) would be disqualified. All parties contesting a general election must endorse a "slate" of 100 candidates so as to help eliminate small splinter factions and, by consolidating

Table 1 Per capita GDP (US\$) 2002 (estimated), selected countries in Southeast Asia

Country	Per capita GDP (\$US)
Singapore	24,000
Malaysia	9,300
Thailand	6,900
Philippines	4,200
Indonesia	3,100
Vietnam	2,250
Myanmar (Burma)	1,660

http://www.worldfactsandfigures.com/gdp_country_desc.php.

power in the hands of a few dominant parties, help ensure governments with a degree of long-term stability.

ASIAN FINANCIAL CRISIS

The spark igniting the Asian financial crisis in July 1997 was the Thai government's failed attempt to defend the Thai baht, which was linked to the U.S. dollar, against speculators. Thailand's booming economy ground to a halt amid massive layoffs in finance, real estate, and construction that resulted in huge numbers of workers returning to their villages in the countryside and 600,000 foreign workers being sent back to their home countries. The currency was floated and the government agreed to a US\$17.2 billion bailout by the International Monetary Fund (IMF)/World Bank subject to conditionalities such as passing laws relating to bankruptcy (reorganizing and restructuring) procedures and establishing strong regulatory frameworks for banks and other financial institutions.

Thailand's economy is once again doing well and the extra tax revenue being generated should allow Thailand to balance its budget in 2004, four years ahead of schedule. Real GDP growth is expected to average 5.9% in 2003–2004, while inflationary pressures are expected to remain low and the baht to continue to be stable (<http://www.economist.com/countries/Thailand/profile.cfm?folder=Profile-Forecast>).

PUBLIC SECTOR

The Thai public sector is not particularly large. Except for Myanmar, Thailand spends a smaller portion of its GNP on the public sector than its Asian neighbors (Table 2).

Over the years, legislation has been enacted to facilitate the development of a modern civil service.

Table 2 Total expenditure on the public sector, selected countries in Southeast Asia (2000)

Rank order	Country	Expenditure (% of GDP)
1	Vietnam	23.4
2	Indonesia	20.5
3	Malaysia	19.7
4	Philippines	19.5
5	Singapore	19.1
6	Thailand	18.0
7	Myanmar	8.7

<http://www.worldbank.org/data/databytopic/publicsector.html>.

The 1975 Civil Service Act introduced a position classification system and required new recruits to take an entry examination, and the 1992 Civil Service Act outlined the role of the Civil Service Commission as the central human resources agency for the public service. However, the bureaucracy is a traditional pillar of Thai society that has exercised considerable power and enjoyed high status under both military and civilian regimes. Therefore, reform initiatives have had a limited impact on rectifying a host of shortcomings such as overcentralization, poor coordination, functional overlap (particularly at the subnational level), antiquated work processes, too much delay and discretion in implementing decisions, and lack of responsiveness and accountability to the citizenry. During the 1990s, the public sector's attempt to recruit and retain professionals was undermined by more attractive private sector salaries, particularly in key areas such as law, accounting, and computer science.

The military is also a traditional pillar of Thai society, although its role since 1992 has changed from one of exercising absolute power to that of "system stabilizer." As a more professional military has emerged, it has been less inclined to become involved directly in political affairs. Even the severe economic meltdown of the Asian financial crisis and the strong opposition within the traditional bureaucratic/military establishment to the new constitution did not provoke military intervention.

Public sector reforms designed to reduce the role of the state predate the initiatives flowing from the new constitution, although privatization of state assets was part of a package of reform measure announced when Thailand accepted its financial bailout in 1997. The long-term trend of selling off state-owned enterprises (SOEs), developing joint public/private ventures, and deregulation to encourage market competition has been boosted by the Thaksin administration's determination to move ahead with state enterprise privatization, with the caveat that enterprises responsible for key public services are to remain at least 50% owned by the government. Public

utilities, such as electricity and water, are to be under government control even after corporatization.

The 1997 constitution is designed to reduce the control of the central bureaucracy through initiatives such as decentralization and to strengthen good governance principles. The Official Information Act (1997) provides greater access to official information. One of the most visible changes in transparency is the policy on asset declaration by government officials—an issue highlighted frequently by the print media. In May 1999, the Cabinet issued its Public Sector Management Reform Plan, outlining the government's vision for institutional change under the supervision of a high-level Public Sector Reform Committee. Greater decentralization of the public sector has been a controversial issue of long-standing between politicians and bureaucrats, and it is now a key component of the democratic reform vision. A Decentralization Committee has been established, and responsibilities and funding are beginning to move away from the center of government toward special or general organizations at the local level. This is particularly the case with health, education, and local government financing.

A new bureaucratic system was launched in October 2002, consisting of 20 ministries, 61 bureaus, and 60 departments. This reform was designed to enhance government efficiency and flexibility, and reduce the chain of command. Primary objectives are to improve interdepartmental coordination and to bring together similar functions from multiple ministries.

Because of his deep distrust of, and increasing impatience with, the Thai bureaucracy's measured procedures, the Prime Minister, Thaksin Shinawatra, has now thrown his support behind the development of a chief executive officer (CEO) system, which would require senior government officials to manage their areas of responsibility like corporate CEOs. Among the first group targeted for training in this approach have been ministry permanent secretaries, department heads, chiefs of state enterprises, and some provincial governors (<http://www.kellogg.northwestern.edu/news/hits/030702tnt.htm>).

Corruption is a long-standing governance issue in Thailand, where traditional patron–client relationships and gift-giving have been widely accepted in public appointments, dealing with public officials and political campaigning. It has been the norm for government ministers to exercise patronage in relation to promotions and appointments, in particular to semigovernment Boards and authorities. Transparency International's Corruption Perceptions Index 2003 ranks Thailand 75th out of the 133 countries whose levels of corruption were charted (<http://www.transparency.org/cpi/2003/cpi2003.en.html>) (Table 3).

A major concern about the public reform decentralization initiatives concerns the capacity of the watchdog National Counter Corruption Commission (NCCC) to operate effectively against widespread corruption at the subnational level. However, unlike the corruption agency it replaces, the NCCC is an autonomous agency established apart from the Office of the Prime Minister with its own secretariat, budget, and personnel administration. Its main responsibilities are to ensure the declaration and inspection of assets and liabilities for officials holding political positions and state officials, to undertake corruption prevention and suppression activities (<http://www.nccc.thaigov.net/nccc/eng.php>). It has the power to overrule the Attorney General and initiate prosecution.

CONSTITUTIONAL COURT

The Constitutional Court is an independent body consisting of 15 full-time judges selected by a complicated process that seeks to ensure they are above political and business interests. It renders judgments on the constitutionality of the provisions of law and other powers as provided for in the Constitution and other laws.

In 2000 the NCCC found Thaksin Shinawatra, Thailand's prime minister, guilty of filing false asset statements when he was deputy prime minister in 1997 and the case was referred to the Constitutional Court for further action. A controversial eight-to-seven

Table 3 Transparency international corruption perceptions index 2003, selected countries in Southeast Asia

Country rank	Country	CPI 2003 score	Surveys used	Standard deviation	High–low range
5	Singapore	9.4	12	0.1	9.2–9.5
37	Malaysia	5.2	13	1.1	3.6–8.0
70	Thailand	3.3	13	0.9	1.4–4.4
92	Philippines	2.5	12	0.5	1.6–3.6
100	Vietnam	2.4	8	0.8	1.4–3.6
122	Indonesia	1.9	13	0.5	0.7–2.9
129	Myanmar	1.6	3	0.3	1.4–2.0

<http://www.transparency.org/cpi/2003/cpi2003.en.html>.

split in Thaksin's favor, in August 2001, averted a political crisis by allowing the prime minister to remain in office. It did, however, raise serious questions about whether the rule of law was being undermined by political expediency.

OMBUDSMAN

The Ombudsman of Thailand (<http://www.ombudsman.go.th>) established in April 2000, under the provisions of the 1997 constitution, is empowered to refer any case in violation of the constitution to either the Administrative Court or the Constitutional Court. It also has the authority to inquire into complaints about maladministration by civil servants, members or employees of government bodies, state agencies, state enterprises, or local governments.

CIVIL SOCIETY

Civil society has, during the past 20 years, expanded substantially. Participation in Thai civil society appears to be strongest among rural, older citizens with more traditional attitudes although there is a continuing growth in the numbers of non-governmental organizations (NGOs), with and without legal status, supporting issues such as environmental and wildlife protection, women's/children's rights, and student's concerns. Private foundations, business groups, and professional organizations have also developed as vehicles of expression for the concerns of the middle class.

Many civil society groups and organizations participated in the consultation exercises that led to the drafting of Thailand's 1997 Constitution, which recognizes public participation as having a legitimate role to play in shaping national policies (Chapter V, Section 76—http://www.kpi.ac.th/en/con_th5.asp). Under the new constitution, the Senate, composed of 200 members elected from geographic constituencies, is designed to reduce the influence of the bureaucracy, military, and business groups in favor of individuals more closely aligned with the community.

Thai civil society appears to be increasingly integrated into the participatory processes developed and controlled by the state in line with its "good governance" agenda rather than operating as a wholly independent critical force. For example, the government recently established the Community Organizations Development Institute (CODI) as a new type of public organization under the supervision of the Ministry of Finance (<http://www.codi.or.th>) to promote the development of community organizations and civil society by coordinating the efforts of stakeholders involved in community development.

NATIONAL ECONOMIC AND SOCIAL DEVELOPMENT BOARD

The National Economic and Social Development Board (NESDB) (<http://www.nesdb.go.th>) is Thailand's economic development agency and "think tank." It is responsible for developing the 5-year National Plans that provide a guideline for the country's medium-term economic development. A series of participatory planning exercises with the private sector, NGOs, and civil society at the regional and local levels was undertaken to develop the 9th National Plan (2002–2006) (http://www.un.or.th/Thailand_Info/Development/plan/plan.html). Its main goals of poverty alleviation, recovery with sustainability and stability, good governance, and strengthening development foundations are essentially Thailand's national priorities. In the latest plan the growing significance of civil society in the decision-making process is emphasized.

MEDIA

The print and non-print media have been credited with the role of assisting to heighten political consciousness as for example, during the televising of the constitutional debates prior to the approval of the 1997 constitution. However, Freedom House's Global Survey of Media Independence reduced Thailand's 2002 rating from Free to Partly Free because of increased pressure on media outlets by Prime Minister Thaksin Shinawatra's administration (<http://www.freedomhouse.org/research/pressurvey.htm>) (Table 4).

Table 4 Freedom house freedom of the press survey 2003,^a selected countries in Southeast Asia

Country	Status
	Free (0–30)
	Partly free (31–60)
	Not free (61–100)
Philippines	Free/21 to 30
Thailand	Partly free/31 to 40
Indonesia	Partly free/51 to 60
Singapore	Not free/61 to 70
Malaysia	Not free/71 to 80
Vietnam	Not free/81 to 90
Myanmar	No data

^aLegal environment: 0–30 points

Political influences: 0–40 points

Economic pressures: 0–30 points

Total score: 0–100 points

<http://www.freedomhouse.org/research/pressurvey.htm>.

Most radio and broadcast television stations are directly or indirectly owned or overseen by either the government or the armed forces, and, by law, radio stations must renew their licenses annually. As for newspapers, which do scrutinize official policies and report allegations of corruption and human rights abuses, they appear to be exercising a higher level of self-censorship.

CONCLUSION

The electoral system and framework of agencies established by Thailand's enlightened 1997 constitution provide a sound basis for the continuing development of good governance principles in the country. However, critical issues surround how to ensure that either the majority government or the many different "vested interests" in the public administration system do not compromise, either overtly or covertly, the potential effectiveness of the new watchdog agencies, progress toward public services reform, freedom of the media, or the active engagement of civil society in policy making processes.

BIBLIOGRAPHY

1. Asian Development Bank. 2003. <http://www.adb.org/thailand>.
2. Asia Foundation. 2003. <http://www.asiafoundation.org/Locations/thailand.html>.
3. Bello, W.; Cunningham, S.; Li, K.P. *A Siamese Tragedy: Development and Disintegration in Modern Thailand*; Zed Books Ltd.: London, 1998.
4. Bunbongkarn, S. Thailand: democracy under siege. In *Driven By Growth: Political Change in the Asia-Pacific Region (Revised Edition)*; Morley, J.W., Ed.; M. E. Sharpe: Armonk, NY, 1999; 161–175.
5. CountryWatch. *Thailand: 2003 Country Review*. Country Watch: Houston; 2002. <http://www.countrywatch.com>.
6. Dixon, C., Drakakis-Smith, D., Eds.; *Uneven Development in South East Asia*; Ashgate: Aldershot, 1997.
7. Klein, J.R. *The Constitution of the Kingdom of Thailand, 1997: A Blueprint for Participatory Democracy*; The Asia Foundation: San Francisco, CA, 1998. (Working Paper #8).
8. Office of the Council of State, Royal Thai Government. *Constitution of the Kingdom of Thailand*; 1997.

Theory, Types and Attributes of

Jeffrey A. Weber

Department of Political Science, East Stroudsburg University, East Stroudsburg, Pennsylvania, U.S.A.

INTRODUCTION

Theory has come to mean different things to different people.^[1] Kerlinger,^[2] for example, views theory as a “set of interrelated constructs, definitions, and propositions that present a systematic view of phenomena by specifying relations among variables, with the purpose of explaining and predicting the phenomena.” While Kaplan^[3] viewed theory as an “explanation of something” through its “relationship with other elements” in such a way that it “formed a unified whole.” Weick^[4] offers another view of theory as an approximation of a phenomenon that begins as a “non-linear vision” and is converted into related constructs and propositions. He contends that a theory is never complete but always a work in progress.^[1]

Despite these different meanings, there does appear to be some commonality in the overall understanding of what is a theory. Therefore, based on some of the commonalities, a theory is defined as: An explanation of a phenomenon or phenomena through a set of inter-related constructs.

TYPES OF THEORY

Just as theory can have different meanings, so there are different typologies for theories.^[3,5] Some categorize theories by analytical process—deductive, inductive, pattern, or grounded.^[1,6] Others categorize theories by scope and level of generalizability.^[1,6]

Drawing from these works, theories can be grouped into four types: grand, general, middle-range, and substantive.

Grand Theory

Grand theory is a system of interrelated abstract concepts and propositions that comprehensively describe and explain large categories of phenomena non-probabilistically. A grand theory lays out proposed universal laws that guide the interrelationship of the concepts comprising the phenomena.^[1,6]

There are many examples of grand theories in both the natural and social sciences. Etzioni^[7] developed a

grand theory concerning the “societal and political processes.” Einstein^[8] and Hawking^[9] developed grand theories that explained matter, space–time, and energy. Parson’s^[10] “structure of social action” was a grand theory explaining a societal system. While, most recently, Wilson^[11] has presented a grand theory involving the “unity of knowledge.”

A grand theory involves a possible explanation of phenomena. While this explanation is broad and generalizable, in order to be useful, the grand theory must be complete, meaning addressing the whole of the phenomena. Grand theories serve the body of knowledge by providing an overall guide to the study of the individual phenomenon that comprise the phenomena.

General Theory

The term, “general theory” has previously been used (i.e., general theory of relativity), but, while the theory-building literature addresses what a theory is in the broad sense, as well as different types of theory and multiple definitions of theory, a specific definition for general theory has yet to be offered.

A general theory is a form of grand theory, except instead of dealing with phenomena, a general theory focuses on a phenomenon. Synthesizing from previous use of the term general theory, the following definition was developed:

- A coherent narrative composed of assumptions, abstract reasoning, and speculation that describes, defines, and explains the whole of an observed or experienced phenomenon’s main constructs, their interrelationships, and their boundaries in a broad and generalizable scope that is applicable to the whole of the phenomenon regardless of academic discipline.^[4,12,13]
- General theory would be a type of theory that lies in between grand theory and middle-range theory. While grand theory deals with interrelationships of phenomena, a general theory deals with a specific phenomenon. It is different from middle-range theory, in that it covers the whole of the phenomenon and it is not discipline specific.

Middle-Range Theory

Middle-range theory is “more limited in scope than a grand theory” or a general theory.^[6] It is derived from a general theory by a particular academic discipline that is studying a single concept, which is a part of the phenomenon being explained by the general theory.

Middle-range theories are sometimes referred to as formal theories or theoretical models.^[5] Formal theories and theoretical models are composed of boundaries, constructs, and propositions. These form a detailed explanation of a concept, which can be tested and verified.

The boundaries of middle-range theories are more defined and specified than general theories. Middle-range theory boundaries are within the general theory boundaries but narrow to focus on a single concept and its constructs. A construct is defined as a “mental configuration,” which “though not observational either directly or indirectly, may be applied or even defined on the basis of observables.”^[12] The constructs of middle-range theory flow directly from the concepts that comprise the grand theory.

Furthermore, the interrelationships between constructs in middle-range theories are defined through propositions. Quantitative research methods are easily applied because the unit of analysis is limited in time and space.

Examples of middle-range theories are Trist and Emery's^[14] sociotechnical theory explaining the concept of effective organizations; also, Downs's^[15] bureaucratic life-cycle model and Lindblom's^[16] incrementalism. The verification of a middle-range theory involves breaking it down further into a substantive theory so that it may be tested.

Substantive Theory

Substantive theory is “restricted to features of populations, settings, and times that can be concretely identified.”^[6] A substantive theory is based on a middle-range theory, and it seeks to describe, in great detail, the interrelationships among the middle-range theory's constructs. Substantive theories represent the operationalizing of a grand theory.

Variables comprise the primary building block's of substantive theory. Variables are “an observable entity” that “may be viewed as an operational configuration derived from a construct.”^[12] Furthermore, variables consist of indicators that are specific measurable components.

Hypotheses define the relationships between the variables. Hypotheses are refinements of proposition statements in the middle-range theory but are

“more concrete and operational statements of these relationships.”^[12]

Additionally, narrow “typologies or categorical systems” can be defined as substantive theories. Typically, if one is building theory, from data, such as grounded theory, the data must first be categorized, and this marks the beginning of a substantive theory.

ATTRIBUTES OF THEORY

Basically, there are two schools of thought concerning the attributes of theory. One school of thought contends that there is a single set of rigid attributes that must be met in order to be classified as a theory. The opposing school of thought argues that a theory is ever changing and developing, therefore, it can have a variety of attributes.

Single Set of Qualities

Sutton and Staw^[17] have argued that theory is much maligned and misused, and this situation has resulted in practically any speculation being called a theory. Furthermore, Sutton and Staw argue that lack of consensus on what is a theory has led to disorganization within the body of knowledge. Therefore, theory must be defined clearly because just as “a collection of words does not comprise a sentence,” a collection of thoughts, speculations, assumptions, and data does not make a theory.^[12]

Therefore, what qualifies as a theory is very limited in scope and must meet stringently defined characteristics. Theories must contain boundaries, constructs, propositions explaining the relationship between constructs, variables that comprise the constructs, and hypotheses that explain the relationship among the variables. Furthermore, a theory must have “logical adequacy,” be non-tautological, and have empirical adequacy.^[12,18] Finally, a theory must have utility and provide a mechanism for “prediction beyond chance.”^[12]

While specifically defining what is a theory prevents every metaphor or speculation from being termed a theory, there is a risk of having too stringent a definition, which then limits one to only substantive theory. Thus, details are explained without ever explaining the overall phenomenon that connects them.

Variety of Qualities

Seeming to contradict the above viewpoint is the idea that we are afraid to use the word, “theory.” Weick^[4] contends that academics will go to any extreme

possible to state that they are not putting forth a theory because it does not fit some particular definition of “good theory.” Furthermore, there is the fear of merciless criticisms that the theory is not deep enough in its explanation or that the theory is not broad enough to be generalizable.

DiMaggio^[19] proposes a solution to the above dilemma by pointing out that “theory is a continuum.” He states that theories are approximations of phenomenon that begin as “non-linear visions,” and are converted into detailed constructs and propositions. As one moves a theory from a vision to detailed constructs and propositions, the theory loses accuracy and precision and becomes more of an approximation, but at the same time, it becomes increasingly concrete. Thus, it is “difficult to judge” what is a theory, because “theory work can take a variety of forms,” and it is a work in progress.^[4,20] Therefore, to recognize what is a theory requires considering the process by which the theory is being developed.

The process of theorizing consists of the following activities: abstracting, generalizing, relating, selecting, explaining, synthesizing, and idealizing.^[4] To know if one is developing a theory, one has to know the “context of what came before and what comes next.”^[4] If a researcher, seeking to explain a phenomenon, moves from one part of the process to another part, then theory building is taking place. “Few things are full-fledged theories” but, there are many developing theories, which are moving through the process.^[4]

Common Attributes of Theory

Regardless of the school of thought followed in trying to understand the attributes of theory, there does appear to be some commonalities between the “single set of attributes” and the “variety of attributes” schools of thought. First, a theory does have the attribute of some kind of context or boundary. Consequently, there appear to be three boundaries for a theory: first, the paradigm of the theory builder (paradigm is used in the sense of a overarching perspective that encompasses the values and thought process of the theory builder); second, the amount of time encompassed by the theory; and, third, the amount of space taken by the phenomena or phenomenon the theory seeks to explain.

Second, within those boundaries, a theory is composed of constructs, which are abstract conceptualizations, which serve as a part of the explanation within the theory. As defined earlier, a theory is an explanation of a phenomenon or phenomena. The explanation is composed of parts, and it is these parts that are defined as constructs.

Third, there are interrelationships among the constructs that, when combined with the boundaries and the constructs, serve to complete the explanation of the phenomenon or phenomena. These interrelationships can take numerous forms and many different words are used to express them, such as hypotheses or associations. What the interrelationships are ultimately

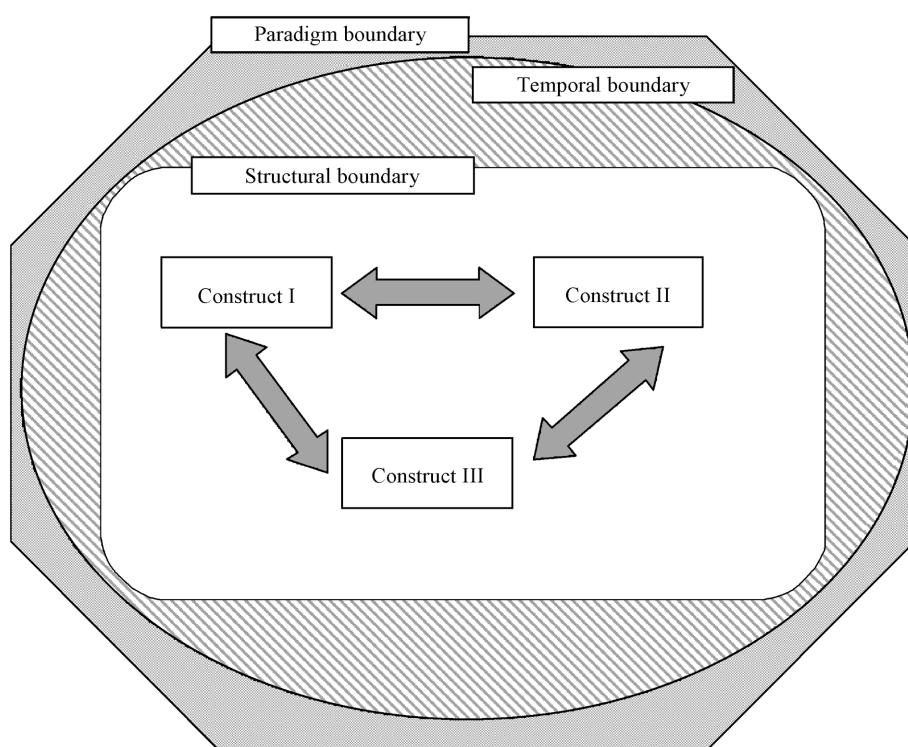


Fig. 1 Components of a theory.

called is based on the values and perspective of the theorist(s).

Fig. 1 depicts the attributes of a theory. The boundaries (paradigm, time, and space) are shown as an oval. Three constructs are depicted within the boundaries, along with their interrelationships.

CONCLUSION

The division of theory into four types is meant as a heuristic device rather than a rigid framework. This fourfold typology allows categorizing of theories in order to facilitate their study. It has been stated that almost all theories easily fall into one of the four types: grand, general, middle-range, and substantive. Furthermore, the typology permits one to understand how theories contribute to the body of knowledge. Regardless of the perspective of the theorist, theories consist of the following attributes: boundaries (paradigm, time, space), constructs, and interrelationships among the constructs.

REFERENCES

1. Merriam, S.B. *Case Study Research in Education: A Qualitative Approach*; Jossey-Bass: San Francisco, CA, 1988; 57 pp.
2. Kerlinger, F.N. *Behavioral Research: A Conceptual Approach*; Holt, Rinehardt, & Winston: New York, NY, 1986; 9 pp.
3. Kaplan, A. *The Conduct of Inquiry*; Chandler: San Francisco, CA, 1964; 335 pp.
4. Weick, K.E. What theory is *not*, theorizing *is*. *Adm. Sci. Q.* **1995**, *40*, 387, 389 pp.
5. Glaser, B.G.; Strauss, A.L. *The Discovery of Grounded Theory: Strategies for Qualitative Research*; Aldine De Gruyter: New York, NY, 1967.
6. Goetz, J.P.; LeCompte, M.D. *Ethnography and Qualitative Design in Educational Research*; Academic Press: New York, NY, 1984.
7. Etzioni, A. *The Active Society: A Theory of Societal and Political Process*; The Free Press: New York, NY, 1968.
8. Einstein, A. *Relativity: The Special and General Theory*, 15th Ed.; Crown Publishers: New York, NY, 1952.
9. Hawking, S. *A Brief History of Time: From the Big Bang to Black Holes*; Bantam Books: New York, NY, 1988.
10. Parsons, T. *The Structure of Social Action*; The Free Press: New York, NY, 1937.
11. Wilson, E.O. *Consilience: The Unity of Knowledge*; Alfred A. Knopf: New York, NY, 1998.
12. Bachrach, S.B. Organizational theories: some criteria for evaluation. *Acad. Manage. Rev.* **1989**, *14* (4), 496–515.
13. Gioia, D.A.; Pitre, E. Multiparadigm perspectives on theory building. *Acad. Manage. Rev.* **1990**, *15* (4), 584–602.
14. Emery, F.E.; Trist, E.L. The causal texture of organizational environments. In *Sociotechnical Systems: A Sourcebook*; Pasmore, W., Sherwood, J.J., Eds.; University Associates, Inc.: LaJolla, CA, 1978; 13–27.
15. Downs, A. *Inside Bureaucracy*; Rand Corporation: Palo Alto, CA, 1967.
16. Lindblom, C.E. The science of muddling through. *Public Adm. Rev.* **1959**, *19*, 298–312.
17. Sutton, R.I.; Staw, B.M. What theory is not. *Adm. Sci. Q.* **1995**, *40*, 371–384.
18. Dubin, R. *Theory Building*, Revised Ed.; The Free Press: New York, NY, 1978.
19. DiMaggio, P.J. Comments on “What theory is not.” *Adm. Sci. Q.* **1995**, *40*, 391–397.
20. Weick, K.E. Theory construction as disciplined imagination. *Acad. Manage. Rev.* **1989**, *14* (4), 516–531.

Third-Party Interventions

R. Wayne Boss

Leeds School of Business, University of Colorado at Boulder, Boulder, Colorado, U.S.A.

Mark L. McConkie

Graduate School of Public Affairs, University of Colorado at Colorado Springs, Colorado Springs, Colorado, U.S.A.

INTRODUCTION

Third-party intervention is a method of resolving conflicts between two or more people when those conflicts involve interpersonal issues that significantly disrupt necessary processes and work relationships, typically with a trained resource. Not all conflict is bad: Creative conflict can enhance motivation and innovation and help group members understand unfamiliar ideas or points of view. Also, not all conflict is interpersonal. When conflict involves substantive issues, such as work methods, pay rates, conditions of employment, and competition for scarce resources, conflict resolution interventions often involve settling labor-management disputes through arbitration and mediation—methods that require considerable training and expertise in law and labor relations. Even when the conflict is interpersonal and counterproductive, it may not be severe or disruptive enough to warrant attention. In some instances, the parties may resolve the conflict themselves, or managers may resolve the conflict unilaterally—for example, when they are under heavy time pressures, when the disputants are not expected to work together in the future, or when resolution of the dispute has a broad effect on the organization.

The third-party intervener must be highly sensitive to his or her own feelings and to those of the others. He or she also must recognize that some tension and conflict are inevitable and that although there can be an optimum amount and degree of conflict, too much conflict can be dysfunctional for both the people involved and the larger organization. The third-party consultant must be sensitive to the situation and able to use a number of different intervention strategies and tactics when intervention appears to be useful. Finally, he or she must have professional expertise in third-party intervention and must be seen by the parties as neutral or unbiased regarding the issues and outcomes of the conflict resolution.^[1]

THE ROLE OF CONFRONTATION

Central to successful third-party interventions is the concept of confrontation. People must be willing to both accept the fact that conflict exists between them and to confront that conflict in constructive ways. Walton describes the elements for successful confrontation:^[2]

1. Mutual positive motivation: both parties want to resolve the conflict
2. Balance in the situational power of the two principles (the parties hold approximately equal power positions)
3. Synchronization of their confrontation efforts: the parties are ready to resolve the problem and agree on a process for doing so
4. Appropriate pacing of the differentiation: clarifying their differences and integration (identifying commonalities and mutual goals) phases of a dialog ensure that they allow adequate time to work through and clarify their negative feelings
5. Conditions favoring openness in dialog: agreement beforehand about honesty and openness
6. Reliable communicative signs: making certain that people can understand one another
7. Optimum tension in the situation (both parties should be experiencing stress in the relationship, since no meaningful change can take place without it).

THIRD-PARTY CONSULTANTS

When the conflict is interpersonal in nature, however, and when it destroys trust and prevents people from working together, third-party approaches may help control and resolve it. Interpersonal issues include personality differences, task orientation, communication problems, and misperceptions.

Third-party consultants help the parties interact with each other directly, facilitate the diagnosis of their conflict, and help identify ways to resolve it. Cummings and Worley summarize the characteristics of effective third-party consultant in the following way:

The third-party consultant must develop considerable skill at diagnosis, intervention, and follow-up.

A MODEL FOR THIRD-PARTY FACILITATION

A basic and typical model for third-party facilitation includes perception sharing, problem identification, contracting, and follow-up meetings.^[3]

Perception Sharing

The parties are asked to take a few minutes and respond to such questions as these:

1. What does the other person do well?
2. What do I think I do that bothers the other person?
3. What does the other person do that bothers me?

Both parties then take turns explaining their responses to the facilitator and the other person.

This design has several benefits, as an outline reflects:

- The process of writing down specific behaviors that each person likes and dislikes forces an element of rationality into an otherwise emotionally charged interchange; subjective, evaluative statements and gross generalizations are no longer acceptable.
- Just taking the time to write things down helps diffuse the hostility and allow each party to cool off. Since the design guarantees that both will be able to tell their side of the story without interruption, they are more inclined to listen.
- In listing their observations, conflicting parties realize that the things that actually bother them about each other are not nearly as unchangeable as they had imagined.
- The design allows both parties to concentrate on the positive aspects of their relationship, which often surprises them. Those positive comments help mitigate the hostility and establish a climate of cooperation.
- The second question forces them to examine their own behavior and to specify what they did that caused problems for the other person—something neither party is typically willing to do.

Problem Identification

The responses to question 3 are then listed, and the parties assign vectors of different lengths to indicate the seriousness of each: the longer the vector, the more serious the problem. In most cases, the parties are surprised with what they see, since people often have no idea what they are doing to cause problems for the other person. Usually they find that some behaviors that cause serious problems can be corrected immediately, while others take longer.

Contracting

The purpose of contracting is to ensure that specific steps will be put in place to resolve the problems. Specifically, for each problem, the parties are asked to respond to two questions:

1. What will I do to resolve this problem?
2. What will the other person do to help me succeed?

This approach gives both parties a vested interest in resolving the problems and helps promote a spirit of cooperation. Both parties define the specific behaviors they want the other to change and then negotiate the things that they are willing to change. The consultant facilitates the process and documents the decisions.

Follow-Up Meetings

Since meaningful change is impossible without mechanisms for reinforcement, the parties schedule follow-up meetings to monitor their progress, to hold one another accountable, and to address issues that may subsequently arise. These meetings usually last no more than a couple of hours and, depending on the level of progress, the third-party consultant may or may not attend.

Results of the Intervention

When properly conducted, this intervention can have a long-term, successful impact on dysfunctional relationships. For example, Boss, McConkie, and Golembiewski^[4] report the results of a third-party intervention between the director of an operating room and the chief surgeon in a major medical center. Positive outcomes included significant improvements in their levels of interpersonal trust and cooperation with one another and in 10 key behaviors that impacted their relationships. Three hospital administrators, who also evaluated their effectiveness, reported similar results. Further, these positive results were maintained for 14 years, as indicated by reports from both the participants and observers. Additional positive outcomes in the operating room included a resolution to the orthopedic prostheses problem, improved scheduling of surgical procedures, a decrease in verbal abuse by physicians, the reconvening of the operating room standards committee, and a significant reduction in turnover of operating room nurses.

WORDS OF CAUTION

Third-party interventions in conflict situations have the potential not only to control the conflict but to also resolve it, to the point that conflicting parties can develop long-term, positive relationships with one

another. However, the “intervention into interpersonal conflict situations requires a highly skilled professional or a highly skilled layperson who understands the dynamics of conflict. This intervention should not be undertaken by a novice in human and social processes of organizations.”^[5]

REFERENCES

1. Cummings, T.G.; Worley, C.G. *Organization Development and Change*, 7th Ed.; South-Western College Publishing: Cincinnati, OH, 2001; 229 pp.
2. Walton, R.E. *Interpersonal Peacemaking: Confrontations and Third Party Consultation*; Addison-Wesley: Reading, MA, 1969; 94 pp.
3. Fordyce, J.K.; Weil, R. *Managing With People: A Managers Handbook of Organization Development Methods*; Addison-Wesley: Reading, MA, 1971.
4. Boss, R.W.; McConkie, M.L.; Golembiewski, R.T. *The Durability of OD Interventions: The 14-Year-Impact of a Third-Party Intervention*, Paper Presented at the Annual Meeting of the Academy of Management, Aug. 5–7, 2002.
5. French, W.L.; Bell, C.H., Jr. *Organization Development*, 5th Ed.; Prentice-Hall: Englewood Cliffs, NJ, 1995; 203 pp.

Time Value of Money

John D. Wong

*Hugo Wall School of Urban and Public Affairs, Wichita State University,
Wichita, Kansas, U.S.A.*

INTRODUCTION

According to economics, the value of a commodity is directly related to its scarcity relative to demand. The concept of the *time value of money* is based on the fact that most people would prefer to receive a given sum of money today than the same amount at some time in the future. If nothing else, receiving the money now potentially affords additional opportunities that may not be available later. Also, money held today is worth more than a promise of payment in the future because of the risk that the payment may not be made in the future and because cash held today could be invested to earn a return in the future. Time value analysis has many applications, ranging from setting up schedules for paying off debt to decisions about whether to acquire new equipment. When an organization acquires a capital asset, it must either borrow the funds needed for the acquisition or use resources currently held by the organization. In the first case, the organization would be paying interest over the life of the loan. In the second case, the organization would be foregoing the return that it could have earned from using those resources in their next best alternative use.

CONCEPT OF INTEREST

The time value of money derives from the concept of interest. *Interest* is the return earned by or the amount paid to another party that has foregone other opportunities and chooses instead to make money available to the other party. In other words, interest is the fee assessed for the privilege of borrowing the money of another or the fee obtained for loaning money to another. The *principal* is the amount of money borrowed, loaned, or invested. The *term* is the length of time or number of time periods during which the borrower can use the principal, i.e., the duration of the loan. The *rate of interest* is the percentage of the principal that the borrower must pay the lender for each time period as compensation for foregoing other opportunities. The *nominal interest rate* is the periodic rate of interest that is stated in the loan agreement

or security. The *market interest rate* is comprised of three components:

- The *real rate of interest* compensates the lender for not being able to use the money during the term of the loan.
- An *inflation premium* compensates the lender for the possibility that inflation may erode the purchasing power of the money during the term of the loan.
- A *risk premium* compensates the lender for the possibility that the loan may not be repaid in full.

The real rate of interest and an inflation premium are sometimes referred to collectively as the *nominal risk-free rate*. In the United States, the nominal risk-free rate can be approximated by the rate on 13-week U.S. Treasury bills, because they are generally considered to be risk free.

An important concept involving the time value of money is that a single lump sum of money or a stream of equal, evenly spaced payments or receipts promised in the future can be converted to an equivalent value today. Conversely, a single lump sum or a series of future payments can be converted to its equivalent value in the future. To take into account the time value of money, the present stock of money may be converted to a future flow of money, or the future flow of money may be converted to a present stock of money.

The process of converting the present stock of money into a future flow of money focuses on the concept of *compounding*, while the process of converting the future flow of money into a present stock of money focuses on the concept of *discounting*. Computationally, both methods will produce equivalent results for decision-making purposes. However, there are conceptual and pedagogical advantages to discounting future flows to present value.

However, it should be kept in mind that the concepts of compounding and discounting are not premised specifically on adjusting for the effect of inflation or uncertainty, but rather focus on the recognition that a given amount of money possessed today is more valuable than that same amount to be received in the future. Current value cash flows should be discounted by the nominal interest rate, while real cash flows should be discounted by the real interest rate.

Table 1 Simple interest

Year	Beginning value	Interest	Ending value
1	\$10,000	10% of \$10,000 = \$1000	+\$10,000 = \$11,000
2	\$10,000	10% of \$10,000 = \$1000	+\$11,000 = \$12,000
3	\$10,000	10% of \$10,000 = \$1000	+\$12,000 = \$13,000
4	\$10,000	10% of \$10,000 = \$1000	+\$13,000 = \$14,000
5	\$10,000	10% of \$10,000 = \$1000	+\$14,000 = \$15,000

Simple Interest

Simple interest is paid or received on the principal sum only. Accumulated interest from prior periods is not used in calculations for the following periods. The amount of simple interest (I) is equal to the product of the principal or present value (PV_0) times the interest rate per time period (i) times the number of time periods (n). Although the examples presented here are denominated in dollars (USD, \$), using alternative currency units [e.g., British Pounds Sterling (GBP, £), European Union Euros (EUR, €), or Japanese Yen (JPY, ¥)] will not alter the result.

Example: $PV_0 = \$10,000$; $i = 10\%$; $n = 5$ years

$$\begin{aligned}
 I &= PV_0 \times i \times n \\
 &= \$10,000 \times 0.10 \times 5 \\
 &= \$5000
 \end{aligned}$$

In this example, \$10,000 invested or borrowed at a 10% simple interest rate will result in \$1000 of interest at the end of one year. After five years, the investment or loan will result in a total of \$5000 in interest plus the original principal amount (see Table 1).

Compound Interest

Compound interest is paid or received not only on the principal but also on any interest earned and not withdrawn during earlier periods. Compound interest may be thought of as a series of single-period, simple-interest contracts. The interest earned in each period is added to the principal of the previous period

to become the principal for the next period. Because interest is paid on interest as it accumulates, compound interest will result in more interest than simple interest.

In the example shown in Table 2, \$10,000 invested or borrowed at a 10% compound interest rate will result in \$1000 of interest at the end of one year. For the second through fifth years, the compound interest computations are based on the original principal plus the interest accrued during the following period. After 5 years, the investment or loan will result in a total of \$6105 in interest plus the original principal amount.

Rule of 72

The *rule of 72* may be used to approximate the time period (n) needed for money to double in value or the interest rate (i) needed for money to double in a given number of years. According to the rule of 72, the number 72 may be divided by the interest rate to determine the approximate number of years it would take for a sum of money to double in value. For example, because $n = 72/i$, it would take 9 years for a sum of money to double in value at an interest rate of 8%. Likewise, the number 72 may be divided by the number of years to determine what interest rate would be necessary to double the value of a given sum of money, e.g., because $i = 72/n$, an interest rate of 8% would be required for a sum of money to double in value after 9 years. The closer the rate of interest is to 8%, the more accurate is the estimate given by the rule. However, the accuracy of the rule diminishes progressively as the interest rate rises over 8%. For interest rates of 20% or higher, an analogous rule of 78 is sometimes used for approximation.

Table 2 Compound interest

Year	Beginning value	Interest	Ending value
1	\$10,000	10% of \$10,000 = \$1000	+\$10,000 = \$11,000
2	\$11,000	10% of \$11,000 = \$1100	+\$11,000 = \$12,100
3	\$12,100	10% of \$12,100 = \$1210	+\$12,100 = \$13,310
4	\$13,310	10% of \$13,310 = \$1331	+\$13,310 = \$14,641
5	\$14,641	10% of \$14,641 = \$1464	+\$14,641 = \$16,105

A more precise result may be obtained by using natural logarithms. To find the number of periods (n) necessary for a given present value (PV_0) to increase to a certain future value (FV_n) at an interest rate of (i), the following may be used: $n = \ln(FV_n/PV_0)/\ln(1 + i)$. For example, it will take 9 years for \$10,000 to double in value at an interest rate of 8% compounded annually.

$$\begin{aligned} n &= \ln(20,000/10,000)/\ln(1 + 0.08) \\ &= \ln(2)/\ln(1.08) = 9 \text{ years} \end{aligned}$$

FUTURE VALUE

Future value (FV_n) is the amount to which an investment made today (present value) will grow after earning compound interest for a number of periods. The difference between the future value and the present value depends on the number of compounding periods involved and the interest rate. The future value includes the principal plus interest accumulated at the end of the term of the loan. The relationship between future value and present value can be expressed as follows:

$$\begin{aligned} FV_n &= PV_0 + I \\ &= PV_0(1 + i)^n \end{aligned}$$

where:

FV_n = Future value.
 I = Annual interest payment.
 PV_0 = Present value.
 i = Interest rate per period.
 n = Number of periods.

Example: $PV_0 = \$10,000$; $i = 10\%$; $n = 5$ years

$$\begin{aligned} FV_n &= PV_0(1 + i)^n \\ &= \$10,000(1 + 0.10)^5 \\ &= \$10,000(1.10)^5 \\ &= \$10,000(1.6105) \\ &= \$16,105 \end{aligned}$$

In this example, \$10,000 invested or borrowed at a 10% compounded interest rate will be worth \$16,105 at the end of 5 years.

Future value interest factors (FVIFs) are commonly programmed into spreadsheet functions and financial calculators and reproduced in tables. The FVIF is the future value of \$1 at the end of (n) periods at a compound interest rate of (i) percent per period.

Table 3 presents FVIFs for up to 30 periods and interest rates up to 10%. To use the table, simply go down the left-hand column to locate the appropriate number of periods. Then go out along the top row until the appropriate interest rate is located.

Multiple Compoundings per Period

Although the interest may be stated as a yearly rate, the compounding periods can be *annually*, *semiannually*, *quarterly*, or even *continuously*. If a cash flow is compounded more frequently than annually, then intrayear compounding is being used. To adjust for intrayear compounding, the interest rate per compounding period must be found as well as the total number of compounding periods. The interest rate per compounding period is found by dividing the annual rate by the number of times per year the cash flows are compounded. The total number of compounding periods is found by multiplying the number of years by the number of times per year cash flows are compounded. Notice that the formula reflects more frequent compounding (n^*m) at a lower interest rate per period (i/m). As (m) approaches infinity, the term $(1 + i/m)^{m^*n}$ approaches (e^{i^*n}), where (e) is equal to Euler's number (approximately 2.71828), and the future value becomes ($PV_0 * e^{i^*n}$). As such, continuous compounding results in the maximum possible future value at the end of (n) periods for a given rate of interest (i). However, note that although compounding multiple times during a time period will increase the interest yield, compounding continuously will not yield an infinite amount of interest.

$$FV_n = PV_0(1 + i/m)^{m \times n}$$

where:

FV_n = Future value.
 PV_0 = Present value.
 i = Interest rate per period.
 m = Compoundings per period.
 n = Number of periods.

Example: $PV_0 = \$10,000$; $i = 10\%$; $m = 12$ compoundings per period; $n = 5$ years

$$\begin{aligned} FV_n &= PV_0(1 + i/m)^{m \times n} \\ &= \$10,000(1 + 0.10/12)^{12 \times 5} \\ &= \$10,000(1 + 0.008333)^{60} \\ &= \$10,000(1.008333)^{60} \\ &= \$10,000(1.6450) \\ &= \$16,450 \end{aligned}$$

Table 3 Future value interest factors (FVIF) for \$1 at interest rate (*i*) for (*n*) years

Year	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
1	1.0100	1.0200	1.0300	1.0400	1.0500	1.0600	1.0700	1.0800	1.0900	1.1000
2	1.0201	1.0404	1.0609	1.0816	1.1025	1.1236	1.1449	1.1664	1.1881	1.2100
3	1.0303	1.0612	1.0927	1.1249	1.1576	1.1910	1.2250	1.2597	1.2950	1.3310
4	1.0406	1.0824	1.1255	1.1699	1.2155	1.2625	1.3108	1.3605	1.4116	1.4641
5	1.0510	1.1041	1.1593	1.2167	1.2763	1.3382	1.4026	1.4693	1.5386	1.6105
6	1.0615	1.1262	1.1941	1.2653	1.3401	1.4185	1.5007	1.5869	1.6771	1.7716
7	1.0721	1.1487	1.2299	1.3159	1.4071	1.5036	1.6058	1.7138	1.8280	1.9487
8	1.0829	1.1717	1.2668	1.3686	1.4775	1.5938	1.7182	1.8509	1.9926	2.1436
9	1.0937	1.1951	1.3048	1.4233	1.5513	1.6895	1.8385	1.9990	2.1719	2.3579
10	1.1046	1.2190	1.3439	1.4802	1.6289	1.7908	1.9672	2.1589	2.3674	2.5937
11	1.1157	1.2434	1.3842	1.5395	1.7103	1.8983	2.1049	2.3316	2.5804	2.8531
12	1.1268	1.2682	1.4258	1.6010	1.7959	2.0122	2.2522	2.5182	2.8127	3.1384
13	1.1381	1.2936	1.4685	1.6651	1.8856	2.1329	2.4098	2.7196	3.0658	3.4523
14	1.1495	1.3195	1.5126	1.7317	1.9799	2.2609	2.5785	2.9372	3.3417	3.7975
15	1.1610	1.3459	1.5580	1.8009	2.0789	2.3966	2.7590	3.1722	3.6425	4.1772
16	1.1726	1.3728	1.6047	1.8730	2.1829	2.5404	2.9522	3.4259	3.9703	4.5950
17	1.1843	1.4002	1.6528	1.9479	2.2920	2.6928	3.1588	3.7000	4.3276	5.0545
18	1.1961	1.4282	1.7024	2.0258	2.4066	2.8543	3.3799	3.9960	4.7171	5.5599
19	1.2081	1.4568	1.7535	2.1068	2.5270	3.0256	3.6165	4.3157	5.1417	6.1159
20	1.2202	1.4859	1.8061	2.1911	2.6533	3.2071	3.8697	4.6610	5.6044	6.7275
21	1.2324	1.5157	1.8603	2.2788	2.7860	3.3996	4.1406	5.0338	6.1088	7.4002
22	1.2447	1.5460	1.9161	2.3699	2.9253	3.6035	4.4304	5.4365	6.6586	8.1403
23	1.2572	1.5769	1.9736	2.4647	3.0715	3.8197	4.7405	5.8715	7.2579	8.9543
24	1.2697	1.6084	2.0328	2.5633	3.2251	4.0489	5.0724	6.3412	7.9111	9.8497
25	1.2824	1.6406	2.0938	2.6658	3.3864	4.2919	5.4274	6.8485	8.6231	10.8347
26	1.2953	1.6734	2.1566	2.7725	3.5557	4.5494	5.8074	7.3964	9.3992	11.9182
27	1.3082	1.7069	2.2213	2.8834	3.7335	4.8223	6.2139	7.9881	10.2451	13.1100
28	1.3213	1.7410	2.2879	2.9987	3.9201	5.1117	6.6488	8.6271	11.1671	14.4210
29	1.3345	1.7758	2.3566	3.1187	4.1161	5.4184	7.1143	9.3173	12.1722	15.8631
30	1.3478	1.8114	2.4273	3.2434	4.3219	5.7435	7.6123	10.0627	13.2677	17.4494

Example: $PV_0 = \$10,000$; $i = 10\%$; $m = 365$ compoundings per period; $n = 5$ years

$$\begin{aligned}
 FV_n &= PV_0(1 + i/m)^{m \times n} \\
 &= \$10,000(1 + 0.10/365)^{365 \times 5} \\
 &= \$10,000(1 + 0.0002740)^{1825} \\
 &= \$10,000(1.0002740)^{1825} \\
 &= \$10,000(1.64869) \\
 &= \$16,486.90
 \end{aligned}$$

Example: $PV_0 = \$10,000$; $i = 10\%$; $m = \text{continuous compounding}$; $n = 5$ years

$$\begin{aligned}
 FV_n &= PV_0 \times e^{i \times n} \\
 &= \$10,000(e)^{0.10 \times 5} \\
 &= \$10,000(2.71828)^{0.5} \\
 &= \$10,000(1.64871) \\
 &= \$16,487.10
 \end{aligned}$$

As can be seen from [Table 4](#), simple interest increases linearly, while compound interest grows

geometrically. The effect of compounding is modest over short periods of time but becomes quite significant over longer periods of time. As can be seen in [Table 4](#), compound interest grows much more rapidly than simple interest. However, the increased growth from compounding interest monthly rather than annually is significantly less than the difference between simple interest and compounding annually. The increased growth from compounding interest daily rather than monthly is almost negligible.

Effective Interest Rate

The *effective interest rate* (i_e) is the actual rate of interest paid by the borrower or earned by the lender. To make a valid comparison between two interest rates when different compounding periods are used, both nominal interest rates must be converted to their equivalent effective rates. The effective rate will always be higher than the nominal or stated interest rate, when interest is compounded more than once per year. As the number of compounding periods increases, the difference between the nominal and effective rates will also increase. However, the size of the increases diminishes with each additional compounding. In many cases, higher effective annual

Table 4 Future value of a \$1 investment at 10% interest

Year	Simple interest	Compounded annually	Compounded monthly	Compounded daily	Compounded continuously
1	\$1.10	\$1.10	\$1.10	\$1.11	\$1.11
2	\$1.20	\$1.21	\$1.22	\$1.22	\$1.22
3	\$1.30	\$1.33	\$1.35	\$1.35	\$1.35
4	\$1.40	\$1.46	\$1.49	\$1.49	\$1.49
5	\$1.50	\$1.61	\$1.65	\$1.65	\$1.65
6	\$1.60	\$1.77	\$1.82	\$1.82	\$1.82
7	\$1.70	\$1.95	\$2.01	\$2.01	\$2.01
8	\$1.80	\$2.14	\$2.22	\$2.23	\$2.23
9	\$1.90	\$2.36	\$2.45	\$2.46	\$2.46
10	\$2.00	\$2.59	\$2.71	\$2.72	\$2.72
11	\$2.10	\$2.85	\$2.99	\$3.00	\$3.00
12	\$2.20	\$3.14	\$3.30	\$3.32	\$3.32
13	\$2.30	\$3.45	\$3.65	\$3.67	\$3.67
14	\$2.40	\$3.80	\$4.03	\$4.05	\$4.06
15	\$2.50	\$4.18	\$4.45	\$4.48	\$4.48
16	\$2.60	\$4.59	\$4.92	\$4.95	\$4.95
17	\$2.70	\$5.05	\$5.44	\$5.47	\$5.47
18	\$2.80	\$5.56	\$6.00	\$6.05	\$6.05
19	\$2.90	\$6.12	\$6.63	\$6.68	\$6.69
20	\$3.00	\$6.73	\$7.33	\$7.39	\$7.39
21	\$3.10	\$7.40	\$8.10	\$8.16	\$8.17
22	\$3.20	\$8.14	\$8.94	\$9.02	\$9.03
23	\$3.30	\$8.95	\$9.88	\$9.97	\$9.97
24	\$3.40	\$9.85	\$10.91	\$11.02	\$11.02
25	\$3.50	\$10.83	\$12.06	\$12.18	\$12.18
26	\$3.60	\$11.92	\$13.32	\$13.46	\$13.46
27	\$3.70	\$13.11	\$14.71	\$14.87	\$14.88
28	\$3.80	\$14.42	\$16.25	\$16.44	\$16.44
29	\$3.90	\$15.86	\$17.96	\$18.17	\$18.17
30	\$4.00	\$17.45	\$19.84	\$20.08	\$20.09

yield may result from a lower nominal interest rate combined with more frequent compounding than from a higher nominal interest rate with no or less compounding. To convert a nominal rate to an equivalent effective rate: $i_e = [1 + (i/m)]^m - 1$, where i_e = effective interest rate, i = nominal or stated interest rate, and m = number of compounding periods per year. For example, a nominal interest rate of 10% compounded daily would result in an effective interest rate of 10.52%.

$$[i_e = [1 + (i/m)]^m - 1 = [1 + (0.10/365)]^{365} - 1 = 0.1052 = 10.52\%]$$

PRESENT VALUE

Present value is the value today of a future cash flow or cash flow equivalent evaluated at the appropriate interest rate. The process of computing present values is also sometimes referred to as discounting. Discounting is just the opposite of compounding. Because sums of money are worth different amounts at different times, they can be compared only if they have the same time

reference. The most convenient time reference is usually the present. The farther the payment date is in the future, the less it is worth today. The present value of a sum to be received in the future decreases and approaches zero as the payment date is extended further into the future, and also, the rate of decrease is greater the higher the discount rate. At relatively high discount rates, money to be received in the future is worth little today, and even at a relatively low interest rate, the present value of money due in the distant future is still small. As such, the present value increases whenever the interest rate decreases. Also, the present value increases when the number of periods of discounting decreases. In other words, the change in the present value is inversely related to changes in the interest rate or the number of periods.

When computing present values, the relevant interest rate is referred to as the *discount rate*. The present value of a future payment is the amount of money that would make the recipient feel indifferent to receiving the present value now or receiving the future value later. All other things being equal, a safe sum of money is more valuable than a risky one. The *opportunity cost rate* is the expected rate of return on the next best use of the money with comparable risk. Therefore, the appropriate

discount rate should approximate the opportunity cost of using the money. The discount rate is not the expected inflation rate. A common discount rate used by many governments is the tax-exempt rate of borrowing. Alternatively, other governments employ the rate received for the investment of public funds. Yet other governments assume that the next best use of the funds is a tax cut. In this case, an even higher discount rate may be appropriate. However, there is no precise way to determine a uniformly appropriate discount rate. The higher the discount rate, the more quickly the impact of future effects is diluted. Thus, increasing the discount rate reduces the present value of any future effect.

Present values can be computed by using the future value relationship employed previously. Because $FV_n = PV_0(1 + i)^n$, present values can be derived by rearranging the relationship and solving for PV_0 :

$$PV_0 = FV_n[1/(1 + i)^n]$$

where:

PV_0 = Present value.

FV_n = Future value.

i = Interest rate per period.

n = Number of periods.

Example: $FV_n = \$10,000$; $i = 10\%$; $n = 5$ years

$$\begin{aligned} PV_0 &= FV_n[1/(1 + i)^n] \\ &= \$10,000[1/(1 + 0.10)^5] \\ &= \$10,000[1/(1.10)^5] \\ &= \$10,000[0.6209] \\ &= \$6209 \end{aligned}$$

In this example, investing \$6209 today at 10% compounded interest would yield \$10,000 at the end of 5 years.

Because of the complexity of present value calculations, *present value interest factors* (PVIFs) are commonly programmed into spreadsheet functions and financial calculators and reproduced in tables. The PVIF is value today of \$1 payable or receivable (n) periods in the future discounted at a compound interest rate of (i) percent per period. Because of this, all PVIFs will be less than one. Therefore, when multiplying a future value by these factors, the future value is discounted down to present value. PVIFs can be derived by taking the reciprocal of the respective future value interest factors ($1/FVIF$) in Table 3.

When the present value, future value, and number of periods are known, the interest rate or rate of return can be computed:

$$i = (FV/PV)^{1/n} - 1$$

Example: $FV_n = \$10,000$; $PV_0 = \$6209$; $n = 5$ years

$$\begin{aligned} i &= (FV/PV)^{1/n} - 1 \\ &= (10,000/6209)^{1/5} - 1 \\ &= (1.6106)^{0.2} - 1 \\ &= 1.1000 - 1 \\ &= 0.10 \\ &= 10\% \end{aligned}$$

In this example, investing \$6209 today in return for \$10,000 at the end of 5 years will result in a rate of return of 10%.

ANNUITIES

An *annuity* is a series of equal deposits, withdrawals, payments, or receipts that occur at evenly spaced time intervals. Monthly salaries, monthly leases or rental payments, payments from retirement accounts, quarterly dividends payments, semiannual bond interest payments, and annual life insurance premiums all involve annuities. An *ordinary annuity* (annuity in arrears) has payments or receipts at the end of each period. An *annuity due* (annuity in advance) has payments or receipts occur at the beginning of each period.

Future Value of an Ordinary Annuity

The *future value of an ordinary annuity* is the value that a stream of future payments or receipts will grow to after a given number of periods at a specific compound interest rate. The future value of an ordinary annuity can be determined by computing the future value of each individual payment in the series using the future value formula and then summing the results. Similarly, the future value of a series of uneven payments or receipts can be found by compounding each payment to the end of the stream and then summing the future values.

$$\begin{aligned} FV_{1...n} &= FV_{n-1} + \cdots + FV_2 + FV_1 + FV_0 \\ &= [R(1 + i)^{n-1}] + \cdots + [R(1 + i)^2] \\ &\quad + [R(1 + i)^1] + [R(1 + i)^0] \end{aligned}$$

where:

FV_n = Future value in time period (n).

R = Annuity or rental payment per period.

i = Interest rate per period.

n = Number of periods.

Example: $R = \$10,000$; $i = 10\%$; $n = 5$ years

$$\begin{aligned}
 FV_{1...5} &= FV_4 + FV_3 + FV_2 + FV_1 + FV_0 \\
 &= [R(1+i)^4] + [R(1+i)^3] + [R(1+i)^2] \\
 &\quad + [R(1+i)^1] + [R(1+i)^0] \\
 &= [\$10,000(1+0.10)^4] + [\$10,000(1+0.10)^3] \\
 &\quad + [\$10,000(1+0.10)^2] \\
 &\quad + [\$10,000(1+0.10)^1] \\
 &\quad + [\$10,000(1+0.10)^0] \\
 &= [\$10,000(1.10)^4] + [\$10,000(1.10)^3] \\
 &\quad + [\$10,000(1.10)^2] + [\$10,000(1.10)^1] \\
 &\quad + [\$10,000(1.10)^0] \\
 &= [\$10,000(1.4641)] + [\$10,000(1.331)] \\
 &\quad + [\$10,000(1.21)] + [\$10,000(1.1)] \\
 &\quad + [\$10,000(1)] \\
 &= \$14,641 + \$13,310 + \$12,100 + \$11,000 \\
 &\quad + \$10,000 \\
 &= \$61,051
 \end{aligned}$$

In this example, depositing \$10,000 at the end of each of the next 5 years at an annually compounded interest rate of 10% will yield \$61,051 at the end of 5 years.

Future value interest factors for annuities (FVIFAs) are commonly programmed into spreadsheet functions and financial calculators and reproduced in tables. The FVIFA is the future value of a series \$1 payments or receipts per period after (n) periods at a compound interest rate of (i) percent per period. FVIFA can be computed directly using the following formula:

$$FVIFA_n^i = [(1+i)^n - 1]/i$$

Example: $R = \$10,000$; $i = 10\%$; $n = 5$ years

$$\begin{aligned}
 FVIFA_5^{0.10} &= [(1+0.10)^5 - 1]/0.10 \\
 &= 6.1051 \times \$10,000 = \$61,051
 \end{aligned}$$

In this example, depositing \$10,000 at the end of each of the next 5 years at a compound interest rate of 10% will yield \$61,015 at the end of 5 years.

When the future value, interest rate, and number of periods of an ordinary annuity are known, the periodic annuity payment may be determined:

$$R = FV_0 / \{[(1+i)^n - 1]/i\}$$

Example: $PV_n = \$61,051$; $i = 10\%$; $n = 5$ years

$$\begin{aligned}
 R &= \$61,051 / \{[(1+0.10)^5 - 1]/0.10\} \\
 &= \$61,051 / \{[(1.10)^5 - 1]/0.10\} \\
 &= \$61,051 / \{[(1.61051) - 1]/0.10\} \\
 &= \$61,051 / \{0.61051/0.10\} \\
 &= \$61,051 / \{6.1051\} \\
 &= \$10,000
 \end{aligned}$$

In this example, in order to accumulate \$61,051 at the end of 5 years, \$10,000 would need to be deposited at the end of each of the next 5 years, earning 10% interest compounded annually.

Present Value of an Ordinary Annuity

The *present value of an ordinary annuity* is the amount which, if invested with compound interest, would generate a stream of equal payments or receipts at evenly timed intervals. More specifically, the present value of an ordinary annuity is the amount that must be invested today at a specific interest rate so that an equal amount may be withdrawn at the end of each period, such that the original principal and all accumulated interest will be completely exhausted at the end of the annuity. The present value of an ordinary annuity can be determined by computing the present value of each payment in the series using the present value formula and then summing the results. Similarly, the present value of a series of mixed value payments or receipts can be found by summing the present values of the individual payments.

$$\begin{aligned}
 PV_{1...n} &= PV_1 + PV_2 + PV_3 + \cdots + PV_n \\
 &= [R/(1+i)^1] + [R/(1+i)^2] \\
 &\quad + [R/(1+i)^3] + \cdots + [R/(1+i)^n]
 \end{aligned}$$

where:

PV_0 = Present value.

R = Annuity or rental payment per period.

i = Interest rate per period.

n = Number of periods.

Example: $R_n = \$10,000$; $i = 10\%$; $n = 5$ years

$$\begin{aligned}
 PV_{1...n} &= PV_1 + PV_2 + PV_3 + \cdots + PV_n \\
 &= [R/(1+i)^1] + [R/(1+i)^2] + [R/(1+i)^3] \\
 &\quad + \cdots + [R/(1+i)^n] \\
 &= [\$10,000/(1+0.10)^1] + [\$10,000/(1+0.10)^2] \\
 &\quad + [\$10,000/(1+0.10)^3] \\
 &\quad + [\$10,000/(1+0.10)^4]
 \end{aligned}$$

$$\begin{aligned}
& + [\$10,000/(1 + 0.10)^5] \\
& = [\$10,000/(1.10)^1] + [\$10,000/(1.10)^2] \\
& \quad + [\$10,000/(1.10)^3] + [\$10,000/(1.10)^4] \\
& \quad + [\$10,000/(1.10)^5] \\
& = [\$10,000/(1.1)] + [\$10,000/(1.21)] \\
& \quad + [\$10,000/(1.331)] + [\$10,000/(1.4641)] \\
& \quad + [\$10,000/(1.61051)] \\
& = [\$9091] + [\$8264] + [\$7513] + [\$6830] \\
& \quad + [\$6209] \\
& = \$37,907
\end{aligned}$$

In this example, receiving \$10,000 at the end of each of the next 5 years at an annually compounded interest rate of 10% is equivalent to \$37,907 today.

Present value interest factors for annuities (PVIFAs) are commonly programmed into spreadsheet functions and financial calculators and reproduced in tables. The PVIFA is the present value of \$1 received for (n) periods at a compound interest rate of (i) percent per period. PVIFA can be computed directly using the following formula:

$$PVIFA_n^i = [1 - 1/(1 + i)^n]/i$$

Example: $R = \$10,000$; $i = 10\%$; $n = 5$ years

$$\begin{aligned}
PVIFA_5^{0.10} &= [1 - 1/(1 + 0.10)^5]/0.10 \\
&= [1 - 1/(1.10)^5]/0.10 \\
&= [1 - 1/(1.6105)]/0.10 \\
&= [1 - 0.62093]/0.10 \\
&= [0.37907]/0.10 \\
&= 3.7907 \times \$10,000 = \$37,907
\end{aligned}$$

When the present value, interest rate, and number of periods of an ordinary annuity are known, periodic annuity payment may be determined:

$$R = PV_0 / \{[1 - (1/(1 + i)^n)]/i\}$$

Example: $PV_n = \$37,907$; $i = 10\%$; $n = 5$ years

$$\begin{aligned}
R &= \$37,907 / \{[1 - (1/(1 + 0.10)^5)]/0.10\} \\
&= \$37,907 / \{[1 - (1/(1.10)^5)]/0.10\} \\
&= \$37,907 / \{[1 - (1/(1.6105))]/0.10\} \\
&= \$37,907 / \{[1 - (0.62093)]/0.10\} \\
&= \$37,907 / \{[0.37907]/0.10\} \\
&= \$37,907 / \{0.37907/0.10\} \\
&= \$37,907 / \{3.7907\} \\
&= \$10,000
\end{aligned}$$

In this example, if \$37,907 is deposited today earning 10% interest compounded annually, \$10,000 would be available for withdrawal at the end of each of the next 5 years.

Annuities Due

An annuity due is a constant stream of payments starting immediately. Annuities due are beginning-of-year annuities. The future value of an annuity due is identical to an ordinary annuity except that each payment occurs at the beginning of a period rather than at the end. Because each payment occurs one period earlier, the present value of an ordinary annuity can be computed and then multiplied by $(1 + i)$. Likewise, the present value of an annuity due is identical to an ordinary annuity, except that each payment occurs at the beginning of a period rather than at the end. Because each payment occurs one period earlier, the present value of an ordinary annuity can be computed and then multiplied by $(1 + i)$.

BIBLIOGRAPHY

1. Brealey, R.A.; Myers, S.C. *Principles of Corporate Finance*, 6th Ed.; McGraw-Hill, 2000.
2. Brealey, R.A.; Myers, S.C.; Marcus, A.J. *Fundamentals of Corporate Finance*, 3rd Ed.; McGraw-Hill, 2001.
3. Brigham, E.F.; Houston, J.F. *Fundamentals of Financial Management*, 9th Ed.; South-Western, 2001.
4. Finkler, S.A. *Financial Management for Public, Health, and Non-for-Profit Organizations*; Prentice-Hall, 2001.
5. Helfert, E.A. *Techniques of Financial Analysis: A Practical Guide to Managing and Measuring Business Performance*, 9th Ed.; Irwin, 1997.
6. Mikesell, J.L. *Fiscal Administration: Analysis and Applications for the Public Sector*, 5th Ed.; Wadsworth, 1999.
7. Moyer, R.C.; McGuigan, J.R.; Kretlow, W.J. *Contemporary Financial Management*, 8th Ed.; South-Western, 2001.
8. Rachlin, R. *Handbook of Budgeting*, 4th Ed.; John Wiley, 1999.
9. Rosen, L.R. *The McGraw-Hill Handbook of Interest, Yields, and Returns*; McGraw-Hill, 1995.
10. Ross, S.A.; Westerfield, R.W.; Jordan, B.D. *Fundamentals of Corporate Finance*, 5th Ed.; McGraw-Hill, 2000.
11. Scott, D.L.; Moore, W.K. *Fundamentals of the Time Value of Money*; Praeger, 1984.
12. Shim, J.K.; Siegel, J.G. *Encyclopedic Dictionary of Accounting and Finance*; Fine Communications, 1996.
13. Woelfel, C.J. *Desktop Guide to Money, Time, Interest, and Yields*; Probus, 1986.

Times Series Analysis

Bruce D. Rogers

Institute of Government, Tennessee State University, Nashville, Tennessee, U.S.A.

INTRODUCTION

Times series analysis enables public administrators and policy analysts to examine values of a variable over equally spaced intervals of time (unemployment figures, monthly or yearly). Using times series analyses, administrators and analysts can discern patterns in the values that then enable them to forecast future values based on historical and existing patterns.

Times series analyses are important in producing the economic forecast indicators reported by government agencies each month. Federal, state, and local budget analysts continuously track revenues and expenditures over time. Planners make infrastructure decisions based upon population shifts over time, and census officials make projections about trends in society and lifestyles using times series analysis. Police and national crime agencies analyze data collected over time to draw conclusions about trends regarding the security of residents and to look at the impact that new anticrime laws may have had. Agricultural governmental planners are concerned with the variability of crop yields and price supports over time. The continuing national debate about the performance of public schools relies on the times series analysis of performance indicators. Indeed, almost every field in public management and policy analysis collect and analyze data using times series analysis.

TIME FRAMES AND STATISTICAL FORECASTING

While public administrators and policy analysts rely on times series analysis to forecast future values, the weakness of statistical forecasting can be related to time frames for forecasting. Confidence in statistical forecasting is dependent upon the time frames for forecasting and the nature of the information being forecast. Short-term forecasts are generally more accurate in policy analysis than are long-term forecasts. Different disciplines permit greater long-range forecasting than others. Economists may accurately forecast ahead for a year. Those forecasts are fine tuned by the scores of economic indicators that are produced in the course of any year. Educators may be able to forecast school physical facility needs for a decade or more after

analyzing new census data (assuming that the numbers of immigrants remains close to the numbers anticipated). Meteorologists can accurately forecast weather successfully for several days, while longer-term forecasts are more subject to change. The science of astronomy permits astronomers to forecast certain events ahead, or back, for thousands of years. Forecasts are, therefore, evaluated by the time frames of the forecasts and the disciplines in which the forecasts are made.

REGULARITY AND HISTORICITY

Critical for times series is the concept of regularity: recording values at regular intervals of time. It is more productive to look at values over regular time intervals than to look at many values taken all at once, if our purpose is to examine values over time. The concept of historicity is also important: What occurred in the past is a predictor for the future. These two concepts are at the core of times series analysis.

Historical patterns can be effective predictors in the absence of major transforming events. Predictions about terrorists and terrorism were stood on their heads by the events of September 11, 2001: the attacks on the World Trade Center in New York City, the Pentagon in Washington, DC, and the airplane crash in Pennsylvania. Prior to that date, profiles of those who committed acts of terror over the previous three decades would have included only a few who were among the more educated, the more affluent, or older than 30. Prior to that date, acts of terror against the United States did not include coordinated, multiple series of acts in several states at one time. Those responsible for guarding against terrorism relied on the historical patterns until that time and concentrated on profiles different than those described above.

Forecasting in certain policy areas is dependent upon the creativity of analysts to envision alternative scenarios not predicted by past data and to creatively look at multiple interpretations of historical patterns. When analysts look at a diversity of scenarios, forecasters may have greater confidence that the forecast will include a range of outcomes. The World Trade Center experience demonstrates that forecasting must rely upon multiple types and methods of data in cases where alternative outcomes are possible.

Another example of forecasting during a change in historical patterns occurred in the 1980s after large numbers of mental health patients were released from inpatient facilities. Prior to that, planning and forecasting for the care of outpatients assumed continued reliance on inpatient facilities. After the new policy was implemented, prior times series data was no longer relevant for forecasting future outpatient needs.

NO ASSUMPTION OF INDEPENDENCE

The independent variable in times series analysis is “time,” measured in units such as: seconds, minutes, hours, days, weeks, months, quarters, or years. Observations of the dependent variable in times series analysis are often a function of the temporal order in which the data were collected. Since values in a later time period may be dependent upon previous values collected (referred to as “autocorrelation” or “serial correlation”), the assumption of “independence” is not applicable in times series analysis. Statistical techniques such as the Box-Jenkins Times Series Analysis help correct for this.^[1]

COMPONENTS OF TIMES SERIES

Forecasts can be improved if the analyst can understand and identify the patterns in the data. Breaking down the data into its component parts is a technique called decomposition. If you plotted the number of catalog mail orders over the past 10 years by month, there would appear to be an upward linear trend for the entire period. At the same time, there would also be a pattern recurring annually between the Thanksgiving and Christmas season. The seasonal “decomposition” method enables the analyst to break down and to separate the times series analysis to look at trend, seasonal, cyclical, and other variations in the data.^[2]

To assist in understanding past occurrences and to help forecast the future, times series analysis has been assumed to have four component elements: 1) trends; 2) seasonal variations; 3) cyclical patterns; and 4) unexplained reasons. A forecast is made considering each of the four elements separately and then together.

1. Trends over time imply that the data might increase or decrease over time. If one plotted the U.S. population between 1960 and 2000, one would observe that while the trend shows increases, it has not been in a straight line for the entire period.
2. Seasonal variations imply that certain patterns recur at regular intervals. Employment in state tourism industries exhibits regular patterns.

The pattern is what constitutes the seasonal variation in the data. While seasonal may refer to the seasons of the year, it can also refer to the increase in sales of the American flag just before federal holidays or to the variations in applications for state hunting licenses just before deer hunting begins each year.

3. Cyclical patterns occur somewhat irregularly and in durations longer than seasonal variations but shorter than trends. If one analyzes FBI Crime Reports that list incidences of various crimes, certain cyclical patterns may be discerned. During periods of high unemployment, there may be greater incidences of burglaries. During periods of high employment, as in the period 1998–2000, crime indices declined.
4. Times series data that cannot be explained in terms of trends, seasonal variations or cyclical patterns are often referred to as random, the fourth component of times series.^[3]

METHODS OF IDENTIFYING TRENDS IN TIMES SERIES ANALYSIS

There are a number of statistical methods used to identify trends. Three techniques are particularly common:

1. Moving Averages technique takes data from predefined past periods (e.g., every three-year or five-year period) and combines them as a way of forecasting future periods. Each new data point examined thereby becomes an average of past values. This technique gives equal weight to every data point and serves to “smooth” out the data.
2. Box-Jenkins Models referred to as the Autoregressive Integrative Moving Averages technique looks at the past behavior of the data to select a forecasting model, from among several, to be used in the times series analysis. Box-Jenkins provides good short-term forecasts. To be used, though, large amounts of data are required. The models are:
 - Autoregressive models provide forecasts based on a linear examination of past observations.
 - Moving Average models forecast future values based on combining past values and averaging them into new values.
 - Autoregressive–Moving Average models, which are a combination of the first two models, listed above.
3. Exponential Smoothing is a third times series analysis technique. Unlike the moving averages

model that gives equal weight to all observations, greater weight is given in the exponential smoothing model to more recent data observations. This model assumes that the more recent observations are more important in forecasting than are older points.^[3]

These techniques require a degree of statistical sophistication to be utilized.

EXAMPLES OF TIMES SERIES ANALYSIS

There are several types of times series analysis:

1. Describing a trend over time. Have the number of motorcycle highway fatalities declined over time? Have eighth-grade proficiency scores improved over time? Has the rate of growth of state expenditures declined in the last five years?
2. Impact analysis extends times series beyond just description. Through the use of an interrupted times series design, such questions as the following can be addressed: Has the trend in motorcycle deaths declined since a law requiring drivers to use helmets was enacted? Have motorcycle highway deaths declined since medical evacuation helicopter flights were introduced for accidents on state roads? Have eighth-grade proficiency scores improved since the new eighth-grade curriculum was introduced? Did eighth-grade proficiency scores increase after the school board mandated that only those who passed the exam could continue onto high school? Did the rate of growth of state expenditures decline after state voters passed a new proposition limiting state financial mandates for programs at the local level? Did the rate of growth of state expenditures decline after the state legislature reduced state income tax levels? The design is "interrupted" because all of the questions refer to an "intervention," such as the enactment of a new law or rule or regulation. Thus, a new element has been added to the times series, permitting an observer to look at before intervention (baseline) and after intervention times series data. Policy analysts thereby have an additional tool to assess whether a change in policy resulted in a change in behavior.
3. Times series analysis is also concerned with analyzing data by degrees. For example, in states with new helmet usage laws and new laws requiring lower motorcycle speeds on highways, is the decrease in motorcycle highway fatalities greater than in states with only new helmet usage laws? Do eighth-grade proficiency scores increase more in areas where the new eighth-grade curriculum is introduced all at once or where it is phased in? Do scores increase more in areas where both the new curriculum and student promotion requirements are phased in at the same time? Is the decrease in the rate of growth in state expenditures greater in states with strong legislative control over the state budget as compared with states with stronger executive control over the state budget? Regression techniques permit the researcher to answer such questions.
4. Finally, times series analysis is used for forecasting purposes. Based on the data, can we predict how many fewer fatalities might occur if there was another law passed further reducing the speed limits that motorcycles can travel? Can we forecast the results of student performance if even stricter promotion requirements are put into place?

ECONOMIC TIMES SERIES ANALYSIS FOR GOVERNMENT

Government managers and policy analysts are major consumers and users of times series analyses. Examples of available data can be found at Ref.^[4]. The comprehensive times series data library available at the Princeton web site provides an annotated listing of available data sets. Instructions on how to use those data sets are available at the individually listed web sites. The sites include times series of economic indicators, consumer price index data, Bureau of Labor Statistics data, trade surpluses or deficits, stock trading data on multiple exchanges, Organization for Economic Cooperation and Development data, and data for both the United States and foreign nations.

The Economic Report of the President,^[5] contains data covering a 30-year period. For the last 3 years, there are quarterly breakdowns, and for the last 2 years, monthly data is provided for many data sets. Population, government finance, trade, agriculture, and budget data are available. There is an excellent report at the site, with times series tables, about the demographic changes in the United States.

The Economic Time Series Page at Ref.^[6] offers access to wide ranging, frequently requested domestic times series data sets from many government agencies. Data is also broken down by states. This web site permits the user to transform the data from, for example, monthly to quarterly or monthly to annually. Depending on the government data set employed, users can even convert data to percentages. Such manipulations permit the user to transform the data for their own

research purposes. Also at this site, the Bureau of Economic Analysis publishes the Regional Economic Information System, which provides economic data for states, counties, and metropolitan areas for the 25-year period from 1969–1994. Earnings data, employment information, personal income, farm income, and transfer payments are available.

ADDITIONAL GOVERNMENT TIMES SERIES DATA

Using Ref.^[7], permits users to access over 100 U.S. government agencies at one site. Numerous government agencies listed there have reports, which utilize times series data. For example, an analyst in a state department of education may want to compare overall state trends with that of the nation as a whole. By going to Fedstats, the user could click successively on:

- National Center for Education Statistics
- Encyclopedia of ED Stats
- Digest of Educational Statistics 1999
- Preface
- Introduction^[8]

In this example, the user would find, in the introduction to the report times series data on enrollment at the elementary/secondary/higher education levels, times series data on teachers, public perception of education, and student performance in reading, mathematics and science, as well as international comparisons and data on graduates and degrees. For many agencies other than education, there is a similar breadth of times series data available at the site. Fedstats is an excellent resource for students in public administration and policy analysis who need to be informed users and consumers of times series data.

OFFSTATS: Official Statistics on the Web web site,^[9] provides wide-ranging data sets that are organized by country and about 60 topical areas. The country and topical indices offer a rich array of data sets and sources for the United States and for all countries of the world. Students in comparative and international public administration will benefit from a study of the times series resources available at OFFSTATS.

The University of Michigan Document Center's Statistical Resources on the Web,^[10] offers a vast data resource, including times series analysis, for data users who are working in public administration and public policy. Broad topics are divided into more manageable subject areas. If agricultural policy analysts need to see trends in crop yields or government payments to farmers they could go to this site and navigate to find the data that they need. There are data sets for many government function areas.

The Inter-University Consortium for Political and Social Research at the University of Michigan conducts a number of repeated surveys: the General Social Survey, the National Interview Health Survey, the Current Population Survey, and the American National Election Surveys, as well as Consumer Surveys.

GRAPHING TIMES SERIES DATA

Plotting the data is the first step in doing times series analysis. Trend lines can be applied to the graph. Cyclical patterns (seasonal variations if applicable) or any other major changes in the data can be noted.

The following simple times series analyses are taken from the U.S. Department of Education, National Center for Education Statistics, Digest of Education.^[11] The data show that the postsecondary enrollment patterns for men and women both show

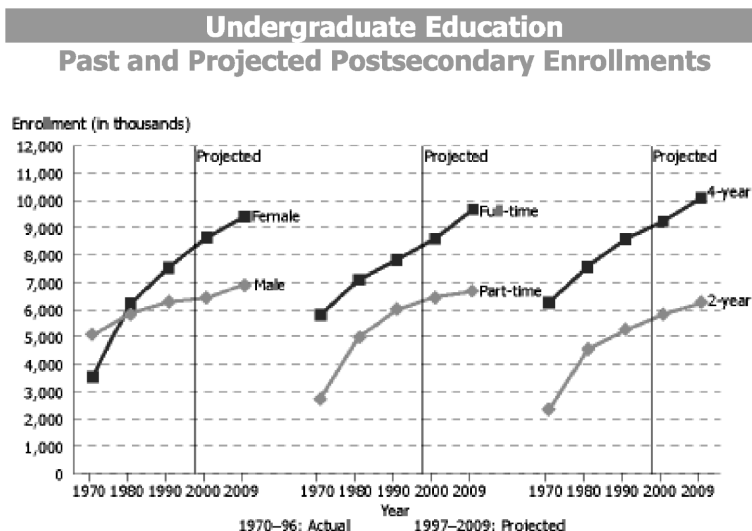


Fig. 1 Postsecondary enrollment: Total enrollment in degree-granting 2- and 4-year postsecondary education institutions, with projections, by sex, enrollment status and level: Fall 1970–2009. Source: U.S. Department of Education, NCES. *Digest of Education Statistics 1998* (NCES 1999-036), 1999, and *Projections of Education Statistics to 2009* (NCES 1999-038), 1999.

upward trends between 1970 and 2000, but with a steeper rise for women over the period, and a widening gap in projections for 2009. Indeed in 1980, the gap was practically closed (Fig. 1).

While the gap between full-time and part-time enrollment decreased between 1970 and 1990, the numbers of full-time enrollments outpaced part-time enrollments between 1990 and 2000 and the projections for 2009 show the same wide difference as in 1970. Similarly, the trends in enrollments at two-year institutions showed a narrowing of the gap with four-year institutions between 1970 and 1980, with an increasingly wider difference between 1980 and 1990 and between 1990 and 2000. The projected enrollment in 2009 forecasts as large a difference between the types of institutions as in 1970.

Graphing the data enables managers and researchers to visually assess the patterns in the data, assisting in making decisions about the techniques to be used in the analysis of the data. If the trend line is relatively stable, then simple moving averages or exponential smoothing will be called for. Fluctuating trend lines may require decomposition analysis or Box-Jenkins models.

REFERENCES

1. Box, G.E.P.; Jenkins, G.M. *Times Series Analysis: Forecasting and Control*; Holden-Day: San Francisco, 1976.
2. Avent, S.R. *Times Series Analysis An Introduction*; <http://userwww.sfsu.edu/~efc/classes/biol710/time%20series/Time%20Series%20Analysis> (accessed August 24, 2001).
3. Makridakis, S.; Wheelwright, S.C.; Hyndman, R.D. *Forecasting: Methods and Applications*; Wiley: New York, 1998.
4. <http://www.princeton.edu/~data/datalib/timeser.html> (accessed September 10, 2001).
5. <http://w3.access.gpo.gov/eop/> (accessed August 11, 2001).
6. <http://www.economagic.com> (accessed August 11, 2001).
7. <http://www.fedstats.gov> (accessed July 30, 2001).
8. <http://nces.ed.gov/pubs2000/Digest99/introduction.html> (accessed September 12, 2001).
9. <http://www2.auckland.ac.nz/lbr/stats/offstats/OFFSTATSmain.htm> (accessed August 14, 2001).
10. <http://www.lib.umich.edu/govdocs/stats.html> (accessed August 10, 2001).
11. <http://nces.ed.gov/pubs2000/coe2000/charts/chart7.html> (accessed September 20, 2001).

Trade Policy

Philip A. Mundo

Political Science Department, Drew University, Madison, New Jersey, U.S.A.

INTRODUCTION

U.S. trade policy is both a matter of foreign policy and domestic politics. Mainly because of its constitutional roots in commerce and revenue, trade policy is made largely through the domestic policy process. The dynamics of trade policy making resemble other kinds of major legislation (e.g., budgets) more than foreign policy. Yet trade policy counts in international affairs, and public officials must devise a careful strategy that achieves foreign policy objectives, while accommodating domestic political forces. Although this is usual for other countries, the peculiarities of American domestic politics have serious consequences for U.S. trade policy.

This article outlines the history of U.S. trade policy, its main components, and the political process that produces it. Throughout the article, the reader should pay attention to the delicate balance between foreign policy goals and the messiness of the domestic policy process.

BRIEF HISTORY

Trade policy is as old as the nation. From the beginning, it reflected the basic assumptions in Alexander Hamiltonian's views of the proper role between government and the economy. For Hamilton, an activist government should intervene in the economy to facilitate economic growth. Centralized and focused, the national government was in the best position to do such things as construct infrastructure, provide a centralized and reliable financial system, and promote American commerce abroad.^[1] Toward this end, the national government established high tariffs from the outset for two purposes: initially to raise revenue for the fledgling federal government, and later to protect new American industries from potentially damaging foreign competition.^[2]

American tariffs ebbed and flowed throughout the nineteenth century. Tariffs rose in the first part of the century, fell in the 1830s and early 1840s, then fluctuated until the Civil War.^[3] Beginning in the 1880s, national leaders began to see the value in opening foreign markets to American producers. The fact that American industry had experienced a burst of growth

following the Industrial Revolution made many companies competitive abroad. Threatening to limit access to the American market was one way to gain access to foreign markets; American officials threatened to raise tariffs if a foreign government did not give U.S. manufacturers more access to its market.^[4] Tariffs rose and fell throughout the 1910s and 1920s as public officials responded to various domestic and international pressures. A wave of protectionism dominated the 1920s in tandem with post-World War I isolationism that characterized U.S. foreign policy.^[5,6]

Following the stock market crash of 1929 and the onset of the Great Depression, Congress and the president reacted by passing the Smoot-Hawley tariff. Resulting largely from members of Congress working together to accommodate various organized interests,^[7] the Smoot-Hawley tariff added to national economic problems. As the United States raised tariffs for the purpose of blocking imports and protecting American industries, other nations did the same. The result was a collapse in international trade, which was certainly a contributing factor to the further decline of the U.S. economy, as well as others.

Learning the hard lesson, Congress and the president reversed course 4 years later, enacting the Reciprocal Trade Agreements Act (RTAA) of 1934. The act gave the president more discretion over trade policy, and it reflected the establishment of the liberal economic idea of free trade in the American policy-making machinery.^[4,5] The RTAA set in motion a commitment by U.S. policy makers to freer trade; tariffs would come down in negotiated deals with other nations on the basis of reciprocity.^[2]

Following World War II, the United States reiterated its commitment to free trade by putting its considerable weight behind the establishment of a global international trade organization—the General Agreement on Tariffs and Trade (GATT). After the war, the United States was the major economic, political, and military power in the West. In part because policy makers were committed to the idea of liberal economics on a global scale and because it was in American interests to help the world economy recover, the United States pushed for the creation of the GATT, along with international organizations dealing with finance and aid—the World Bank and the International Monetary Fund.

The original goal was to create a supranational body that would have the authority to adjudicate disputes between member countries. Named the International Trade Organization (ITO), the international body would ensure that the liberal economic idea of free trade became the dominant theme in postwar international business and finance. The goal was not fully realized, however, because nations could not agree to turn over a measure of sovereignty to the supranational ITO. In its place, a compromise emerged that was institutionalized in the GATT. The GATT provided a means for making rules governing international trade, but it did not have the authority to adjudicate disputes between member nations.^[8] Nevertheless, the GATT's activities through the 1990s were the major developments shaping international trade.

1960s Through the 1980s: New Concerns, Greater Complexity

The 1960s and 1970s saw the emergence of new concerns in international trade. Tariffs remained significant as a barrier to trade, but they were reasonably easy to handle. Because tariffs were calculated as a percentage of the price of a good, the task of lowering them involved a relatively simple calculation that could easily be negotiated. The next major barrier to trade was much more difficult to manage. Non-tariff barriers (NTBs) consist of a range of practices that make it more difficult for a foreign producer to operate in a market than domestic producers. NTBs include, for example, domestic regulations over such things as health, environment, safety, and food standards. Lowering NTBs is a more complicated task than simply cutting tariffs, and NTBs themselves are embedded in matters that can be important to a society, quite apart from their effects on trade. In the 1960s and 1970s, GATT negotiations sought partly to address the problem of NTBs, moving ahead only slowly. Nations did not want to sacrifice control over regulating, say, food safety for the sake of expanding the flow of goods and services across national borders.

The picture became even more complicated in the 1980s, as national governments began to wrestle with trade-in services. Services are products that are not based in some physical thing. They include such industries as film making, construction, medicine, legal services, and software. As the importance of these industries became more obvious, national policy makers sought ways to facilitate their flow across national borders. Reducing barriers to trade-in services, mostly in the form of NTBs, proved to be tricky business, and much work remains to be done.^[9] The point here is that the complexity and challenge presented by trade-in services further complicated the task of policy makers.

Add the problem of intellectual property protection to this increasingly messy picture. Producers, largely but not exclusively American, wanted to establish international rules to protect their intellectual property—patents, copyrights, trademarks, software protection, and so on. Intellectual property protection was and is notoriously lax in some nations, and producers found that doing business in these countries was costly. The connection between intellectual property concerns and trade led some governments to pursue the issue through the GATT, resulting in the creation of trade rules governing intellectual property in the Uruguay Round of GATT negotiations.^[10]

Where are we Now?

The World Trade Organization (WTO) replaced the GATT in the 1990s. To some extent fulfilling the promise of its stillborn predecessor, the ITO, the WTO is a supranational body with the authority to adjudicate disputes among member countries. It is also a club many nations want to join. China, for example, sought for years to gain membership to the GATT, and now the WTO, to benefit from the access membership would provide to foreign markets and the promise of foreign investment at home. The WTO has become the target of opponents of globalization. Viewed as an instrument of international capitalism, the WTO is one of the main international agencies, along with the World Bank and the IMF, held responsible for the negative consequences of globalization.

THE MAIN COMPONENTS OF U.S. TRADE POLICY

The United States pursues its interest in international trade through several means: global negotiations, regional pacts, bilateral agreements, and administrative trade relief. Each approach is rooted in the domestic policy process, although in different ways.

Global Negotiations

As stated earlier in the article, the United States has been a major proponent of the global trade regime, first in the form of the GATT, then in the WTO. The domestic politics connected to these efforts were initially fairly straightforward. As the negotiations mostly involved tariff reductions, policy makers had to deal with domestic producers as they sought to secure their interests in the new negotiations. Thus, American agriculture, among the most durable and powerful interests in American politics, has repeatedly protected its interests, still winning subsidies

from the federal government.^[11] As labor unions saw the prospects of key unionized industries (steel, automobiles) in the international economy sink, they switched from a largely protrade stance to opposition to free trade.

As GATT negotiations took on NTBs, services, and intellectual property protection, the domestic politics scene changed. New players in international trade (e.g., environmental groups) jumped into the fray as they saw American environmental regulation potentially being considered as nettlesome NTBs to be negotiated away under the GATT/WTO framework. Add to these issues concerns for labor rights abroad and human rights more generally. American interest groups focused on these issues increasingly came to see free trade initiatives as undermining basic rights. These shifts in the political landscape appeared in sharp relief at the December 1999 WTO meeting in Seattle. Protestors of every stripe took to the streets, impressing delegates to the meeting and national policy makers of the seriousness of their intent and the practical value in paying attention to their concerns.

Regional Trade Agreements

The United States has entered into several multilateral regional trade agreements, the most noteworthy of which is the North American Free Trade Agreement (NAFTA). The three-country pact reduced tariffs between Canada, the United States, and Mexico. As a result of pressure from American unions and environmental groups, NAFTA includes two side agreements, one each for labor and the environment. The purpose of these side agreements is to prevent freer trade among the three nations from leading to labor abuses and accelerated environmental degradation as governments turned away from such activities to increase the competitiveness of their producers.^[12] Whether the side agreements have been successful remains an open question, but the fact that they became a codified part of the formal agreement at all attests to their importance and the growing complexity of international trade agreements.

In addition to the potential for fostering labor and environmental abuses, NAFTA and other regional agreements are controversial with respect to how much they contribute to global free trade. Do trade pacts simply move the locus of protection from national governments to the intergovernmental agency governing the pact? Or, because trade is freer within them, do they lead to the goal of free trade internationally?^[13,14] The economic debate as it relates to NAFTA has become nearly moot in recent years, as the United States' plans for expanding NAFTA in South America have faltered.^[15]

Bilateral Agreements

The United States has entered into numerous bilateral agreements—pacts between two nations typically focused on trade outcomes. These agreements deal with all sorts of industries, from automobiles to semiconductors, and they are usually the result of concerns that American producers may suffer the results of competitive foreign companies. Sector-specific agreements—those dealing with only one industrial sector—are suspect in the eyes of true free traders. They promise certain outcomes—a quota, for example—and aim to control trade. The United States' successive deals with Japan on semiconductors are examples of this kind of agreement. Its proponents argue that the agreements thwarted unfair trade practices on the part of Japanese firms and the Japanese government, whereas detractors say these deals distorted trade and ultimately damaged a major segment of the U.S. semiconductor industry.^[16]

Administrative Trade Relief

This is an enormously complex yet extremely useful tool for U.S. policy makers. Crafted by Congress and the president to provide a means of protection for affected industries, administrative trade relief procedures are essentially bureaucratic and regulatory in nature, and they can be used with precision. Several major forms of administrative trade relief play a significant role in U.S. trade policy. Section 201 is a straightforward provision of U.S. trade law that gives American producers adversely affected by foreign imports protection in the form of higher tariffs. The producers in this case do not have to demonstrate that foreign competitors or their governments have engaged in unfair trade practices. Countervailing duty is used when government agencies determine that a foreign producer has been subsidized by its government, giving it an unfair advantage over American competitors. A duty (or tariff) is levied on the products of the offending company. Antidumping, perhaps the most controversial of these methods, refers to American government efforts to counter a foreign producer's selling of products in the American market at below fair market value. As in the other cases, a tariff is the remedy.^[12]

The United States also uses forms of administrative trade relief to open foreign markets to American products. Section 301 engages when government agencies determine that a foreign government is unfairly blocking the entry into its market of American competitors. The remedy for this action is reciprocal impairment of trade on products produced in the offending country. Section 337 deals with intellectual property protection

in foreign companies, and it is used when an American company's goods or services have been reproduced without appropriate permission. A tariff on selected products coming from that country is the remedy.^[12]

In all these administrative trade relief methods, many decision criteria are based in findings of fact by bureaucratic agencies. Although the president makes the final decision in some cases (e.g., the Section 201 case against steel in 2002), the procedure is nearly automatic once injury and violations have been found. This aspect of administrative trade relief takes the load off the shoulders of elected politicians, especially members of Congress, in the sense that they can maintain a commitment to free trade, while also allowing for protection of key industries, many of which may be part of their constituencies.

THE DOMESTIC POLICY PROCESS AND POLITICS

U.S. trade policy is a product of the domestic policy process and politics. This presents some interesting problems because, in many ways, trade policy is part of the nation's range of foreign policies. It is best to consider this sweeping topic in two pieces: first, the role of the different parts of government in the development of trade policy; and second, the influence exerted by political parties and interest groups in this process.

Government

Congress and the presidency, of course, are principally responsible for the development of U.S. trade policy. Congress's central role is based in its constitutionally prescribed responsibility for commerce and revenue.^[2] Because of its constitutional roots, Congress is inevitably involved in trade policy, prohibited from handing it over entirely to the presidency. The constitutional basis also means that trade policy cannot be treated in the same way as other parts of foreign policy. For example, trade agreements must be approved by both houses of the Congress, not just the Senate, as is the case for other treaties.

Congress deals with trade policy much as it treats other major policy areas. Major trade bills are subject to considerable debate, committee hearings, lobbying, negotiations between the House and Senate, and ultimately, bargaining between Congress and the president. Because economic issues have always been at the heart of trade policy, economic interests represented by their groups fight hard to protect their gains and avoid losses. The arrival of environmental and labor rights issues complicate congressional treatment

of trade policy still further because the topics under consideration are not limited to the competitiveness of American producers and jobs. Dealing with complex environmental and labor rights issues that necessarily have implications for foreign countries makes trade policy an enormous challenge for Congress.

By the end of the nineteenth century, the presidency had become a proponent of free trade.^[17] Representing and accountable to the entire electorate, the president has the leeway to sacrifice the interests of one segment of society to enhance those of the rest of society. The chief executive is also in the position to take the lead with new initiatives that expand free trade. Presidents do not always act as though this is the right calculus; presidents have protected specific industries in response to various forms of political pressure.^[18]

Still, on balance, the president is generally inclined to favor free trade, but must rely on Congress to grant the executive the authority to engage in new initiatives to expand free trade. Congress requires the president to receive trade negotiating authority from the legislature before initiating negotiations. Congress has the constitutional authority to make this requirement because its role in trade policy is constitutionally based. Congress may exact concessions from a president before granting trade negotiating authority in the form of protection for specific industries, for example, or mandating that any new trade deal include provisions to protect the environment. This aspect of congressional–presidential relations with respect to trade policy opens the door for including all sorts of other issues into trade policy and, as a result, complicates the task of trade policy making.^[18]

Presidents since the Carter administration have relied on the Office of the U.S. Trade Representative (USTR) for policy initiatives and to take the lead position among top executive branch agencies in developing trade policies, and shepherding them through Congress and the bureaucracy. The USTR, the head of the agency, serves as the chief negotiator representing the United States in international trade negotiations. Created by Congress to be a toehold for the legislature in the executive branch on trade policy, the USTR now serves the president directly, superceding the authority of other agencies that have interest in trade policy, especially the Department of State.^[19]

Bureaucracy

As explained earlier, bureaucracy plays the major role in administrative trade relief. Government agencies determine the existence and cause of injury to American producers by foreign competitors. These responsibilities play to the strengths of government bureaucracies. They are staffed by experts in a wide range of areas who apply their technical skills to

assigned problems. These tasks also give bureaucrats enormous influence over U.S. trade policy because of their mastery and control of information. Outside experts have certainly challenged the government agencies involved in trade policy, mainly the International Trade Administration of the Department of Commerce (ITA) and the U.S. International Trade Commission. For example, some analysts complain about the methodology used by the ITA in antidumping calculations.^[18] Agencies can also serve as advocates of various trade policy positions and sources of ideas that can be used to modify existing policies. As is true for numerous other issue areas, bureaucrats are unelected policy makers with significant authority over an important set of policies.

Parties and Interest Groups

Political parties and interest groups have played an important role in trade policy development since the nineteenth century. At that time, the Republican party favored protection, while the Democratic party favored free trade. As America's industrial sector became more competitive with foreign producers, the party that represented their interests gradually shifted to a free trade stance. Thus, by the 1960s, the Republican party had become the party of free trade. Meanwhile, the Democratic party switched to the other side, especially as labor unions, among the Democratic party's most important supporters, began to see their interests in protection rather than free trade. Union leaders saw American jobs disappearing as American firms set up manufacturing plants abroad and as imports began to supplant home-grown products in the American market. The United Automobile Workers, for example, traditionally a free trade proponent, switched by the 1980s as foreign competition pushed Detroit to the economic wall. The Democratic party continues to look upon free trade with skepticism, although a centrist part of it, organized under the roof of the Democratic Leadership Council, favors free trade. For the most part, the Republican party supports free trade, although some Republican law makers quickly jump to the protectionist side if an industry in their constituency is threatened by foreign competition.

Interest groups had a great deal to do with the enactment of the Smoot-Hawley tariff. Lobbying legislators for their specific concerns, they pushed through a bill that had something for each special interest but not much for the public interest. Economic interest groups dominated the playing field on trade policy until the 1980s and 1990s. Some American producers saw advantage in opening foreign markets, as others sought protection from rapidly encroaching foreign

competition. Thus, business does not necessarily speak with one voice when it comes to trade policy.^[20] The addition of environmental groups and other non-economic groups to the fray has complicated the trade policy picture. They bring new issues to the table as it increasingly appears to them that trade policy is the cutting edge of globalization. At the start of the twenty-first century, labor unions are almost entirely protectionist, fearing that jobs will go where labor is cheaper and that foreign products will push American firms out of business.

CONCLUSION

U.S. trade policy is a product of the domestic policy process and politics, as well as a part of the United States' overall foreign policy. This arrangement creates some treacherous going for trade policy makers in the United States. They must pay attention to domestic forces and processes as much as they do in any other major domestic policy area, while also finding ways to negotiate credibly with other nations toward the end of expanding free trade. Free trade has won the day as a concept; most politicians pay at least lip service to supporting it, and many policy makers have spent a great deal of time and political capital seeking policies that expand free trade. Yet, there are political hazards. Elections are usually not won or lost on trade policy, but keeping certain interests in one's electoral camp may require using the various mechanisms available to policy makers to protect their interests by protecting their industry.

REFERENCES

1. Hamilton, A. Report on the subject of manufactures. In *The Papers of Alexander Hamilton*; Columbia University Press: New York, 1966; Vol. 10, 230–340.
2. Rothgeb, J.M., Jr. *U. S. Trade Policy: Balancing Economic Dreams and Political Realities*; CQ Press: Washington DC, 2001; pp. 14, 15, 19, 43–55.
3. Taussig, F.W. *The Tariff History of the United States*, 8th Ed.; Putnam: New York, 1931; 102 pp.
4. Goldstein, J. *Ideas, Interests, and American Trade Policy*; Cornell University Press: Ithaca, NY, 1993; 93–94.
5. O'Halloran, S. *Politics, Process, and American Trade Policy*; University of Michigan Press: Ann Arbor, MI, 1994; pp. 81, 82, 85, 86.
6. Schwab, S.C. *Trade-Offs: Negotiating the Omnibus Trade and Competitiveness Act*; Harvard Business School Press: Boston, 1994; 29 pp.
7. Schattschneider, E.E. *Politics, Pressure, and the Tariff*; Prentice Hall: New York, 1935.

8. Low, P. *Trading Free: The GATT and U. S. Trade Policy*; The Twentieth Century Fund: New York, 1993; 39–41.
9. Alic, J.A.; Conca, K.L.; Fletcher, W.W.; Hart, J.; Jarboe, K.P.; Miller, J.R.; Mogue, M.E.; Mundo, P.A.; Ostrom, D.R. *International Competition in Services*; Congress of the United States Office of Technology Assessment: Washington, DC, 1987; OTA-ITE-328.
10. Spero, J.E.; Hart, J.A. *The Politics of International Economic Relations*, 5th Ed.; St. Martin's Press: New York, 1997; pp. 76, 77, 86, 87.
11. Cohen, R.E.; Hegland, C. Farm bill winners and losers. *Natl. J.* **2002**, *34* (19), 1389–1390.
12. Mundo, P.A. *National Politics in a Global Economy: The Domestic Sources of U. S. Trade Policy*; Georgetown University Press: Washington DC, 1999; 151–189, 191–223 pp.
13. Dornbusch, R.W. Policy options for freer trade: the case for bilateralism. In *An American Trade Strategy*; Lawrence, R.Z., Schultze, C.L., Eds.; Brookings Institution: Washington DC, 1990; 125–134.
14. Lawrence, R.Z. Discussion. In *An American Trade Strategy*; Lawrence, R.Z., Schultze, C.L., Eds.; Brookings Institution: Washington DC, 1990; 136–138.
15. The original plan was to add Chile to NAFTA.
16. Tyson, L.D'A. *Who's Bashing Whom? Trade Conflict in High-Technology Industries*; Institute for International Economics: Washington DC, 1992; 85–154.
17. Lake, D.A. The state and American trade strategy in the pre-hegemonic era. In *The state and foreign economic policy*; Ikenberry, G.J., Lake, D.A., Mastanduno, M., Eds.; Cornell University Press: Ithaca, NY, 1988; 40–56.
18. Destler, I.M. *American Trade Politics*, 3rd Ed.; Institute of International Economics: Washington DC, 1995; pp. 11–38, 65–104, 122–124, 169.
19. Dryden, S. *Trade Warriors: USTR and the American Crusade for Free Trade*; Oxford University Press: New York, 1995.
20. Bauer, R.A.; de Sola Pool, I.; Dexter, L.A. *American Business and Public Policy: The Politics of Foreign Trade*; Atherton Press: New York, 1963.

Training and Development

Joan E. Pynes

*Department of Government and International Affairs, University of South Florida,
Tampa, Florida, U.S.A.*

INTRODUCTION

The demands placed on public organizations keep changing. Public agencies are threatened with budget cuts and reductions in staff, while citizens are requesting increases in the level of services and new services. The federal government is increasingly sharing responsibility with other levels of government, private companies, and non-profit organizations.^[1] Changes in technology are requiring new skills. Technology has taken on much of the workplace's mentally and physically repetitive tasks, so jobs today are requiring employees to assume more challenging responsibilities. For example, the downsizing of many public agencies has left them without the skills they need.

IMPORTANCE OF TRAINING AND DEVELOPMENT

Technology is being used to communicate with many people across large geographic areas, eliminating the need to be nearby for personal interactions. Demographics are changing; for example, there has been an increase in the number of racial and ethnic minorities employed in public agencies. Jobs have become less specialized, forcing employees to work in teams to deliver services, and productivity needs to be improved despite declining personnel and fiscal resources. Changes in goals, the purchase of new equipment, the enactment of new laws or regulations, and fluctuations in the economy are some of the other variables that influence change.

Many public organizations are facing the impending retirement of experienced public servants. By 2006, nearly 31% of the 24 federal agencies that fall under the Chief Financial Officer Act will be eligible to retire.^[2] State and local governments are facing the impending retirements of senior employees as well.

Change has become an inevitable part of organizational life. To remain viable, organizations must learn how to manage change. As the demands placed on public organizations keep changing, organizations must implement training and development activities to ensure that their staffs have the requisite knowledge, skills, abilities, and other characteristics (KSAOC) to confront these new challenges. The comptroller

general of the United States, David Walker, believes that government cannot maximize its resources and accountability without focusing on its most important asset, which is its employees.^[3]

Training can be targeted to help employees learn new job-specific skills, improve their performance, or change their attitudes. Changes need to be anticipated; training and development needs should be identified, planned for, and budgeted. Developing a comprehensive long-range training program requires a strategic human resources management (HRM) plan and recognition that employees are the most valuable resource. Agencies wishing to be viable must develop strategies to maximize their human capabilities. Training and development must be integrated into the core HRM functions.

Training and development have been defined as a planned effort by an organization to facilitate the learning of job-related behavior on the part of its employees.^[4] Training and development programs seek to change the skills, knowledge, and/or attitudes of employees. Programs may be focused on improving an individual's level of self-awareness, increasing an individual's competence in one or more areas of expertise, or increasing an individual's motivation to perform his or her job well. It is important to note the word "planned" in the above definition. Training and development efforts need to be thought out, and the following questions need to be answered:

1. How can you develop a comprehensive training plan to address the needs of managers, elected officials, support staff, direct service providers, and volunteers?
2. What methods can you use to assess your agency's training needs?
3. How can you design and implement the training program?

NEEDS ASSESSMENT

The first step in the training process is to determine the specific training needs. A need can be defined simply as the difference between what is currently being done and what needs to be done. This difference can be determined by conducting a needs assessment of the

skills and knowledge currently required by the position and those anticipated as necessary for the future. A needs assessment is critical to discerning whether performance deficiencies can be eliminated by training. Without a needs assessment, it is possible to design and implement a training program as the solution to a problem that is not related to a training deficiency.

For example, it comes to the attention of higher management that one department's productivity is very low. Higher management attributes the low productivity to a lack of motivation on the part of the employees. Employees are offered performance incentives and sent to training designed to inspire them. Despite the training, productivity is still low. The problem turned out not to be employee motivation but the first-line supervisors unable to plan and delegate assignments and prepare and budget equipment maintenance schedules.

A true needs assessment would have discovered that the problem was different than originally thought and that the solution may involve a different kind of training. In this case, the employees did not lack motivation; rather, it was the supervisors' skills that needed to be improved. Training the supervisors and not the employees would have been more appropriate. It is very important for a needs assessment to be accurate if training is to be successful.

Organizations can determine training needs through a variety of techniques. A strategic job analysis performed prior to the needs assessment is useful. The job analysis should identify the KSAOCs that incumbents need to effectively perform their jobs. Surveys and interviews with incumbents and supervisors; performance evaluations that identify performance deficiencies; criticisms or complaints from clients, staff, or personnel in agencies working with your employees; changes in regulations or operating procedures; and requests for additional training by incumbents can all provide clues to what training is needed.

DEVELOPING TRAINING OBJECTIVES

Training objectives are statements that specify the desired KSAOCs that employees will possess at the end of training. The objectives provide the standard for measuring what has been accomplished and for determining the level of accomplishment. For training objectives to be useful, they should be stated as specifically as possible. For example, learning objectives for software project management would state the following.

Participants in this class will learn to:

1. Identify the best software development methods.
2. Assure proper time is allocated for defining requirements.

3. Schedule realistically and avoid costly delays.
4. Estimate software size and development time frames.

The development of training objectives should be a collaborative process incorporating input from management, supervisors, workers, and trainers to ensure that the objectives are reasonable and realistic.

DEVELOPING THE CURRICULUM

After assessing the training needs and developing objectives, a training curriculum must be developed. Before developing the content and the manner of presenting the information, an analysis of the trainees must first be done. This step is crucial because the trainees often prescribe the kind of training that is likely to be effective. Some of the relevant issues to examine include the following:

- What are the participants' levels of education? For example, classroom instruction may be intimidating for employees with limited formal education.
- What are participants' expectations? Will all participants come to training with the same concerns?
- What are participants' knowledge levels, attitudes, and relationships with one another?
- Are participants prepared to receive technical instruction?
- Is the training voluntary or imposed from above?
- If the training is mandatory, will the participants be threatened by it?

The answers to these questions will provide some guidance to the trainers when developing the curriculum. For example, the Department of Defense uses digital game-based training for its young recruits, who grew up with computers and tend to get bored with traditional classroom instruction.^[5] The curriculum should provide the necessary information and be developed to maximize the imparting of KSAOCs.

One of the first decisions to be made is whether to provide on-the-job instruction or off-the-job classroom instruction or a combination of the two. On-the-job instruction takes place while the employee is actually working at the job site. It is usually provided by supervisors who instruct subordinates in the correct way to perform a task, such as filling out new purchase-order requisitions. Or a representative from the management information systems (MIS) department could demonstrate how to load and set up a new software package. On-the-job training is useful when employees are expected to become proficient in performing certain tasks or using equipment found at their work stations. Because the training is directly related to the

requirements of the job, transferring skills is easier. Employees learn by actually doing the job, and they get immediate feedback as to their proficiency. Other examples of on-the-job training include job rotation, in which employees move from job to job at planned intervals, either within their departments or across the organization. For example, many organizations train managers by placing them in different positions across the agency, so that they are provided with a comprehensive perspective of its operation.

Some KSAOCs are difficult to teach at the work site, so off-site training will be necessary. For example, training probation officers in counseling and listening skills would be difficult to do at their desks because other employees not involved in the training would be distracted by the instruction. Off-site training provides an alternative to on-the-job training. Employees receive training away from their work stations. In addition to avoiding disruptions to the normal routine at the job site, off-site training also permits the use of a greater variety of training techniques. Discussion of some other common training techniques follows.

COMMON TRAINING TECHNIQUES

Lecture

In a lecture format, a trainer presents material to a group of trainees. Lectures have been criticized because the information in them flows in only one direction—from trainer to trainees. The trainees tend to be passive participants. Differences in the trainees' experiences, interests, expertise, and personalities are ignored. Lectures are limited to the transfer of cognitive material. Lectures are beneficial when they are used to introduce new information or provide oral directions for learning tasks that eventually will be developed through other techniques. Lectures are readily adaptable for use with other training techniques.^[4]

Experiential Exercises

Experiential exercises attempt to simulate actual job or work experiences. Learning can be facilitated without the cost and risks of making mistakes while actually on the job. For example, law enforcement agencies use experiential exercises to train officers in emergency and disaster planning.

Role Playing

Role playing gives trainees the opportunity to practice interpersonal and communication skills by applying them to life-like situations. Participants are expected

to act out the roles they would play in responding to specific problems that they may encounter in their jobs. Role playing can be used in a variety of contexts. Law enforcement academies use it when training officers how to interview crime victims, such as sexually abused children or witnesses to a crime. Role playing is frequently used in supervisory training, in which participants are asked to counsel a subordinate who is suspected of having a substance-abuse problem.

Case Studies

The use of case studies in training involves having participants analyze situations, identify problems, and offer solutions. Trainees are presented with a written description of a problem. After reading the case, they diagnose the underlying issues and decide what should be done. Then, as a group, the trainees discuss their interpretations and understanding of the issues and the proposed solutions.

Audiovisual Methods

Videos are often used for training in a variety of contexts. There are videos to educate employees on legal topics such as sexual harassment, hiring disabled applicants, and using progressive discipline; videos that focus on interpersonal and communication skills; and videos that simulate a grievance arbitration hearing, permitting trainees to view the process, hear witnesses testify, see the behavior of management and union representatives, and learn how arbitrators conduct proceedings. Videos can be used to demonstrate particular tasks, such as the procedures to follow when apprehending a suspect or extinguishing a chemical fire. Videos are often used in orientation sessions to present background information on the agency—its history, purpose, and goals. This use eliminates the need for trainers or supervisors to repeat themselves for all new employees and ensures that the same information is always presented.

Trainees may also be videotaped. They may be asked to make a presentation or to provide performance feedback to colleagues. They may then view the videotape to identify their strengths and weaknesses related to the topic.

An advantage of video is that it provides the opportunity to slow down, speed up, or stop the video to review specific activities and enable questions to be asked and answered. A disadvantage to the use of videos is that they can be expensive to purchase or make.

Equipment Simulators

Anybody who has taken a driver's education course probably remembers the driving simulator that

replicated a car's dashboard, and gas and brake pedals. Simulators are used to bring realism to training situations. For many jobs, such as law enforcement, on-the-job training can be too dangerous, such as training police officers when to discharge firearms. So equipment and scenarios that replicate the shadows and noises of alleys are used to train police officers not to over-react. Fire departments use burn buildings, which are designed to withstand repeated fires, to give firefighters opportunities to practice rescue attempts while battling heat and smoke. It would be too expensive or dangerous to burn vacant or decayed buildings. With the use of simulators, new procedures can be attempted without the risk of endangering human lives.

Videoconferencing

Improvements in technology have made videoconferencing less expensive and more accessible to an increasing number of public and non-profit organizations. Two-way video cameras and high-speed networks are able to transport interactive live images across large geographic distances. For example, when training dollars became tight and it could not afford to send individual employees for training, the City of Phoenix contracted for a satellite and videotape service to train employees in its MIS department. New York City's information technology (IT) and personnel departments worked together to establish a video training network, using a channel of the city's subscriber cable system coupled with fiber-optic links, to provide closed-circuit audio and video service to city agencies. A pilot program was designed to teach fire academy drills at 55 department locations.^[6] The advantages of video training include more flexibility in meeting training needs, a reduction in time lost in travel to training sites, and increased uniformity of training.

Community Resources

Public agencies should be aware of the many community resources that are often available to provide training for nominal costs or even for free. Many health-care facilities offer workshops on topics that are targeted to specific issues, such as adolescent or elderly depression, sexual abuse, or substance abuse.

Programmed Instruction and Computer-Based Training

Programmed and computer-based instructions are self-teaching methods designed to enable trainees to learn at their own pace.

The International City/County Management Association (ICMA) provides a variety of professional development resources. It offers self-study courses, training packages, group courses, multimedia training, handbooks, and workbooks. One of its training packages is titled "Measurement for Results: Implementing Performance Measures in Local Government." This training package is designed to help local governments develop a set of performance measures for managing, assessing, and improving local government programs. Students are assigned readings and are then asked to complete written assignments covering the material learned in the lessons, to review questions based on the material, and to carry out a series of activities that apply what they have read.

Computer-based training uses interactive exercises with computers to impart job skills. Training materials are on compact disks. Employees read information, instructions, and diagrams or other graphics on the computer screen and then respond accordingly. The ICMA offers multimedia training in ethics. *Ethos: Multimedia Ethics Training for Local Governments* CD-ROM is available to purchase. Trainees explore and decide how to resolve 21 real-life ethics scenarios in this interactive training program.

Within the last few years, web-based training has become very popular. For many employers and employees, it is convenient and typically less expensive than some other training methods. Sending an employee to a classroom session can be expensive, depending on tuition, travel, hotel, and per diem expenses.

Web-based training can be synchronous or asynchronous. Because the courses can be accessed through a computer with an Internet connection, employees can even receive training at home.

At the federal level, e-learning is facilitated by the Gov Online Learning Center—GoLearn, government-wide resources that support the delivery of the federal workforce through simplified and one-stop access to high-quality e-training products and services. The site acts as a virtual campus and offers free and fee-based courses.

In Washington State, the state Department of Personnel has launched the E-Learning Network (ELN). ELN includes approximately 1200 off-the-shelf courses in IT skills and certification as well as topics such as accounting, sexual harassment and diversity; it serves as a centralized provider for 83 state agencies and more than 70,000 employees.

In Illinois, the state's upward mobility program trains people for positions that the state has trouble filling because of high competition in IT positions from the private sector. It received a 3 yr grant from the Alfred P. Sloan Foundation for the development of asynchronous online curriculum for IT job titles. The state worked with Lincoln Land Community College to put the courses into an online format.^[7]

The American Society for Training and Development found that employees were less likely to take advantage of web-based training if the training is only offered at their desks. There are more distractions there and people are also used to going to a classroom to learn. People like to go to a training center to do the learning even if that learning is online.^[8]

Not everything is appropriate for web-based training. Skills such as how to work effectively as a team member or how to make presentations are better taught in a class where participants have to work together and where employees actually give presentations and receive live feedback. Web-based instruction allows trainers to provide background information and theories, but the skills are best developed through interpersonal interactions. Each technique has advantages and disadvantages, which need to be weighed in relation to time constraints, staff resources, finances, and desired outcomes.

DELIVERING TRAINING

Other issues must be addressed in addition to curriculum. Should the training take place for short periods of time spread over many days, or should it encompass long periods over fewer days? What time of day should the training take place? What size group should be involved? No one right answer exists. It depends on the information being presented or the skills that need to be taught, as well as the aptitudes of the participants and the techniques that are used. Failing to consider any of these factors can negatively influence the results of training efforts. Most public-sector employers cannot afford such waste.

The delivery of the training program is the stage where the trainers and the participants converge. At a well-organized work site, the employees selected for training understand what the objectives of the training are and what they can expect.

It is up to the trainers to create a climate in which individual learning styles are recognized and considered in the delivery of the content. This is especially important for employees who resist training or perceive it as a punishment and not as an opportunity.

Sims suggests that if public-sector trainers understood how individuals learn, they might be able to design programs that enhance learning and that, therefore, are more effective.^[9] Training programs must address not only substantive content or material but also how people learn and, therefore, incorporate different learning strategies.^[10]

As the workforce becomes more diverse, there will be more variation in employees' ability to learn, in their learning styles, in their basic literacy skills, and in their functional life skills. It will be even more

important for training to take into account individual backgrounds and needs. Organizations and trainers will also need to recognize that adult employees learn differently than children. Adult learners see themselves as self-directed and expect to be able to answer part of their questions on the basis of their own experiences.^[11] Instruction tailored to adult learners allows the trainees the opportunity to participate in the process. The role of the training instructors is to facilitate learning; they use lots of questions, guide the trainees, and encourage two-way communication between the instructor and the class, as well as communication among the class's participants.

EVALUATING TRAINING

An evaluation of the training program is necessary to determine whether the training accomplished its objectives. Unfortunately, this is often the most neglected aspect of training, especially in the public sector. Evaluation improves training programs by providing feedback to the trainers, participants, and managers, and it assesses employee skill levels.

Researchers suggest that there are four primary levels at which training programs can be evaluated.^[9,12] The first level is measuring the participants' reactions to the training program. Kirkpatrick refers to this step as a measure of customer satisfaction.^[12] Participants are asked to answer questions such as: Was the trainer knowledgeable? Was the material/information relevant? Will the information they learned assist them in performing their jobs? Data are gathered through the use of surveys distributed at the conclusion of the training session. Asking these questions provides data on the training program's content and the trainer's skill.

Learning has taken place when one or more of the following occurs: Attitudes are changed; knowledge is increased; and skills are improved. The second level of evaluation measures whether learning has occurred as a result of attending the training. Did the participants acquire the skills or knowledge that was embodied in the objectives? Did the training impart the KSAOCs that were deemed important? To determine whether learning took place, participants can be tested on the information presented, follow-up interviews can be conducted, skill demonstrations can be required, or case studies can be developed, which test the competencies that were intended to be taught.^[12] Three to six months after training, the participants can be asked how they would handle situations provided in the case studies if they arose as part of the job.^[13] Supervisors can evaluate the participants on their use of the learned competencies to solve the work-related problems. It is important to note that the methods used should be selected on the basis of the level of mastery desired.

The third level of evaluation attempts to measure the extent to which on-the-job behavioral change has occurred because the participants attended the training program. Evaluation activities are aimed at determining whether the participants have been able to transfer to their jobs the KSAOCs they learned in training. Measurement at this stage is more difficult; it requires supervisors to collect work samples or observe employees' performance. Another technique would be to utilize performance evaluations designed to measure the new competencies.

For change to occur, four conditions are necessary: 1) the employee must have a desire to change; 2) the employee must know what to do and how to do it; 3) the employee must work in the right climate; and 4) the employee must be rewarded for changing. Kirkpatrick notes that a training program can accomplish the first two requirements, but the right climate is dependent on the employee's immediate supervisor. Some supervisors may prevent their employees from doing what was taught in the training program; others may not model the behaviors taught in the training program, which discourages the employees from changing; some supervisors may ignore the fact that employees have attended the training program, and thereby not support employees' efforts to change; others may encourage employees to learn and apply their learning on the job; and finally some supervisors may know what the employees learned in the training and make sure that the learning transfers to the job.^[12]

The fourth condition that the employee must be rewarded for changing can include the feelings of satisfaction, achievement, proficiency, and pride that can occur with change. Extrinsic rewards such as praise from the supervisor, recognition from others, and possible merit rewards or promotions can also result.

The fourth level of evaluation attempts to measure the final results that occurred because employees attended the training. Ideally, training is linked to improved organizational performance. At this level, evaluation is concerned with determining what impact the training has had on the agency. Satisfactory final results can include such things as fewer grievances filed against supervisors, greater employee productivity, a reduction in the number of client complaints, a decrease in workplace accidents, increased dollars raised through fundraising, improved board relations, and less discrimination in the workplace. Some final results will be easier to measure than others. For example, the dollars raised from fundraising activities, the number of workplace accidents, or the number of grievances filed can be easily quantified and compared to times before the training. Other final results, like eliminating discrimination, changing attitudes, and improving leadership and communication are less tangible and more difficult to measure. Such results

will have to be measured in terms of improved morale and attitudes.^[12]

As a final step, organizations must determine whether the benefits of the training outweigh its direct and indirect costs. Examples of direct costs include expenses for instructor fees, facilities, printed materials, and meals. Indirect costs include the salaries of participants who are away from their regular jobs. Has there been a reasonable return on this investment? Basically, was the training worth its costs? Did it accomplish what it was designed to accomplish?

The potential benefits from evaluating training programs include improved accountability and cost effectiveness for training programs, improved program effectiveness (Are programs producing the intended results?), improved efficiency (Are they producing the intended results with a minimum waste of resources?), and information on how to redesign current or future programs.

CAREER DEVELOPMENT

Training has been defined as "the acquisition of knowledge and skills for present tasks, which help individuals contribute to the organization in their current positions. . . . To be successful, training must result in a change in behavior, such as the use of new knowledge and skills on the job."^[13] Career development, however, provides the employee with knowledge and skills that are intended to be used in the future. The purpose of career development is to prepare employees to meet future agency needs, thereby ensuring the organization's survival.^[14]

Career development is used to improve the skill levels of and to provide long-term opportunities for the organization's present workforce. Career development programs provide incumbents with advancement opportunities within the organization, so that they will not have to look elsewhere. Federal managers have found that training opportunities, challenging jobs, and flexible work environments are more successful in attracting workers with IT skills than higher pay.^[15,16] Taking the time and spending resources to develop employees signal to them that they are valued by the agency. As a result, they become motivated and assume responsibility for developing their career paths.

The focus of career development plans is on where the agency is headed and where in the agency incumbents can find future job opportunities. Employees and supervisors should produce a development plan that focuses on employee growth and development. The plan should have measurable development objectives and an action plan. For example, supervisors should review their employees' skills with the job descriptions of higher-level positions within the same

job family or of positions within the organization to which the employee might be able to cross over. By comparing employees' skills with the skill requirements of other positions, the employees and supervisors can determine what experience and training might still be needed for advancement or lateral movement. Supervisors should direct employees to relevant training opportunities and, when possible, delegate additional tasks and responsibilities to employees so that they may develop new competencies.

A number of career development programs can be found in the public sector. Some of them focus on moving employees from clerical or paraprofessional positions into higher-paying administrative jobs. Others focus on developing supervisory and management skills.

MANAGERIAL AND EXECUTIVE DEVELOPMENT

Problem solving and initiative are examples of some of the critical skills managers and executives of public agencies need. Technical experience and competence are no longer enough; public organizations need leaders with the vision to direct and guide their agencies.

To assist in the development of these skills, many state HRM departments and the federal government have developed career development programs, to impart the types of competencies that leaders need to meet the challenges faced in a constantly changing environment. The Michigan Department of Civil Service has identified and provides training on the following competencies: adaptability; aligning performance for success; developing a successful team; building partnerships; building trust; communication and customer focus; decision making; delegating responsibility; facilitating change; innovation; leading through vision and values; planning and organizing work; and strategic planning.

Strategies available to provide training and development opportunities include assigning employees to various jobs; providing on-the job coaching; allowing inexperienced employees to shadow another person around to observe what they do, this can be helpful in learning about a particular area or function; expanding present responsibilities to include a wider variety of assignments and duties; and assigning employees to department or agency taskforces to make them aware of complex issues or issues beyond their usual scope of responsibilities.^[17]

CONCLUSION

Training is typically associated with improving the performance, knowledge, and/or skills of employees in their present positions. Career development is viewed as a continuous process consisting of evaluating

abilities and interests, establishing career goals, and planning developmental activities that relate to the employees' and organization's future needs. Public organizations must recognize the importance of both training and career development planning.

Agencies that are serious about training and career development should continue to monitor, evaluate, and revise their training and career development programs. New approaches to training need to be considered, and organizational reward structures should encourage individual growth and development that benefit both the employee and the organization.

REFERENCES

1. Kettle, D. What's next? Gov. Exec. Mag. **2001**, <http://www.govexec.com/features/0101/0101sl.htm> Gov Exec.com [WWW document] URL.
2. U.S. General Accounting Office. *Federal Employee Retirements: Expected Increase over the Next 5 Years Illustrates Need for Workforce Planning*; Washington, DC, 2001; GAO-01-509.
3. Weinstock, M. Virtual learning. Gov. Exec. Mag. **2000**, <http://www.govexec.com/features/10000/1000s3.htm> Gov Exec.com [WWW document]. URL.
4. Wexley, K.N.; Latham, G.P. *Developing and Training Human Resources in Organizations*; Harper Collins: New York, 1991.
5. Lippman, H. Top marks for employee training. Bus. Health **2001**, 19 (3), 52.
6. Cranford, J. A guide to award-winning technology. Governing **1995**, 8, 61–70.
7. Wilson, C. Training & Development: E-learning States, federal government lead the way with e-learning initiatives. IPMA-HR News **2004**, 70 (1), 1, 3, 5–7.
8. Zimmerman, E. Better training is just a click away. Workforce **2001**, 80 (1), 36–42.
9. Sims, R.R. The enhancement of learning in public sector training programs. Public Pers. Manage. **1993**, 22, 243–255.
10. Milano, M.; Ullius, D. *Designing Powerful Training: The Sequential-Iterative Model*; Jossey-Bass: San Francisco, 1998.
11. Laird, D. *Approaches to Training and Development*; Addison-Wesley: Massachusetts, 1985.
12. Kirkpatrick, D.L. *Evaluating Training Programs: The Four Levels*, 2nd Ed.; Berrett-Koehler: San Francisco, 1998.
13. Erickson, P.R. Evaluating training results. Train. Dev. J. **1990**, 44, 59–65.
14. Fitzgerald, W. Training versus development. Train. Dev. J. **1992**, 5, 81–84.
15. Robb, K. Other perks outweigh salary for IT recruits. **2001**, <http://www.federaltimes.com/jobs/car032601.html> Federal Times.com, [WWW document]. URL.
16. U.S. General Accounting Office. *Results-Oriented Cultures: Modern Performance Management Systems Are Needed to Effectively Support Pay for Performance*; Washington, DC, 2003; GAO-03-612T.
17. Ibarra, P. How to get better all the time through training and development. IPMA-HR News **2005**, 71 (2), 17, 19.

Training for Productivity

Montgomery Van Wart

Department of Public Administration, California State University-San Bernardino,
San Bernardino, California, U.S.A.

INTRODUCTION

There are four challenges for organizations that want to use training and development for greater productivity. First, training and development must be used strategically, which in turn, often means a strategic overhaul of many training systems today. Training systems must be assessed, both for what they can do and what they should do in a robust organization bent on producing both routine operational quality as well as flexibility and innovation in operations and programs. Second, because almost all the reform initiatives being undertaken require more decentralization, empowerment, job expansion, and power sharing, more broadly trained employees are necessary. Technical skills are no less important; however, advanced general skills (e.g., skills formerly largely restricted to supervisors and systems analysts, but now devolved to line workers) generally need greater emphasis. This occurred initially with the wave of total quality management programs in the early and mid-1990s, but has fallen off dramatically. Third, the more strategic use and design of training systems, and a broader training agenda, cannot detract from a strong emphasis on training basics. Organizations and training systems may change over time, but human nature and how people learn does not. Core learning and teaching principles must be constantly reinforced. Untutored trainers, more often than not, are poor trainers. Finally, there must be an emphasis on curriculum and instructional design as training programs and systems are rebuilt. This is particularly true with new stress on self-training, certification systems, higher levels of accountability, and immense changes in technology that are often poorly understood and thus badly used.

With these four challenges in mind, we now turn to a brief review of organizational learning needs, organizational competency clusters and training types, individual learning needs, and individual training needs. In keeping with a productivity orientation, we begin with organizational needs—that is, strategic needs—and move to individual needs, which are covered more briefly.

ORGANIZATIONAL LEARNING NEEDS

As always with strategic planning, the assessment phase is the most important. Without the right questions

and good data, how can good decisions be made? Five questions should be asked with skepticism because ongoing systems frequently “make do” for such a long time that more robust questioning is unaccustomed.

What is the quality of training? This is a technical question and needs a technical analysis.

How appropriate is the training that is being offered? Even more than most systems, training must be constantly updated and upgraded if it is not to be irrelevant and wasteful.

Are the resources and support adequate for the training system? Needless to say, all systems can use more resources in their quest for improvement or excellence, but the chronic underfunding of development activities in the public sector with its heavy personnel-driven workload makes this question pressing.

Is the suitability and substitutability of training properly balanced? For any given problem, training may be the correct, partial, or incorrect response.

Does the training system have the requisite comprehensiveness or complexity for its current stage of development? Answering this question requires a big picture perspective. Are different types of training available both for organizational and individual learning needs (addressed immediately below)?

One way of assessing the training system in a strategic analysis, especially the issue of comprehensiveness, is to look at tiers of learning that may be met in an ascending order of learning needs. The two tiers are basic and advanced learning.

Basic learning occurs primarily by acquiring basic knowledge and skills, and dominates much of the literature and our thinking about training and development. It can be further divided into the routine level and the professional level. New initiates to first-tier learning may be thought of as the young child, the first-year student, the preservice Master's in Public Administration student, the new trainee, or inexperienced user of technology. Still in the basic learning tier but at the high-performing end of the continuum are those at the professional level who are seen, after long periods of hands-on learning and experience, as craftspeople, masters, and operational experts.

We learn by acquiring knowledge and practicing skills throughout our lives. In the world of work, this is especially true of the vast areas of technical material that we must routinely master. Technical material includes the content area in which we specialize—such as law enforcement, social welfare, health, or land management—as well as the processes by which we apply that knowledge—such as laws, regulations, organizational policies, or job requirements. Technical information doubles every 7 yr according to many experts and is clearly increasing at a faster pace. Today, unlike the past in which most basic learning was completed by the end of formal schooling, we must continue to learn throughout our lives if we are not to become obsolete.

The differences between routine training and professional training (also commonly referred to as continuing education) are a matter of sophistication rather than a matter of kind. Experts still acquire knowledge and refine skills, but at a higher level of complexity and intricacy. Public agencies have many programs for mature professionals who are reviewing bodies of knowledge and practice areas for up-to-date information, detailed explication of rules and policies, and cutting-edge solutions to operational issues. Indeed, continual learning needs to take place at this level for a lifetime, without even considering advanced learning. Yet, despite its fundamental importance and the new emphasis on lifelong learning, this is widely bemoaned as an area that is weaker than it should be.

The basic tier of learning, including both the routine and the professional levels, is roughly what Argyris and Schön^[1] called single-loop learning or what Senge^[2] referred to as personal mastery. Without knowledge and skills we cannot be experts, and to give them their due, public bureaucracies generally excel at acquiring expertise.^[3] The next tier, *advanced learning*, goes beyond simple knowledge and skill acquisition to involve learning that expands, creates, and transmits knowledge.^[4] There are many images used in second-tier learning, such as perfecters, improvers, adapters, and innovators, and at a still more high-performing level, artists, inventors, applied researchers, and scientists. The two levels within this tier can be thought of roughly as problem solvers at the simpler level and as systems designers at the more sophisticated level. This tier of learning involves four of Senge's five disciplines for learning organizations, and is what Argyris and Schön referred to as “double-loop” learning. The five types of learning included in this advanced tier are learning by sharing, comparing, systems thinking, competing, and suspending disbelief.^[5]

- *Learning by sharing* is particularly powerful in the affective domain, which is often the gateway to the

other two domains of learning—cognitive and psychomotor. Affective learning is where we learn to believe in and value some things as opposed to others, and therefore is generally a key to motivation. Learning by sharing takes a number of forms, including teaching, discussing, and teaming. Learning by teaching forces people into more advanced modes because it requires teachers to be clear, logical, and organized; to practice themselves and anticipate problems; to respond to questions; to see things through others' eyes; and even, as the old saying goes, to know things backward and forward. Learning by discussing is one of the most important forms of advanced learning because it leads to the discovery of non-routine problems, anomalies, contradictory perspectives, hidden assumptions, and complementary ideas. Learning by teaming was introduced as an antidote to excessive hierarchical tendencies in organizations in the 1980s, and became popular in the 1990s.

- *Learning by comparing* is the basis of the scientific method that compares the past with the present (various types of time series), the “normal” with deviations and anomalies and, in some restricted settings, control groups with experimental groups. From these comparisons come hypotheses about the relationships of phenomena, which subsequently can be proven, disproved, or modified by further systematic observation. Given the tremendous success of the scientific method in modern history, there can be little doubt of its power. Learning by systematic examination of past experience refers to one's own experience, whether that is defined as the person, group, or organization. Learning from systematic observation of others' experience is commonly called benchmarking in the management literature. Learning by experimenting is indeed a robust form of comparison, but it is not always possible. Individuals/organizations perform experiments by changing a practice and measuring the change, such as a change in an intake procedure or technology, with both experimental and control groups.
- Systems thinking is the hallmark of advanced learning for Peter Senge who makes this “discipline” the one that fuses all the other advanced forms of learning “into a coherent body of theory and practice.”^[2] Systems thinking is the antidote to tunnel vision, turfism, stove-piping, and other bureaupathologies that accrue from overspecialization of function and responsibility. Despite the contributions of specialization in the past, contemporary organizational learning finds it more often than not a major inhibitor.
- The ability to learn through competition is captured by expression “necessity is the mother of

invention,” and through the underlying capitalistic belief that competition causes innovation and creativity to flourish in the long run. What Rosabeth Moss Kanter noted for the last decade still holds true, “the question most appropriate for the 1990s is not *what* the competitive world organization should look like but *how* to become one.”^[6] Competition can lead individuals to learn by requiring them to determine what they want to achieve, by driving them toward that achievement or level of excellence, and by creating both incentives for succeeding and disincentives for floundering. Competition itself can be broken down into numerous component parts, including goal setting, striving and risk taking, and learning through failures.

- Of all the modes of advanced learning, none is more commonly discussed, but more difficult to do, than suspension of disbelief. This mode of learning requires the ability to suspend current beliefs about the proper way to do something, the proper way to value, or the nature of truth. We unquestioningly use habits, routines, and belief systems most of the time or we would need to go back to first principles every time. Yet, some of the greatest learning is by seeing the world afresh, discarding (at least temporarily) our mental models, and trying out new ideas, perceptions, and beliefs. Most great scientists and inventors are masters at this type of learning.

ORGANIZATIONAL COMPETENCY CLUSTERS AND TRAINING TYPES

Organizations have various types of functions they must perform that translate into different types of training.^[7] There are three metacompetency clusters: technical training, general skills training, and management training. No strategic analysis of an organization would be complete without an assessment of the strengths and weaknesses of the various types of training, or in some cases, an assessment of the types of training that are altogether absent.

Three types of technical training are identified here: procedural, mechanical, and professional. Procedural training focuses on laws, codes, rules, or procedures that are required for compliance with agency mandates, and for coordination and flow of work. It is sometimes called operations or administrative training. The most common type of procedural training is new employee orientation, which focuses on employee benefits and agency policies. Some of the other common types of procedural training cover topics such as agency overviews, personnel procedures, selection/hiring procedures, and procurement processes.

Mechanical training (also known as mechanical-technical training) focuses on how things operate, how they are built, how they can be fixed when broken, and how they can be maintained. Manual or physical skills are a large part of the training. One of the most common types of mechanical training is training on computers, such as learning word processing programs, graphics packages, spreadsheets, and customized data systems for the organization. Mechanical training is a frequent part of other types of training, and can be quite sophisticated in and of itself (e.g., pilot training).

Professional-technical training focuses on the select knowledge, skills, and abilities needed by practicing professionals in performing their jobs in the public sector. Because professional workers apply principles from mathematics, physical and natural sciences, law, social sciences, administrative science, education, political science, and so on, their training is inherently more theory and principle driven than are most types of training.

Although technical skills focus on the specific knowledge and skills required to perform a specific job, general skills focus on the skills that are needed to perform any or many jobs in organizational life. Most employees come to the organization with adequate basic skills; however, this assumption must be verified, especially for lower-paying front-line positions, some physical jobs, or when a significant number of employees do not have English as a first language. Basic skills include not only the classic “reading, writing, and arithmetic,” but also basic listening and speaking skills and even some basic advanced skills, such as speaking in public and meeting skills. Advanced general skills training focuses on making employees more effective outside their roles as subject-matter specialists. General skills training aims to improve employees’ ability to learn, work with others, adapt, and be productive. General skills training includes courses on learning to learn, interpersonal skills, negotiation, teamwork, problem solving, creativity, goal setting, self-esteem, customer relations, and basic management tools [e.g., flowcharts, PERT charts (performance, evaluation, and review technique), statistical process control, and check sheets].

The last metacluster is management training. Managers at different levels of the organization tend to have different responsibilities and functions and need different competencies,^[8] therefore, this cluster is divided into supervisory, (middle) management, and executive training. Supervisory skills training focuses on supervisors’ direct interaction with subordinates and on getting work done by, with, or through other people. Specific interpersonal skills are one target and so are clearly identifiable personnel processes.

Common topics for supervisory programs are how to evaluate employees, how to conduct selection/hiring interviews, and motivating people.

Management training focuses on complex interpersonal skills (e.g., labor relations) and inter-related group processes (e.g., organizational climate) in an attempt to find organizational solutions within a changing and complicated environment. The focus shifts from individuals to groups and from problems of a unit to the management of programs. There is often a focus on the improvement of general analytical skills (as opposed to the specific analytical skills of technical professionals) and the ability to make balanced judgments based on various data sources. Management training has numerous standard topics, including administrative leadership, workforce planning, labor relations, and project management.

Executive development training is the most conceptual, broadest, and most externally oriented type of training. Programs for executives tend to focus on the role of the organization in the public sector environment, and facilitate the executives' skills in coping with external opportunities and threats and public relations. Typical courses for executive programs focus on media relations, public speaking and contact skills, leadership and organization assessment (generally with databases collected specifically for the training), and strategic planning.

INDIVIDUAL LEARNING NEEDS

Those responsible for designing and implementing learning in the more formal aspects must pay attention to at least three major learning inputs: trainee characteristics, work environment, and learning principles. Trainee characteristics include the knowledge, skills, abilities, and motivation that trainees bring to the training situation. These characteristics are important because they affect the need or lack of need for training, the design of the training, and the outputs of training. Without an assessment of the training needs of the targeted group, it is easy to make false assumptions in designing training. This is nearly as important is the work environment from which trainees come and to which they will return. The work environment affects training in two important ways: 1) through the support that comes from the work environment for training and development activities and 2) through the opportunity given to the trainees to apply new skills and knowledge gained from the training.

In the actual design and implementation of a training program, seven general principles are important to keep in mind to maximize the effect and efficiency of

training. In substantial training programs, the omission of even one of these principles can lead to dismal results in many cases:

- The first principle is the need for *goal setting*, which is an important way to focus and enhance motivation. One level of goals includes those that bring the employees into training. Another level of goals is employed in the training itself, often expressed as desired learning objectives.
- A second need that is important and significantly more central in training settings than in educational settings is *increasing the similarity of training to the work environment*. Three progressive steps increase the similarity of training to the work environment: the use of examples, modeling, and simulation. The use of examples is a cognitive technique that includes simple allusions to applications to full-blown discussions of relevant cases. Modeling is a visual or graphic technique that actually shows the correct behavior to the trainees. Simulations are opportunities for trainees themselves to perform the behaviors.
- The third training principle is *using underlying principles*. Underlying principles emphasize universal or generalizable aspects and enhance transfer of learning from one training situation to another. Therefore, the use of underlying principles is extremely pronounced in educational venues because of the longer durability of the training. However, underlying principles do not, by themselves, teach the specifics that people need to anchor their learning and their more concrete needs. Therefore, it is easy to overemphasize underlying principles to the detriment of the learning endeavor.
- A fourth, surprisingly complex principle, is *increasing the organization of the material*. A brilliant lecture may dazzle listeners, but little may be retained the next day when questions are asked of them, unless the structure of the lecture is readily apparent and reinforced. Neither is organization of the material solely determined by logical structure or cognitive attributes. In fact, one of the most common mistakes is excessive cognitive emphasis at the expense of opportunities to rehearse or question.
- Although *active involvement of learners* in the learning process has been re-emphasized in modern research, it is actually an ancient learning principle. The Socratic method teaches by using questions to discover truths. It requires the learner to constantly question and probe their experiences and understanding. Whether active involvement is stressed as a part of the initial instruction or is stressed separately after instruction, it is basically practice.

- Related to, but separate from, learner involvement is *feedback*. The term feedback refers to the knowledge of results of practice. If there is no instructor-monitored practice, there can be no external feedback. Practice includes exercises (with answers or oversight), tests, discussions, papers, and in-training observation of behavior of behavior in simulation or apprenticeships.
- The final training principle is *using various techniques and stimuli*. By varying techniques, a trainer can engage several of the learners' senses. For example, if something is heard, seen, and written down (the tactile sense), the chances of retention increase greatly. Further, if the learner has the opportunity to verbalize the material, the chances of remembering are at their greatest.

INDIVIDUAL TRAINING NEEDS

Only two issues regarding training methods are reviewed here briefly: some of the basics of good curriculum design, and some of the strengths and weaknesses of instructional families.

Although full-blown lesson plans may not always be required, anyone designing or evaluating any training needs to know the basics of curriculum design. In a training context, curriculum design is how material is organized to maximize learning. There are four simple elements, all of which are critical for effective curriculum design where robust learning is important. First, what are the learning objectives? Second, how will the learning occur (what will the didactic inputs be)? Third, how will the learners practice the information? Finally, what are the performance measures that will be used and to what standard?

Instructional design looks at the specific methods that are used in the didactic input, practice, and performance measurement phases of curriculum design. Different instructional families tend to emphasize different learning principles, so although there is an endless variety of good configurations for good instructional design, there is also an endless variety of bad configurations. Lecture methods include the standard lecture, team teaching, guest speakers, panels, and student presentations. As a family, they are strong at teaching underlying principles and increasing the organization of the material. Discussion methods include question-and-answer techniques, large-group methods, small-group methods, case studies, debates, and individual conferences. Strictly by themselves, they tend to be strongest at setting goals, actively involving learners, and giving feedback. Printed materials include auxiliary reading materials, training manuals, and programmed instruction. Their strengths and

weaknesses tend to mimic those of lecture methods; they are strong in teaching underlying principles and increasing the organization of the material and weak at setting goals, work similarity, involvement, and feedback. Practice and feedback techniques include note taking, adjunct questions, individual exercises, demonstrations, role plays, simulations, survey and self-assessment techniques, tests, and site visits. Their strengths are generally in their similarity to the work environment, involvement, and feedback. Behavior-shaping methods include coaching, apprenticeships, job rotation, self-training and embedded training, and counseling. These methods have as their strengths setting goals, increasing similarity to the work environment, active involvement, and feedback. Technology-based techniques are a category composed of audiotapes, slides, videotapes and films, computer-based instruction, teleconferencing, and various high-end technology uses, such as optical discs. They tend to be strong at teaching underlying principles and increasing the organization of the material, and sometimes in increasing the similarity to the work environment.

CONCLUSION

Training and development for productivity requires a relatively tough-minded approach that begins with an organizational perspective of learning needs and competency clusters and only then moves to individual needs. It requires asking the tough questions about not only what is desirable to achieve from an organizational perspective, but about the standards that are necessary and who will be accountable, including the trainees and trainers.

REFERENCES

1. Argyris, C.; Schön, D.A. *Organizational Learning*; Addison-Wesley: Reading, MA, 1978.
2. Senge, P. *The Fifth Discipline*; Doubleday: New York, 1990; 12 pp.
3. Weber, M. Bureaucracy. In *Max Weber: Essays in Sociology*; Gerth, H.H., Mills, C.W., Eds.; Oxford University Press: New York, 1946.
4. Garvin, D.A. Building a learning organization. *Harvard Bus. Rev.* **1993**, *71* (4), 73–91.
5. Van Wart, M. Learning in non-hierarchical bureaucracies: the new learning profile Korean Rev. *Public Admin.* **1997**, *2* (2), 117–147.
6. Kanter, R.M.; Stein, B.A.; Jick, T.D. *The Challenge of Organizational Change*; Free Press: New York, 1992.
7. Van Wart, M.; Cayer, N.J.; Cook, S. *Handbook of Training and Development in the Public Sector*; Jossey-Bass: San Francisco, 1993.

8. Katz, R.L. Skills of an effective administrator. *Harvard Bus. Rev.* **1955**, 33 (1), 33–42.

BIBLIOGRAPHY

1. Bane, K.D.; Kopelman, R.E.; Olivero, G. Executive coaching as a transfer of training tool: effects on productivity in a public agency. *Public Personnel Manage.* **1997**, 26 (4), 461–469.
2. Bartel, A. *Formal Employee Training Programs and Their Impact on Labor Productivity: Evidence from a Human Resource Survey*; National Bureau of Economic Research: Cambridge, 1989.
3. Docherty, P.; Shani, A.B. *Learning by Design: Building Sustainable Organizations*; Blackwell Publishing: Oxford, England, 2003.
4. Herriot, P. *Career Management Challenge: Balancing Individual and Organizational Needs*; Sage: Thousand Oaks, CA, 1992.
5. Piskurich, G.M. *The ASTD Handbook of Training and Delivery: A Comprehensive Guide to Creating and Delivering Training Programs, Instructor-Led, Computer-Based, or Self-Directed*; McGraw-Hill Professional: New York, 2000.
6. Prais, S.J. *Productivity, Education, and Training: An International Perspective*; Cambridge University Press: Cambridge, England, 1995.

Transparency and Corruption in Southeast Asia

Habib Zafarullah

*School of Social Science, University of New England, Armidale,
New South Wales, Australia*

Noore Alam Siddiquee

*Department of Political Science, International Islamic University Malaysia,
Kuala Lumpur, Malaysia*

INTRODUCTION

In the contemporary world, effective democratic governance is anchored, among other key factors, on the degree of openness that governments exhibit in political and administrative affairs and the amplitude of integrity public administrators demonstrate in performing their official routine. Democracy cannot thrive in an environment of secrecy, as the free flow of information is imperative for people to follow and scrutinize the operations of a representative government, assess the policies and decisions it makes, and appraise the conduct of its personnel. Open government, which can promote a culture of probity within the public service, facilitates the consolidation and gradual deepening of democracy.^[1] It strengthens the structures of accountability by applying the wherewithal to reduce malfeasance and corruption in public organizations. The more transparent a government is, the fewer will be the opportunities for public administrators and managers to resort to corrupt practices in the tasks they perform and deeper will be people's capacity to exercise control over arbitrary state power. Framing public policies behind closed doors or adhering to obscure procurement and financial practices can lead to making erroneous choices that may have adverse implications for the community and beyond. The immoderate exercise of bureaucratic discretion in a closed secretive environment or the influence of non-transparent corporate culture and the almost unbound volition of oligopolists to control competition in the market place result in corruption. Thus in the symbiotic relationship between corruption and opacity, discretion and monopoly are critical factors,^[2] and the social and economic fallout of inordinate opacity and pervasive corruption can be a bane for democracy and development.^[3]

THE SOUTHEAST ASIAN SCENE

The financial crisis of the late 1990s was a watershed in political governance in Southeast Asia. It exposed

the imperfections of non-democratic political structures and illiberal regimes and the failure of governments to either protect their economies from internal uncertainties and external threats or expose them to economic instability.^[4] In several countries, it brought to the forefront the inadequacies of the state apparatus in fostering a culture of integrity in both government and the economic market for safeguarding the welfare of the people and in sustaining social and economic development.^[5] The crisis, apart from external factors, also stemmed from the recondite operations of governments and markets and their indisposition to share or disseminate information with/to other stakeholders for the effective transaction of governmental or economic business. It also underscored the relevance of openness to accountability for lessening the opportunities for rent seeking, underhand dealings, and other forms of corruption. It highlighted the need for modernizing government by institutional reforms—making the operations of public institutions more transparent and upgrading the mechanisms for enforcing accountability in governments and financial markets.

Among the Southeast Asian nations, Singapore is perceived to be the least corrupt whereas Malaysia, Thailand, the Philippines, Vietnam, Indonesia and Myanmar are alleged to exhibit levels of corruption ranging from high to very high caused by systemic failures and individual dishonesty.^[6,7] The level of governmental transparency is fairly low as citizen access to information is restricted because of overbearing legal and extralegal impediments.^[8,9] Such low levels of transparency have significantly contributed to high levels of corruption in Southeast Asia.

A variety of social, political, and economic factors have contributed to making corruption pervasive, resulting in the effusion of resources and enervating both development activities and the effective delivery of public service.^[10] Most countries in the region have been under authoritarian or pseudo-democratic rule for prolonged periods, and where democratization has been ushered, the consolidation process has been tardy. In government, accountability structures are

imperfect or weak in ensuring vertical and horizontal systems of checks and balances.^[11] Entrenched elites hardly face any competition, giving them a free will to influence officials to obtain undue advantage to serve their parochial interests. These happen out of the public gaze and often escape social and political oversight.

The economic liberalization programs initiated in the last two decades opened the Asian economies to globalization, spurred corporatization within the public sector, and the privatization of public assets and services. Globalization internationalized corruption as corporate power continues to be raised to high levels in the absence or weakness of civil society, particularly in Indonesia and to a lesser extent in Singapore and Malaysia. Transnational corporations have typically corrupted political and bureaucratic structures for “exploiting market opportunities.”^[12] Asian countries are at risk from such international rackets as money laundering, illegal flight capital, financial crime, and people smuggling.^[13]

The media, which can be a powerful influence in creating a sound public integrity system, has not always been free; rather, authoritarian regimes in most of Southeast Asia placed unjustified restrictions on the press in the past and capriciously regulated the electronic media to serve their parochial political interests. Independent information sources have been few and far between, and only recently have they begun to permeate into society and support openness in governance. In countries such as Myanmar, Cambodia, Malaysia, and even Singapore, excessive press censorship is seen as rebuffs on media freedom.^[14]

Thus from a democratic governance perspective, it became essential to put in place structures and mechanisms that would decelerate rent seeking and both systemic and individual corruption in government and the marketplace, facilitate public access to information to comprehend the rationale for and the impact of government decisions, enforce rules to activate disclosure practices in the corporate sector, moderate mismanagement in public organizations, narrow the hiatus between citizens and public agencies, and improve both horizontal and vertical accountability in government. All these are imperative to boost public confidence and trust in government.

MEASURES TO IMPROVE TRANSPARENCY AND COMBAT CORRUPTION

Under both domestic pressure and external intervention to improve governance, the Southeast Asian governments have embarked on wide-ranging programs to enhance transparency and curb corruption within their fold as well as in commerce and industry.

Some of these initiatives were undertaken even before the financial crises laid bare the structural and procedural deficiencies of governance on both the state and corporate planes, mainly as responses to the demands of international financial institutions. Non-governmental efforts, principally by business and industry groups, were heightened postcrisis to institute high levels of corporate ethics, responsibility, and transparency to raise stakeholder confidence in both national and transnational firms' ability to deliver goods and services.

Government, international, regional, and national organizations have initiated open dialogues and comprehensive analyses to locate black spots in governmental structures that incur diminished transparency and weak accountability and to discern the causes of corruption and other forms of political and administrative malfeasance. The international lending institutions are now adopting a hands-on approach to repair institutional weaknesses and freely collaborating with governments in their quest for sound governance. The World Bank, for instance, has been supporting measures that incorporate “deregulation to reduce opportunities for corruption, enforcement of sanctions through development of special watchdog agencies and robust judicial systems, and through strengthened civil society institutions, such as an independent press to raise public awareness of the corruption problem.”^[11] Surveys and analytic reports on corruption in Thailand, Cambodia, Indonesia, and the Philippines have provided directions to combat the malaise as well as enhance transparency.^[9]

Anticorruption Strategies

Several Southeast Asian countries have endorsed the Asian Development Bank/Organization for Economic Cooperation and Development (ADB/OECD) sponsored Anti-Corruption Initiative for Asia-Pacific. This action plan incorporates “a systematic and readily enforceable approach to fight corruption” to promote “economic and social stability” in the region. Basically, this initiative was to underscore the critical significance of developing accountable and transparent systems in public management, designing effective measures to prevent and investigate all forms of corrupt practices, developing an ethical and trustworthy corporate culture, and promoting active citizen participation in anticorruption exercises by stimulating public discussion of the issue and creating mechanisms for easy access to information. Specific country plans are expected to adopt international instruments and standards to design effective public integrity management systems complemented by legal, structural, and administrative reforms.^[15]

Institutions of horizontal accountability such as constitutionally mandated “oversight” bodies, with varying degrees of operational and fiscal autonomy and, in some instances, with quasi-judicial powers, have been created in several Southeast Asian countries to scrutinize government organizations and investigate corruption and malpractices. For example, the Philippines has a Civil Service Commission to enforce “ethics and accountability” among public servants. The Ombudsman “is mandated to investigate and prosecute the criminal liability of public officials and employees involved in graft and corruption,” whereas the Commission on Audit serves as a fiscal vigilante.^[16] The Presidential Commission for Good Government was entrusted with the objective of recovering ill-gotten wealth from individuals, whereas other measures were put in place to ensure public disclosure of assets held by public officials.^[17] In Thailand, the National Counter-Corruption Commission was granted prosecutorial authority to offset unjust governmental edicts, whereas other similar institutions such as the Constitutional Court, the Ombudsmen, the Election Commission, and the State Audit Commission are responsible to promote transparency and accountability and control corruption in the public sector.^[18] Singapore’s strong state capacity and committed political leadership have been instrumental in building and enforcing the mechanisms necessary to keep corruption at a low level. The government has established several instruments and has been strict in enforcing laws through an independent Corrupt Practices Investigation Bureau. Periodic review of government procedures also has the effect of obviating opportunities for corruption in public organizations.^[19]

Like Singapore, Malaysia has had an anticorruption strategy since the 1960s. However, the Anti-Corruption Agency (ACA), the Public Complaints Bureau, and several departmental committees on integrity, quality, and productivity have struggled in keeping the level of corruption low in the public sector. One limitation of the ACA that catechizes its impartiality and objectivity is its lack of independence from political control.^[20] The effectiveness of this strategy is therefore suspect.

Access to Information

Southeast Asian nations have made slow but steady progress in establishing the structures for the free flow of information between government and society. In several cases, the obstacles in information sharing and dissemination have been gradually removed. Thailand and the Philippines have relatively liberal information regimes, and their constitutions are explicit about people’s right to obtain information for a variety of

purpose—either to check the performance of state officials or to comprehend the basis of policy development. These constitutional provisions have been supplemented by laws and regulations such as the Thai Official Information Act and the Philippine Code of Conduct and Ethical Standards for Public Officials and Employees. This is similar to the East Asian (e.g., South Korea, Hong Kong) experience of information disclosure by public agencies and citizens’ access to governmental information for a variety of purpose and especially to ascertain the outcome of decisions bearing on their lives. While some of these measures appear good on paper, they have not always provided the desired level of access to information nor have they always contributed to enhancing transparency or in curbing malfeasance in government. Both Singapore and Malaysia also fare poorly in disclosing governmental information to the public. Official secret acts and tough regulations on Internet access have restricted information access and disclosure.^[10] Other countries in the region are remiss in creating the appropriate structures for freedom of information and in enhancing transparency in government.

ROLE OF CIVIL SOCIETY AND E-GOVERNANCE

In the fight against corruption or in advocating for greater transparency in government and improving public access to information, civil society institutions have been playing a significant role especially in countries where state initiatives have been ineffective in producing results. Similarly, e-governance initiatives have helped improve transparency and information access. Civil society can partner with non-governmental organizations (NGO) and the business sector in creating a culture of integrity and anticorruption in society by persuading and even compelling governments to institute reforms and creating public awareness of the evil and tackling adverse consequences of corruption. In Thailand, civil society groups and organizations have been active in placing pressure on the government to modify constitutional arrangements and create or strengthen public bodies to combat corruption.^[21] Chapters of Transparency International in some Asian countries have initiated purposive moves against corruption through research and surveys, conferences and seminars, publications, and lobbying. Its annual Corruption Perception Index highlights the extent of corruption prevalent in a country and helps mobilize public opinion and stimulate action for strengthening governmental initiatives. As in Korea, where a bold step was taken by the People’s Solidarity for Participatory Democracy to frame on its own initiative an

anticorruption bill and lobby for its enactment by the national legislature,^[22] the Indonesian Coalition for Information Freedom (a coalition of NGOs) drafted a freedom of information act for enabling greater transparency of government business and parastatal and non-governmental organizations. This was in response to the government's failure to initiate legislation on information access.

The information gap that contributes to the lack of transparency can be substantially reduced by employing information communication technology (ICT) in linking organizational units, the government, the private sector, and the state with the people. Information communication technology can empower the poor and the marginalized in societies by affording relevant information to them and enabling their participation in public affairs through their feedback on public services and the exercise of their "voice." The Internet is being widely used not only to disengage citizens from long queues at service counters and simplify procedures, but also to make administrative practices more transparent and comprehensible. Indeed, e-government links citizens to the state by facilitating communication and enabling a positive relationship between them.

Many Asian governments have afforded ICT high priority and have adopted comprehensive policies to overcome the information divide between the state and citizens. Several governments have gone on-line providing not only a wide array of information on policies and procedure, but also a variety of e-services including easy lodgment of complaints via the Internet and quick retrieval of information.^[23] Dedicated Web portals link citizens to ministries and public agencies and the services they provide. The creation of a "cashless" payment system and processing of clearance documents by using ICT has vastly reduced corruption in the Philippine Customs Bureau and advantaged customers from its services.^[24] Similarly, initiatives in other countries have also yielded useful results.

CONCLUSION

The Asian financial crisis has left its impact on the society and public administration in the region. Apart from causing considerable economic hardship, it also exposed the weaknesses and shortcomings of the existing governmental systems and processes and signified the need for reform to improve governance and public integrity. However, reforms are mainly focused on revamping the institutional mechanisms and strengthening anticorruption legislations without seeking to effect corresponding changes in other areas. In essence, the political systems continue to be largely authoritarian and decision-making processes closed and/or

centralized with limited scope for public participation. In some cases, freedom of expression and information is extremely limited, the media continues to be the captive of the government, independent oversight bodies are absent, civil society groups face formidable constraints, and draconian legislations remain firmly in place. With restricted public scrutiny, public officials discreetly cover up mistakes, are arbitrary in making policy decisions, and indulge in corrupt practices. Because exposure is a powerful deterrent to corrupt practices, Southeast Asian nations need to introduce greater transparency and openness in public management, create independent judiciary and anticorruption bodies with wide power, spare the media from undue restraints, and widen the space for civil society groups to play an effective role. These will benefit society, improve social trust, build public confidence in governments, and contribute to political and economic stability.

REFERENCES

1. UNDP. *Human Development Report 2002*; UNDP: New York, 2002.
2. Klitgaard, R. *Controlling Corruption*; University of California Press: Berkeley, 1988.
3. Kaufmann, D.; Kraay, A.; Zoido-Lobaton, P. Governance matters. In *Policy Research Working Paper*; World Bank: Washington, DC, 1999; Vol. 2196.
4. Backman, M. *Asian Eclipse: Exposing the Dark Side of Business in Asia*; John Wiley: Singapore, 1999; 23–41.
5. Vishwanath, T.; Kaufmann, D. Toward transparency: new approaches and their application to financial markets. *World Bank Econ. Obs.* **2001**, *16* (1), 41.
6. Transparency International. *Corruption Perceptions Index 2003*. <http://www.transparency.org>.
7. OECD. *Public Sector Corruption: an International Survey of Prevention Measures*; OECD: Paris, 1999.
8. PCIJ. *The Right to Know: Access to Information in South East Asia*; The Philippine Center for Investigative Journalism: Manila, 2001.
9. World Bank. *Reforming Public Institutions and Strengthening Governance: Implementation Update. Part 2*; World Bank: Washington, DC, 2002; pp. 16, 21.
10. Transparency international. *Global Corruption Report 2003*; 149. <http://www.transparency.org>.
11. World Bank. *Reforming Public Institutions and Strengthening Governance: A World Bank Strategy*; World Bank: Washington, DC, 2000; pp. 86, 88.
12. Goudie, A.W.; Stasavage, D. Corruption: the issues. In *Technical Paper*; OECD Development Center, 1997; Vol. 122, 12 pp.
13. Collins, P. Governance and development: new frontiers. In *Applying Public Administration in Development: Guideposts to the Future*; Collins, P., Ed.; John Wiley: London, 2000; 271 pp.
14. RWB. Introduction: Asia and Pacific. In *Annual Report 2003*; Reporters Without Borders, 2003. http://www.rsf.org/article.php3?id_article=6656.

15. ADB/OECD. *Anti-corruption Initiative for Asia-Pacific: Combatting Corruption in the New Millennium*; Asian Development Bank: Manila, 2001.
16. CSCP. *Ethics and Accountability: the Philippine Experience*; Civil Service Commission of Philippines: Manila, 2000. <http://unpan1.un.org/intradoc/groups/public/documents/apcity/unpan003218.pdf>.
17. Moran, J. Patterns of corruption and development in East Asia. *Third World Q.* **1999**, 20 (3), 569–587.
18. World Bank. Fostering institutions to contain corruption. *PREM Notes* **1999**, 24.
19. Yak, C.C. Good people, good laws. In *Progress in the Fight Against Corruption in Asia and the Pacific: Conference Proceedings*; Asian Development Bank: Manila, 2000; 65 pp.
20. Aziz, T.A. International case study; stamping out corruption in Malaysia. In *Paper Presented at the 113th International Training Course*; United Nations Asia and Far East Institute: Tokyo, 2002; 399 pp.
21. ADB/OECD. *Progress in the Fight Against Corruption in Asia and the Pacific: Conference Proceedings*; Asian Development Bank: Manila, 2000; 5 pp.
22. Lee, H.-B. Government–non-government organization interaction in drafting the republic of Korea’s anti-corruption law. In *Anti-corruption Initiative for Asia-Pacific: Combatting Corruption in the New Millennium*; Asian Development Bank: Manila, 2001; 153 pp.
23. Yong, J. *E-Government in Asia—Enabling Public Service Innovation in the 21st Century*; Times Press: Singapore, 2003.
24. Parayno, G.I., Jr. *Reforming the Philippine Customs Services Through Electronic Governance*. Asian Development Bank: Manila, 1999; 61–69.

Transportation Policy

Jeremy F. Plant

School of Public Affairs, Pennsylvania State University at Harrisburg,
Middletown, Pennsylvania, U.S.A.

INTRODUCTION

Transportation policy is a broad category of public policy, comprising those plans, programs, and actions of all levels of government concerned with the movement of goods and people. Transportation policy goals are numerous and varied. They include the safe and efficient operations of all modes of transportation; the provision of infrastructure, such as highways, ports, airports, and bridges; regulation of private-sector businesses performing transportation functions in society; advancement of science and technology to improve transportation systems; collection of data to measure performance and guide policy formation; measures to secure transportation facilities and operations from the threat of terrorist activity; and the integration of transportation programs with programs in such areas of policy as the natural environment, urban space, parklands, energy usage, and agriculture.

THE FRAGMENTATION OF TRANSPORTATION POLICY

Transportation policy has been characterized in the American setting by a fragmented approach in both its formulation and its implementation.^[1] Despite the creation of a federal department of transportation in 1966, the United States has never had a coherent over-all national transportation policy or plan to guide programming in the various modes and program areas. There are a number of reasons to explain why transportation policy is fragmented. Historically, the nation has relied upon private enterprise to provide most transportation services. The primary role of government has been to regulate the private sector and to provide infrastructure, such as roads, bridges, waterways, port facilities, and airports.

A second factor leading to policy fragmentation has been American federalism. All three levels of government—and, increasingly, a fourth level, regional or metropolitan governance—have traditionally found roles in transportation policy. The most common federal government roles have been as the source of funding for improvements in transportation infrastructure and the regulation of interstate and international

commerce. State and local governments have been the major implementers of policy, actually owning and operating public highways, transit systems, and other transportation facilities. In recent years, stimulated by federal policies and citizen demands for better performance, state and local governments have assumed a major role in planning functions that integrate transportation programs with land-use planning, economic development activities, energy conservation efforts, and a host of other programs that impact transportation or are, in turn, impacted by transportation policies.

Perhaps the most pervasive aspect of fragmentation in transportation policy has been the designing of policies for a specific *mode* of transportation. Modes are important as the basis of most transportation operations. Each has developed in a distinctive historical manner, and each has posed unique problems for policy makers and program managers. The modes of transportation usually considered discrete are water transport, rail, highways, air carriers, transit, and pipelines. Often overlooked is pedestrian travel, which is itself a mode of travel. However, each of the modes contains within it significant variety: railroads include freight and passenger roads, large “Class One” trunk lines, and regional and short lines. Highways carry trucks and automobiles, buses, and taxis. Pipelines may convey liquids or gases. Air policy has to take into account both commercial, scheduled air service, and general aviation (and, increasingly, air freight).

In recent years, *intermodalism* has transformed the transportation sector, raising questions about the traditional mode-based approach to transportation policy. Intermodalism is defined as the use of more than one mode of transportation to perform a transportation service. In the movement of goods, it has become common for freight to be moved in standard containers, often moving from the factory to port by surface transportation, then to a port of entry by air or ship, and then to point of delivery by a combination of rail and trucking. Intermodalism has become a metaphor for the burgeoning of global commerce; it has also been a driving force in changing government policies and priorities. State transportation agencies have been changed from highway departments to multimodal transportation departments, largely to deal with the new realities of the transportation sector.

A HISTORICAL LOOK AT TRANSPORTATION POLICY

Transportation was a policy issue from the beginning of the American republic. This can be explained in two ways. First, the United States in the late eighteenth century was a large and thinly populated nation in desperate need of good transportation. As had been the case for several millennia, the speed of transportation was limited to the speed of sailing ships and the horse. As the population moved inland from the coast and the navigable sections of rivers, the need to conquer distance with better transportation became vital to national development and economic prosperity.

The issue of transportation and other infrastructure improvements divided the early political parties. The Federalist Party under Alexander Hamilton's policy leadership advocated a program of modest national leadership in public investments. It was replaced by the more conservative approach of the Jeffersonians in the early 1800s, although a number of states and some municipalities began to take an active interest in transportation improvements. First canals, later railroads, then, in some instances, transit systems received public *subsidies*. Private companies were given charters or direct assistance to build transportation facilities that, the argument went, would bring benefits not only to the company but the public at large. The Whig Party under the leadership of Henry Clay advocated a national program for internal improvements, many of which were transportation related, but they were unsuccessful in achieving their goals. It was not until the 1860s that a program of aiding railroad companies by awarding them grants of land (a policy supported by President Lincoln, a former Whig) enabled the federal government to achieve its first and still perhaps greatest objective, the opening of the West by supporting construction of a transcontinental railway, which drove the final spike in 1869, giving the nation a system of surface transportation linking every part of the country by rail.^[2]

As transportation companies (principally railroads) matured in the late nineteenth century, a reaction to their power, both economic and political, led to demands for *regulation*. The American preference for the private ownership and operation of transportation services was retained through policies that tried to force private entities to achieve public goals through regulatory restraints. The right of the federal government to regulate interstate commerce pushed the national government into the forefront of transportation policy once again. Beginning with the Interstate Commerce Act of 1887, federal regulatory agencies were set up to do such things as set rates, license carriers, maintain public safety, authorize routes, and other permissible activities of private carriers.

Congress delegated the bulk of the actual policy-making role to the individual regulatory agencies. The structure and operations of these agencies tended to isolate them from policy direction and leadership of the national executive. Lacking a bureaucratic structure of leadership, they were governed by a group of commissioners selected on a bipartisan basis, making it difficult for them to pursue clear policy goals.

Although water-based transportation, both marine and inland, continued to be of major importance throughout the nineteenth and early twentieth centuries, railroads were the dominant mode for both passenger and freight transportation well into the twentieth century.^[3] The Hepburn Act of 1905 empowered the Interstate Commerce Commission (ICC) to set most railroad rates, a major step in providing the public with real control over the railroads. The federal government briefly nationalized the railroads during World War I, but returned them to private ownership shortly after the end of the conflict. The Transportation Act of 1920 recognized that the rail industry was overbuilt and underperforming, and gave the ICC the power to rationalize the national system by creating competitive rail systems within the major regions of the country. However, the ICC was too weak an organization and the railroad industry too apprehensive of public policy direction for the act to lead to major changes. By then, a new modal competition was emerging and with it a changing direction for public transportation policy as well.

FROM RAILROAD TO HIGHWAY TO SKYWAY

The automobile made its first, modest appearance in the years directly before and after the turn of the twentieth century. The automobile transformed transportation policy in a number of ways. Since automobiles were for the most part individually owned and operated, licensing and enforcement of traffic laws increased the role of local and state governments in transportation. The need to improve and expand the primitive road network added another responsibility for government. At first assumed to be a local responsibility, the expansion of the nation's roadways became a national policy goal utilizing a new policy tool, the intergovernmental grant-in-aid, to allow the federal government to support the states financially, while not actually owning or maintaining most of the roadways. This approach was institutionalized in the Federal Highway Act of 1921, which required the states to plan a modern system of roads that would evolve into a national highway system. A series of federal acts since then has helped move the nation toward a degree of uniformity in the design, signage, and maintenance of a national system of highways.

The airplane made its appearance at roughly the same time as the automobile, in the first years of the twentieth century. Commercial use of aviation proceeded swiftly after major technological advances were made during and after World War I. By the 1920s, mail and passengers were being conveyed by fledgling air carriers. The clearly interstate character of air commerce necessitated a federal policy response. The policy approach chosen was regulation, in line with the precedent set by rail policy. The most significant legislation was the Civil Aeronautics Act of 1938, which set up the Civil Aeronautics Authority to regulate routes, license carriers, and set rates.

During the administration of Franklin D. Roosevelt, the regulatory approach was expanded to cover virtually all aspects of transportation in the United States. In addition to aviation, the merchant marine, coastal shipping, trucking, and intercity bus services were all subject to regulation. The private sector, which had initially resisted government regulation, found benefits in the approach.^[4] Regulation protected the private sector from excessive competition and government takeover. As an example, the major transportation services were not nationalized during World War II as the rail industry had been during World War I, even though greater demands were placed upon it in the later conflict.

Regulatory policy helped to stabilize the country's transportation system. Criticism of regulation began to emerge, however, as the decentralized and legalistic approach of regulation was unable to deal with the complexity of transportation's impact on society, or the need to increase economic efficiency to reduce the costs of transportation as a factor in production and personal travel. Nowhere was this more apparent than in the changing transportation needs of cities after World War II.

URBAN TRANSPORTATION POLICY

The growing use of automobiles to commute from home to work began to affect urban areas even before the great movement from city to suburb after World War II.^[5] Trolley companies were the first to feel the impact, with many lines shutting down due to competition from cars, coupled with the economic problems of the Depression years. Urban transportation emerged as a major subset of transportation policy and a major focus of academic interest in transportation.^[6] Urban transportation policy has added greatly to our understanding of the relationship of transportation to other disciplines, including housing, land-use planning, economic development, and architectural design.^[7]

Urban transportation policy has been concerned to a large degree with the impact of the automobile on the

city and alternatives to automobile usage.^[8,9] Experts with an urban policy focus have been at the forefront of arguing for greater expenditures on urban systems, especially ways of providing effective public transit, a major policy issue that has dominated urban transportation planning since the 1940s. Urbanists have also been among the leaders in utilizing sophisticated analytical techniques to measure and model traffic flows, determine citizen attitudes toward transportation, explore the economic impacts of transportation investment, and a host of other important policy questions.^[10]

Urban transportation policy has also been a focal point for consideration of the best type of organizational approach to plan, provide, and manage transportation services. As private bus and rail companies expressed a desire to close due to declining profits or losses, general-purpose governments and regional authorities often stepped in to ensure the continuation of services. This often involved state governments as well. After several decades of little growth in urban rail systems, there has been a renewed interest in alternatives to the highway. Major new systems have been constructed in the San Francisco Bay Area, the Washington, DC metro area, Atlanta, Baltimore, Miami, and Los Angeles. Commuter rail systems also moved from the private to the public sector as private carriers desired to shed money-losing commuter operations.

The assignment of costs is a major policy consideration, before and after construction of the systems. Should the operating costs be covered by ridership fares only, or should government provide subsidies? If so, what is acceptable performance? While not universally recognized as the norm, a 50/50 split between farebox returns and subsidies is a common performance benchmark.

Reversing the trend toward an increased public role is the belief that privatization of public operations leads to greater efficiency or economy.^[11] Privatization efforts have impacted public transportation in areas such as fleet maintenance, policing activities, construction or operation of toll roads, and other support functions. Advocates of privatization have helped expand the policy debate on feasible alternatives to existing transportation programs, the assignment of costs and benefits to groups in society, and means of increasing operational efficiency of transportation operations. Outside the United States, privatization has become a major focus of transportation policy as state-owned transportation organizations have been sold to private interests or have been forced to adhere more to market forces. The privatization phenomenon has profoundly changed transportation systems in both developed nations such as Great Britain and New Zealand and a number of less-developed nations, and promises to

remain an alternative to traditional government-centered programming.

CHANGES IN FEDERAL TRANSPORTATION POLICY

Until the 1990s, most federal legislation dealing with transportation was mode specific. Advocates of a national transportation policy, planned and administered in a comprehensive manner, pushed for the creation of the U.S. Department of Transportation (DOT), which was created in 1966. But the power within DOT remained more with the program agencies organized by modes and programs than with the central office of the secretary. However, the impending completion of the immensely popular and successful Interstate Highway System pulled together a new coalition of groups in the late 1980s to envision transportation policy after the interstate program was terminated. The result was the Intermodal Surface Transportation Efficiency Act of 1991, or as it became known, “ISTEA” (pronounced like the drink). The ISTEA was a major change in federal transportation policy, allowing more flexibility to states and metropolitan planning organizations than had been permitted in prior programs. It also continued funding at a level considered high by most stakeholders. The program increased participation at the metropolitan and state level and became a model for the next surface transportation bill, the Transportation Equity Act for the Twenty-first Century, or “TEA-21.” Influenced more by traditional legislative desires to target money to projects favored by Congress, it retained most of the features of ISTEA. Both ISTEA and TEA-21 helped move state transportation departments along the path away from just a highway construction and maintenance function, and more toward the planning and management of a multimodal, performance-driven system model.

The direction set by ISTEA and TEA-21 was fundamentally retained in the most recent reauthorization of surface transportation programs, dubbed SAFETEA-LU, which passed Congress in 2005. (The official name of the act illustrates the variety of goals now reflected in transportation policy: Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users.) The primary issues involved were issues of quantity more than philosophy: how much total spending would go to transportation, and how it would be divided between programs. A difference between SAFETEA-LU and its predecessors, however, was the growth in the number of congressionally dictated set-asides, suggesting that surface transportation policy is being led as much by congressional interests as from the executive branch.

At roughly the same time that incremental steps toward a more comprehensive approach to surface transportation policy have taken hold at the national and state levels, economic regulation of the different transportation modes has largely been replaced by reliance on market forces. Economic regulatory agencies such as the Civil Aviation Board and ICC were terminated, and many residual regulatory programs, largely related to safety concerns, have been assigned to DOT. The deregulation of transportation industries has had dramatic impacts on the affected industries, leading to a drastic reduction in the number of operating carriers, increasing the diversification of transportation companies, and stimulating the growth of intermodalism and global transportation linkages.

CONTEMPORARY ISSUES IN TRANSPORTATION

As the twenty-first century progresses, a number of educated guesses can be made about the direction of transportation policy.^[12] All signs point to the increasing integration of transportation issues to broader issues of national security, global economic commerce, energy use and conservation, and land-use patterns and planning. One driver of change is technology. Information technology and intelligent transportation systems will add efficiency and a degree of “customer friendliness” to the operations of transportation agencies. More energy-efficient and cleaner vehicles, such as hybrid cars and light trucks, are already creating demand for policies to spur their development and use. Better use of information can also help rationalize demand for urban transportation services and add capacity to a system that is already strained.

Global economic integration will continue, making international protocols on security, information sharing, and uniform regulations on carriers essential. As multicentered urban regions evolve, traditional patterns of suburb-to-city commuting will continue to change, demanding new and innovative approaches to providing the essential transportation service of work-related travel. The aging of the population will raise new issues of service for groups with special needs and for individuals who have passed beyond the age at which they can safely drive.

The threats posed by terrorism and natural disasters have created major changes in transportation policy. The Department of Homeland Security, in particular its Transportation Security Administration, has become a significant new actor in the transportation arena, and is expanding its role from one focused largely on the aviation sector to include more oversight of surface transportation. Port security remains a problem that continues to draw attention from policy makers. Hurricane Katrina exposed the transportation

system to risks beyond terrorist activity. In all, the costs of securing the transportation system from both terrorist and natural threats are staggering, and the need to balance security needs with the needs of efficiency, safety, and conservation have added to the complexity of the transportation policy arena.

In summary, transportation policy in the United States in the postindustrial era is forced to deal with more complex policy issues, as the traditional values of transportation policy, mobility, and efficiency are seen in the context of broader social, economic, and environmental concerns. New approaches for measuring performance and organizing transportation activities have greatly altered the mode-specific, highly regulated system that lasted into the 1970s.

CONCLUSION

Perhaps the greatest policy challenge today remains the imbalance between the critical importance of transportation and the attention it is given in the setting of national political priorities and resources. Transportation accounts for about 16% of the gross domestic product and is arguably the public service most frequently accessed by citizens. Yet it remains a subject that rarely surfaces in national policy debates, and the share of the federal budget devoted to transportation has fallen since 1960, mostly because domestic social programs such as health, social security, and education have grown in importance. The average citizen knows little of the transformation of the transportation sector brought about by globalization, deregulation, and structural change in the transportation industry. The salience of transportation as a policy priority is felt more in state capitals and city halls, as the capacity of the transportation system remains relatively fixed and demand for more and better services increases. However, when terrorism, natural disaster, or cutoffs in oil supplies threaten existing transportation services or lead to significant spikes in costs, we may see the day when transportation policy may once again be

the subject of intense national debate, as it was in the days of Alexander Hamilton, Henry Clay, and Theodore Roosevelt.

ACKNOWLEDGMENTS

I want to thank my colleagues from the Section on Transportation Policy and Administration of the American Society for Public Administration for the many insights I have gained from them over the past decade on transportation policy issues.

REFERENCES

1. Hazard, J.L. *Managing National Transportation Policy*; Eno Foundation for Transportation: Westport, CT, 1988.
2. Taylor, G.R. *The Transportation Revolution 1815–1860*; Rinehart and Co.: New York, 1951.
3. Chandler, A.D. *The Visible Hand: The Managerial Revolution in American Business*; Harvard University Press: Cambridge, MA, 1977.
4. Lowi, T.J. *The Politics of Disorder*; Basic Books: New York, 1971.
5. Pisarski, A.E. *Commuting in America II*; Eno Foundation for Transportation: WA, 1996.
6. Smerk, G.M., Ed.; *Readings in Urban Transportation*; Indiana University Press: Bloomington, IN, 1968.
7. Weiner, E. *Urban Transportation Planning in the United States*; Praeger Publishers: Westport, CT, 1999.
8. Dunn, J.A. *Driving Forces: The Automobile, Its Enemies, and the Politics of Mobility*; Brookings Institution: WA, 1998.
9. Plant, J.F. Transportation and growth management. *Public Works Manage. Policy* **2001**, *6* (2), 91–101.
10. Cervero, R. Unlocking suburban Gridlock. *J. Am. Plan. Assoc.* **1986**, *52*, 389–406.
11. Winston, C.; Shirley, C. *Alternate Route: Toward Efficient Urban Transportation*; Brookings Institution: WA, 1998.
12. Kane, A.R. Transportation in the new millennium. *Transp. Quart.* **2000**, *54* (3), 5–9.

Trust in Organizations

David G. Carnevale

Department of Political Science, University of Oklahoma, Norman, Oklahoma, U.S.A.

INTRODUCTION

Trust is “faith and confidence in the intentions and actions of a person or group to be ethical, fair, and non-threatening concerning the rights and interests of others in social exchange relationships.”^[1] Trust is institutionalized in an organization’s rules, roles, and relations.^[2] That means trust is ubiquitous in organizations and nearly everything an organization does influences it, and is disposed by it, in some way.

There is an inherent physics associated with trust. It can be regenerative with a fair measure of openness, a lessening of a need for control, a willingness to take reasonable risks, and amenable to the influence of system members. Regenerative dynamics are self-reinforcing mechanisms that escalate and intensify. They breed trust. Degenerative systems are the polar opposite. Their mutual reinforcing features spiral down to closed-mindedness, self-protective behavior, deepening control, and low trust.^[3] The nature of organizational processes feeds the regenerative or degenerative vitality of an organization. The correlates of the high-trust regenerative model are embedded in everything an organization performs.

For instance, the organizational processes associated with trust are myriad. Trust is associated with productivity,^[4] group performance,^[5] cooperation and conflict,^[6] leadership styles,^[7] managerial assumptions about employees,^[8] need satisfaction,^[9] organizational development,^[10] communication,^[11] psychological contracts,^[12] the quality of labor–management relations,^[13] and “administrative evil.”^[14]

How an organization chooses to structure and, especially, how it elects to manage the work processes above speaks volumes about underlying, often unconscious, assumptions about trust in employees. Organizations are social systems, and trust is a kind of adhesive that holds the social fabric together. Trust is social capital. It is psychological money in the bank. Paradoxically, trust increases the more it is used and disappears entirely if it is not nurtured or is violated.

TRUST

Trust is implicit in whether or not participation and involvement are permitted. Trust is apparent whether

employees have effective voice. Whether or not organizational rewards and incentives are fairly distributed affects trust. The presence of some kind of forum where conflict can be addressed fairly is a correlate of trust. Whether employees are given the chance to do meaningful, not narrowly specialized high-control jobs, demonstrates trust in staff. In all of these design arrangements and social process arrangements, expressions of trust are articulated, sometimes subtly, often symbolically, but always in ways felt by workers. Most staff can easily assess whether the physics of their workplace is regenerative or degenerative when it comes to the sense of overall trust.

Trust is a reciprocal attitude that explains its regenerative or degenerative dynamics. Actions are self-reinforcing. Organizational routines tend to flow in self-heightening or self-depreciating ways. Low-trust arrangements breed low trust. High-trust work conditions are self-reinforcing.

Douglas McGregor^[8] provided the classic description of regenerative and degenerative forces about trust. Managers can assume Theory X attitudes that legitimize low-trust management. These hold that the average human being has an inherent dislike of work, will avoid it if possible, and prefers to be directed, wishes to avoid responsibility, has relatively little ambition, and wants security above all.^[15]

Theory X justifies the command and control bureaucracies so common at the beginning of the twentieth century. If staff members do not want to work, will not work hard without close supervision, and do not have the best interests of the organization at heart, then it follows that employees need to be watched, pushed, and coerced into acceptable performance. It is a prescription for degenerative employee relations.

There is another set of assumptions about people that are the opposite of Theory X. These too arise from McGregor^[16] and are known as Theory Y beliefs. These posit that employees have the capacity to exercise a relatively high degree of imagination, ingenuity, and creativity in the solution of organizational problems. Given this perspective, managers can work with employees to set objectives, reduce oversight, expand the scope of action giving employees greater control over the conception and execution of their work, and invest in staff training and development. Managers can risk that good work can be accomplished without excessive controls. These are regenerative attitudes.

In summary, the heart of trust in organizations emanates from the predispositions managers hold about the nature of people to accept responsibility. The underlying mindsets are pessimistic or optimistic. Because trust is a reciprocal attitude, the core expectations about the motivation and capacity of persons drive management philosophy that, in turn, creates the nature of organizational trust.

HIERARCHY AND WORK ROLES

All organizations are hierarchical to some degree, even those that have labored to flatten hierarchies, “turn bureaucracies on their heads,” create matrix structures, or foster network arrangements. If there is anything that remains true, organizational change agents have not solved the problem of hierarchy and there are implications for trust.

Creed and Miles^[17] described the evolution of organizational designs over time and demonstrated that as organization forms have changed, so have the managerial philosophies concerning trust. The authors compare the Bureaucratic, with the Human Relations model, and, finally, the Human Resources paradigm. Each category represents a managerial philosophy that establishes expectations about the trustworthiness of staff. The Human Resources model is most advanced. It is designated as the “Human Investment Philosophy,” which assumes that people not only want to contribute, but have unlimited potential to learn and grow. People are seen as trustworthy, and the manager’s basic task is to develop human potential. Expectations are high and so is trust. Since the turn of the past century, structure and philosophy have moved from low-trust Bureaucratic, through a transition state Human Relations that served as a counterforce to the classical hierarchical paradigm, to the current Human Resources model, which has more regenerative possibilities.

FAIRNESS

Justice is difficult to define and no doubt varies among people. It is the same with the idea of fairness. Nevertheless, the issue of fairness covers a number of organizational activities that have a profound influence on levels of trust in organizations.

There are many ways to approach the fairness problem. In simplest terms, one can think of fairness in organizations as being about equality and equity in how reward and incentive systems are governed and the extent of justice in how discipline is administered. Mishandling arouses feelings of distrust among

staff. As any manager realizes, workers pay keen attention to whom gets what in organizations.

There are two perspectives on the exchange of rewards in organizations. The first centers on *distributive justice*, which arises from equity theory.^[18] The distribution of outcomes is seen as fair when people get rewards commensurate with what they invest. The more merit achieved, the greater the reward secured. The less the contribution made, the smaller the outcome. Most people understand this calculus as fair. Trust is violated when people who do more receive less and those who do less receive more. It is not an uncommon problem in organizations that misallocating rewards is a catalyst for the downward degenerative sequence.

Procedural justice takes a different tack in explaining how people assess the administration of rewards and punishments in organizations. In both instances, how things are done matter as much (sometimes more) than what is done.^[19] The fairness connection is that if people do not trust the methods used to make decisions, they will not trust results. Like all the other correlates of organizational trust, this is a process issue.

COMMUNICATION

One of the most important process issues in organizations is communication. The questions are whether communication is open or closed, whether people feel free to speak their minds, and whether secrets are entrenched. Openness of communications allows the truth to be understood, for people to be authentic, and for organizations to deal with reality, not some illusion of it. When the free flow of information is manipulated, fear and suspicion are created. Rumors typically flourish, as employees speculate about things they are not permitted to speak openly about.^[20] The former approach is regenerative, high trust. The latter degenerative, closing off possibilities for meaningful levels of trust.

POLITICS

Politics is in play everywhere in organizations. It is about who gets what, when, where, and how. Politics is degenerative when it is characterized by game playing, inauthentic conduct, personalized power motive, and self-interested behavior. It is inevitable that there will be politics in organizations. There is nothing inherently wrong with trying to influence others concerning the allocation of resources, the ends and means an organization ought to pursue. There are honest differences of opinion about the need for and nature of change. It is ethical to form coalitions, keep promises, be civil, refuse to engage in character assassination, tell

the truth, and not promote zero-sum legitimacy. Principled politics are regenerative and promote a culture of trust. Unethical politics breed distrust. Like every other organizational process, there are choices about courses of action. What is embraced has consequences for trust.

LEADERSHIP

It all starts at the top. There are all kinds of leadership styles.^[21] Of all the types, transformational leadership is strongly associated with trust. The transformational leader generates positive feelings in subordinates. The transformational leader possesses: 1) vision; 2) charisma; 3) powerful communication skills; 4) an empowering attitude; and 5) is considered trustworthy.^[22] Transformational leaders are interested in change. They encourage alternative futures. To convince followers to take risks without the guarantee of reward is the key to overcoming resistance to change. Leadership credibility is a powerful source of influence. Without trust, resistance to change and defensive conduct are exacerbated.

CONFLICT

Because people have dissimilar values, mind-sets, objectives, and visions about what an organization should be doing, disagreements are common. Scarce resources encourage endless rounds of positioning for advantage. Contests of will, ambition, and personality clashes contribute to discord. The physics of conflict are also not helpful. Conflict tends to escalate, positions harden, the dispute spreads as partisans seek allies, communication stops, stereotyping occurs, and, finally, revenge becomes an objective. No organization can be high performing if its social system is awash with suspicion and guarded behavior. Like politics, there will be conflict in organizations. Conflict tears at the social fabric of organizations, and trust is one of the first things destroyed.

There is an upside to conflict. It can clear the air. It can get people to express how they really feel about issues. The healthy interaction of differences can lead to synergy in groups, for instance, and then to creativity. When people know they are safe win-lose, zero-sum conduct is reduced. As trust increases, diversity of ideas becomes an advantage.

It is not the law of gravity at work that discord need be dysfunctional. Persons have a choice about what kind of conflict will exist. There are methods available to help public administrators deal with conflict. These are generally known as third-party peacemaking interventions in organizational development or alternative

dispute resolution (ADR) being mandated in several public jurisdictions. People cannot resolve disputes if they do not trust those they think are interfering with the attainment of their legitimate needs, or if they do not trust the process being used to get at differences, or if they have no confidence in the third parties facilitating the process. Trust is always on the table in conflict resolution.^[23]

CONCLUSION

Trust is a function of regenerative practices. It is a precondition for the establishment of harmonious social relations and the elimination of dysfunctional conflict.^[24] Trust matters, because social systems need to have cultures where people can rely on one another under conditions of dependence and risk. Trust is social capital and a moral resource in organizations. When trust declines, organizational capacity is reduced.

Trust is rooted in nearly every structural and process aspect of organizations. Employees are constantly making judgments about trust depending on the general tone of the managerial philosophy that pervades the organization. Workers are sensitive to the cultural aspects of trust. They experience it in work roles, nature of communication, credibility of leadership, the way conflict is handled, the type of politics, and the amount of confidence expressed in them. High-performing organizations feature trust and work to develop it as an important part of the organization climate. Trust is an asset. Its power should not be underestimated.

REFERENCES

1. Carnevale, D.G. *Trustworthy Government: Leadership and Management Strategies for Building Trust and High Performance*; Jossey Bass Publishers: San Francisco, CA, 1995.
2. Fox, A. *Beyond Contract: Work, Power and Trust Relations*; Hutchinson & Co.: London, 1974.
3. Golembiewski, R. *Approaches to Planned Change Part I*; Marcel Dekker: New York, 1979.
4. Golembiewski, R.; McConkie, M.L. The centrality of interpersonal trust in group processes. In *Theories of Group Processes*; Cooper, C.L., Ed.; Wiley: New York, 1975.
5. Zand, D.E. Trust and managerial problem solving. *Adm. Sci. Q.* **1972**, *1*, 229–239.
6. Deutsch, M. *The Resolution of Conflict*; Yale University Press: New Haven, CT, 1973.
7. Likert, R. *The Human Organization*; Mc-Graw-Hill: New York, 1967.
8. McGregor, D. *The Human Side of Enterprise*; McGraw-Hill: New York, 1960.

9. Maslow, A. *Motivation and Personality*; Harper-Collins: New York, 1954.
10. Golembiewski, R.T. *Humanizing Public Organizations*; Lomond Publications: Mt. Airy, MD, 1985.
11. Mellinger, G.D. Interpersonal trust as a factor in communication. *J. Abnorm. Soc. Psychol.* **1956**, 52, 304–309.
12. Argyris, C. *Understanding Organizational Behavior*; Dorsey Press: Belmont, CA, 1960.
13. Reich, R. *Tales of a New America*; Times Books: New York, 1987.
14. Menzel, D.C. Ethics versus evil: is it before or after midnight in the garden of ethics and evil? *Am. Rev. Public Adm.* **2001**, 31 (3), 340–350.
15. McGregor, D. *The Human Side of Enterprise*; McGraw-Hill: New York, 1960; 33–34.
16. McGregor, D. *The Human Side of Enterprise*; McGraw-Hill: New York, 1960; 48 pp.
17. Creed, W.E.D.; Miles, R.E. Trust in organizations: a conceptual framework linking organizational forms, managerial philosophies, and opportunity costs of controls. In *Trust in Organizations: Frontiers of Theory and Research*; Kramer, R.M., Tyler, T.R., Eds.; SAGE Publications: Thousand Oaks, CA, 1996; 16–39.
18. Homans, G. Social behavior as exchange. *Am. J. Sociol.* **1958**, 63, 597–606.
19. Thibaut, J.; Walker, L. *Procedural Justice: A Psychological Analysis*; Erlbaum: Hillsdale, NJ, 1975.
20. Argyris, C. *Knowledge for Action: A Guide to Overcoming Barriers to Organizational Change*; Jossey-Bass: San Francisco, 1993.
21. Yukl, G.A. *Leadership in Organizations*, 3rd Ed.; Prentice-Hall: Englewood Cliffs, NJ, 1994.
22. Bass, B. *Leadership and Performance Beyond Expectations*; Free Press: New York, 1985.
23. Carnevale, D.G. *Organizational Development in the Public Sector*; Westview Press: Boulder, CO, Forthcoming.
24. Deutsch, M. Cooperation and trust. In *Nebraska Symposium on Motivation*; Jones, M., Ed.; University of Nebraska Press: Lincoln, NE, 1962; 275–320.

Truth and Reconciliation Commission

Daryl Balia

*Professional Ethics and Risk Management, Public Service Commission,
Pretoria, South Africa*

INTRODUCTION

To unearth one's bloody and repressive past is highly risky for a country traveling the road of reconciliation and healing. Not only is the prospect of renewed violence kindled, but also the trauma of painful remembrance, which could so easily infect succeeding generations. South Africa as a sovereign nation decided, however, that a collective search for truth and justice was necessary for uncovering its apartheid past. Apartheid, which is Afrikaans for apartness, was institutionalized racial discrimination and segregation, formally enacted into law in South Africa in 1948 and often repressively enforced. The medium chosen was a Truth and Reconciliation Commission (TRC) which was given legal status through an Act of Parliament in 1995. Such a course was not unique, as the examples of postwar Germany, Chile, Argentina, Haiti, and a handful of other countries illustrate. The emphasis in South Africa's case was to legitimize the moral repugnance of all the traces of apartheid, allow for a degree of justice for victims of human rights abuses, offer qualified amnesty to perpetrators upon individual request, and contribute to building national unity and reconciliation. The truncated past was defined from March 1960 to 10 May 1994, despite the fact that the practice and origins of the apartheid ideology go back centuries earlier.

The TRC itself conducted its work from April 1996 to July 1998 through three committees for human rights violations, amnesty, and rehabilitation and reparation, respectively. It was led by former head of the Anglican Church in South Africa and Noble Peace Prize laureate, Archbishop Desmond Tutu, and signed into law by former President Nelson Mandela.

Tutu was of course assisted by 16 other commissioners, regional officers, a research unit, an investigative unit, a legal department, and a media department. As it was given a two-year time frame to complete its operation, the TRC proceeded with haste in conducting mostly investigations and public hearings. It is noteworthy that only about 10% of the 21,000 statements submitted by victims were used for public hearings. Many victims felt cheated as they would come to a hearing only to learn that their story would not be heard. While for some the Truth Commission seemed to offer hope to form a new collective with those

previously disenfranchised (black), for others, it rendered the poor victims voiceless through a process of translation and contextualization. Mostly privileged black males were given time and space to tell their stories at the expense of women who, when given the opportunity, would speak of the sufferings of the voiceless generally. Ultimately, it seems that no redemptive balance was struck about which stories of survival, healing, defiance, torture, killing, or fighting would prevail in defining the TRC outcome.

MODE OF OPERATION

The founding legislation, which was signed by President Mandela on 19 July 1995 and which made the operation of the TRC possible, was the Promotion of National Unity and Reconciliation Act. The broad mandate defined in the act was as follows:

- To establish as complete a picture as possible of the causes, nature, and extent of the gross violations of human rights during the period from 1 March 1960 to the cutoff date.
- To facilitate the granting of amnesty to persons who make full disclosure of all the relevant facts relating to acts associated with a political objective and who comply with the requirements of this act.
- To establish and make known the fate or whereabouts of victims and to restore the human and civil dignity of such victims by granting them an opportunity to relate their own accounts of the violations of which they are the victims and by recommending reparation measures in respect of them.
- To compile a report providing as comprehensive an account as possible of the activities and findings of the commission, which contains recommendations of measures to prevent the future violations of human rights.^[1]

The three TRC committees were specifically tasked with functions that would enable the TRC to fulfill its mandate. The Human Rights Violations Committee conducted hearings throughout the country to give apartheid's victims a platform to tell their stories of oppression. On the other hand, the focus of the Amnesty Committee was on the perpetrators, allowing

them the possibility of seeking pardon for their individual misdemeanors. The Committee on Reparation and Rehabilitation, apart from making recommendations on such matters, also inquired into measures necessary for the prevention of human rights abuses and the creation of a new society free from racial segregation. Some of the work of the Amnesty Committee would later include outstanding matters of the other two committees when the TRC suspended operations.

MODEL OF MANAGEMENT

A country transitioning from authoritarian rule to democracy faces multiple challenges, one of which is how to deal with a repressive past that could potentially destroy its future. Hayner,^[2] an expert on truth commissions, believes that such commissions do have the capacity to give victims space to voice their suffering, promote reconciliation, prevent a repetition of past horrors, and allow for an acknowledgment of transgressions committed. The potential therefore of the TRC process in South Africa, which lasted more than 2 years, as a strategic link in the national management of conflicts needs to be stressed. This is especially in view of the fact that conventional institutions such as the courts, police, political bodies, or defense establishments are not viewed as being neutral enough in a climate of transition. Truth commissions, on the other hand, if they are generally constituted by men and women of honor, are more objectively able to be scribes of history, pursue claims of justice and truth, contribute to personal healing and reconciliation, and institutionalize the emerging culture of human rights than politicians or state bureaucrats. In hindsight, Tutu was convinced that South Africa would serve as an example to the world in matters of conflict management as the verities of hope, peace, and reconciliation had triumphed through the TRC process, despite the evil nature of human beings.^[3]

Less sanguine were some of his fellow commissioners who judged that the quota of truth to emerge from the TRC hearings to have been rather too small. Within the TRC, internal bickering often buttressed by issues of race proved an albatross as a few commissioners failed to complete their terms. Under tight financial constraints, and within a limited time frame, the TRC clearly succeeded in producing a five-volume report of its work that has laid the foundation for the new society to be built as a reconciled community. The role of the TRC leader was pivotal to the entire process, despite a strongly held view by some whites that the TRC would not be impartial and operate with an ideological bent. Members of the (black) ruling elite were equally grieved that the TRC “erroneously” determined that various acts of the liberation struggle

constituted human rights violations. Tutu lamented the absence of a white leader who would come forward and say, “We had an evil system . . . please forgive us.”^a He celebrated the fact that at last, after years of personal struggle, apartheid was unequivocally relegated in history as a crime against humanity. Little wonder then that Nelson Mandela, in his last opening address to parliament, announced that the “doors of the world have opened to South Africa, precisely because of our success in achieving things that humanity as a whole holds dear” (5 February 1999).

IMAGE AND PERFORMANCE

In reliving the memory of the past, the TRC was able at times to offer a theatrical reenactment of pain, as the case of Jeffrey Benzien illustrates. He used to be an investigator in the security branch of the South African police department and was requested by one of his victims to demonstrate the use of the wet blanket method of torture. For many, this was an “archival moment” that would become iconic of the whole TRC process itself. Another involved Desmond Tutu, leader of the TRC, when he broke down in tears after hearing the public testimonies of pain and anguish. Winnie Mandela, former wife of Nelson Mandela, captured the media spotlight by her appearance and confession, under coercion, that “things went horribly wrong” with her Mandela United Football Club that operated more as a vigilante outfit. Thus apart from documents collected, referenced, and preserved by the TRC, there exists the “unofficial” archive of the media images that remain extremely powerful in causing the memory of the past to be made real in the present.

Beyond the image of a legally constituted hearing, which would often lead to a judgment about amnesty, stood the sacred canopy invoking the sounds of solemn assemblies, where confession was elicited, hymns were sung, and prayers were offered often in the presence of Desmond Tutu who dressed in clerical garb as a rule. Performance at these hearings was therefore designed to achieve a cathartic effect, if not to facilitate the public expression of an individual or collective act of forgiveness. The TRC therefore as public performance sought to recreate the past by facilitating testimony that carried the intention of revealing truth and was thus seen as “a theatrical re-presentation of pain suffered and inflicted by victims and perpetrators of apartheid-era violence.”^[5] Not surprisingly, the work of the TRC became the subject of numerous artistic creations, perhaps provoking individuals to explore deep-seated memories in less traumatic

^aSee foreword by Ref.^[4].

ways. Most noteworthy was a literary classic by white poet Krog^[6] whose “Country of my Skull” won international acclaim. Her “fragmentary” style of writing seems to have been necessitated by the fact that here was a white Afrikaner writing with empathy about suffering and pain endured by blacks. Krog dreaded closure in the labyrinth of memory evocations, as the road ahead required the restoration of such memory of humanity itself.

AMNESTY AND RECONCILIATION

One of the foundational principles of the TRC process was that truth telling by perpetrators could be exchanged for amnesty, something at odds with international conventions. Apart from seeing this as “unconstitutional,” many believed that victims were being denied justice, as perpetrators would go free while the material welfare of victims was slowly being addressed. Because the process was slanted toward individual amnesty, there was little hope of the structural evils of apartheid being held accountable. Apartheid, being a crime against humanity, was to go largely unpunished as the TRC had chosen rather to forgive and forget those making contrite confessions. Many commentators would, as a result, accuse the TRC of sacrificing justice at the altar of truth to the detriment of apartheid’s victims. Furthermore, the TRC’s doctrine of reconciliation, while deeply Christian, was controversial. The three prerequisites for reconciliation are truth, justice, and reparation, but these were highly contested notions even among TRC commissioners themselves.

When it was achieved, reconciliation usually proved to be an individual experience during the public hearings rather than an overnight national achievement. The TRC had set in motion, through concrete example, a pattern of life for erstwhile enemies in a rejuvenated moral landscape. Frustration and bitterness might have been the lot for many, but the TRC, despite its “Christian baggage,” provided an important catalyst for the evolution of the new human rights culture. Hence for Tutu, the TRC is “at its heart a deeply theological and ethical initiative. For people of faith, the experience of honesty and mercy, confession and forgiveness, justice and peace, repentance and reconciliation is what truth and reconciliation are all about.”^[7] Prominent Christian leaders, as a result, saw themselves as being lead agents in a renewed ministry of reconciliation, one where the business of forgiveness was to prevail over the cry for vengeance. Of course, the agreements on amnesty and reconciliation were based on political and moral compromise between leading protagonists of the racial divide who both largely shared the same religious faith. The concept of “truth” that was to emerge from the TRC

process was thus qualified and vitiated by an impinging political and, perhaps, an historical thirst for the better future of South Africa.

CONCLUSION

The publication of the TRC final report did not signal the end of the historical transition as some had hoped. Political parties, including the ruling African National Congress (ANC), remain unconvinced about its pronouncements, especially its reluctance to cast moral judgments about their struggles. To have the TRC paint both victim and oppressor with the same brush of justice was an anathema. Considered in the modern paradigm of restorative justice in a transitional democracy, the TRC ranks very high as having set an international standard. In a situation where punishing perpetrators through criminal trials was not desirable, but rather national healing and reconciliation together with building a human rights culture, the TRC stepped forward with limited resources in a very restricted time and answered the call of leading a nation on the “freedom road.”^[8] That its work remains unfinished is also true. The TRC had instructed that all information collected should be held by the National Archives, but by September 2003, this was still undone. Numerous applications for reparations and amnesty are still being processed while the TRC remains officially in “suspension.” As and when these matters reach finality, it may be possible to begin delving deeper into the lasting significance of the TRC for South Africa and the world.

REFERENCES

1. *Truth and Reconciliation Commission of South Africa Report*; Juta and Co.: Cape Town, South Africa, 2003; Vol. I.
2. Hayner, P.B. *Unspeakable Truths, Confronting State Terror and Atrocity*; Routledge: London, 2000.
3. Tutu, D.M. *No Future Without Forgiveness: A Personal Overview of South Africa's Truth and Reconciliation Commission*; Rider Books: London, 1999.
4. Tutu, D. *TRC Report*; Vol. I–V.
5. Bester, R. *At the Edges of Apartheid Memory; Unpublished Paper*; 6.
6. Krog, A. *Country of My Skull: Guilt, Sorrow, and the Limits of Forgiveness in the New South Africa*; Random House: Parklands, South Africa, 1998.
7. Botman, R.H., Petersen, R., Eds.; *To Remember and to Heal: Theological and Psychological Reflections on Truth and Reconciliation*; Human and Rousseau: Cape Town, South Africa, 1996.
8. Asmal, K.; Asmal, L.; Roberts, R.S. *Reconciliation Through Truth, a Reckoning of Apartheid's Criminal Governance*; David Philip Publishers: Cape Town, South Africa, 1997.

Truthfulness and Lying

Roger Green

Division of Public Affairs, Florida Gulf Coast University, Fort Myers, Florida, U.S.A.

INTRODUCTION

The concepts of truthfulness and lying pose a number of significant concerns for administrators and other public officials. In recent years, the complexity of problems revolving around these concepts has heightened in the wake of demands for greater transparency in government and increasingly stringent reporting and accountability requirements. Traditional principles of representative democracy endorse transparency in government, holding that when public officials are less than fully truthful, they damage a polity's capacity for informed public scrutiny and defy the requirement that public officials be accountable to the sovereign democratic public. However, the dilemmas faced by public officials in a highly competitive pluralist or conflictive international environment oftentimes render the ethics of truthfulness and lying elusive.

Two approaches—one focused on individuals and the paradoxes of the “noble lie” in political life, and the other focused on structural secrecy and deception in complex organizations—play fundamental roles in the analysis of truthfulness and lying in the public sector. Regardless of one's preferred approach, however, it is crucial to note that the forms of deception, as well as their political and ethical implications, are greater in number and variety than is immediately evident in the basic concept of “lying.”

PUBLIC OFFICIALS AND THE “NOBLE LIE”

One of the paradoxes of modern politics is a widespread conviction that on some occasions the exercise of political responsibility (either to ward off public harms or to produce public benefits) may require a public official to engage in deception. An excellent example is provided by New York City Mayor Abraham Beame's decision to withhold information from, and at times to actively mislead, the public as to New York City's looming fiscal disaster during the early 1970s, motivated by a desire to avoid a potentially catastrophic dumping of the short-term city bonds that by that time had become essential to maintaining vital city services and avoiding municipal bankruptcy. Even in such circumstances, questions arise as to whether deception can be accommodated without

damaging the institutions and norms essential to a representative democracy. Moreover, even if a public official uses such methods and by doing so achieves politically praiseworthy results, securing or maintaining the public good, questions are likely to arise as to whether that official nevertheless has been compromised ethically. However ambiguous and difficult such questions might be, they are among the most fundamental ones a democratic public and its officials must wrestle with.

The problems underlying these questions are perennial ones in political philosophy, extending back through the centuries to the works of thinkers such as Plato and Niccolo Machiavelli. Plato, in *The Republic*, endorsed the strategic use of “noble lies” in order to shape the beliefs of people in ways conducive to an effectively regulated class structure and atmosphere of social harmony.^[1] Centuries later in *The Prince*, Machiavelli argued that effective statecraft and political leadership require one to separate the practice of politics from religious values in general and to make strategic use of seemingly unethical practices such as deception. In doing so, Machiavelli was not arguing that “anything goes” in politics, or that a political leader is justified in using whatever methods he pleases to add to his personal wealth and power. Instead, he was pointing toward what he regarded as a fundamental paradox at the heart of politics: that in order to create and maintain the type of stable and secure environment in which citizens can safely live their lives according to religious principles, a conscientious political leader sometimes must be willing to ignore those religious principles. Such a political leader cannot compete successfully against evil without using the weapons of those who are evil, including deception.^[2] In the contemporary political arena, similar justifications are offered on behalf of public officials' use of “noble lies” when faced with the exigencies of a conflictive and relatively disjointed pluralistic domestic arena, as well as when faced with the exigencies of an even more conflictive and near-anarchic international arena.

Philosopher Sissela Bok analyzed the various justifications oftentimes offered by public officials on behalf of such “noble lies.”^[3] These can be summarized as follows:

1. The public has a limited capacity for understanding the “truth” of complex policy issues.

2. The public has a limited capacity for facing the painful realities of politics.
3. Deception is sometimes needed by public officials in order to attain the requisite efficiency, speed, and decisiveness in the face of powerful obstacles or even public inertia.
4. Deception is sometimes warranted *temporarily* as a stage in the education of the public (e.g., “I’ll provide the truth to the public at a later point in time, at which point they will have become sufficiently sensitive to the political realities of the problems I face to appreciate the necessity of my deceptions”).
5. Deception is so pervasive in politics, and there are so many genuinely unscrupulous officials concerned only with self-aggrandizement, that any attempt at “standard-setting” or “line-drawing” is either a useless annoyance or a dangerous competitive disadvantage for a civic-minded public official.

However tempting such justifications might appear, the use of “Machiavellian” or “Platonic” deceptions by public officials is far more complicated ethically in a contemporary representative democracy than it would have been in the context of Machiavelli’s and Plato’s thought. In a representative democracy, citizens are the ultimate source of political authority, and public officials are merely their surrogates. Yet a Machiavellian or Platonic strategy of *principled* deceit would require that public officials sometimes be willing to deceive the sovereign public or their delegates if they felt that doing so could advance long-term public goals. This poses a serious paradox: how can the surrogates ethically justify deceiving the individuals—the public or the public’s delegates—in whose hands political authority rests? Owing to the seriousness of this problem, a strong *prima facie* position must be taken against public officials’ use of deception, and the burden of proof should rest on the shoulders of the public official who thinks it necessary to violate this general principle in a particular instance.

Propaganda

Deception on the basis provided by the “noble lie” or by Machiavellian prescription is far more variegated than might be evident in the concept of lying *per se*, displaying a number of key typological distinctions. One such type is that represented by the deceptive strategy of propaganda: one-sidedly highlighting or providing only those facts that are favorable to one’s policy preferences or administrative positions. Propaganda oftentimes proves an attractive strategy of deception: if a more “truthful” perspective on, or set

of facts relating to, an issue is discovered by potential critics or policy opponents despite efforts to frame the issue in an overtly one-sided way, public officials can proclaim that they did not lie or try to hide anything, but instead offered what was regarded as the appropriate interpretation.^[4] Despite its attractiveness as a strategy, however, propaganda carries with it some risks. In the long run, purveyors of propaganda all too easily become captives of their own propaganda, being forced to continue disseminating the same propaganda in the future (in order to maintain consistency and “face”), or else literally losing their ability to distinguish between truth and propaganda. Both of these maladies, for example, have been diagnosed by critics in assessments of the role played by official disinformation during America’s escalating and ultimately tragic involvement in Vietnam.

Secrecy

A second key type of deception is that of secrecy, or the withholding or active concealment of information. Secrecy also possesses attractive qualities compared to outright lying. A public official’s credibility is less at risk in the event of discovery than would be the case if outright lying had been used. Secrecy can be portrayed equivocally as an absence of candor, rather than stand out as an act of dishonesty and straightforward violation of a public official’s obligations to the public. In addition, secrecy allows a greater range of strategic variations and nuances than does lying. For example, while maintaining secrets, a public official might release selected information from time to time for various strategic purposes. Also, while maintaining secrets, a public official may offer selective access to otherwise hidden information as a way of co-opting potential dissidents or critics, requiring as a *quid pro quo* that they walk away from even more sensitive information the public official is unwilling to disclose.^[4]

However, a public official’s desire for secrecy oftentimes is rooted in genuine political imperatives. First, full and immediate transparency or publicity sometimes can damage policy making, and hence, the public good. Effective policy making (especially when dealing with controversial policy problems) often requires a tentative process of learning about, testing, and weighing controversial alternatives in private before publicizing the chosen policy. Conversely, full and immediate publicity might lead policy makers to avoid consideration of controversial alternatives and to merely “play it safe” or even to pander to the public. Second, secrecy is sometimes necessary to protect the innocent or vulnerable or to respect the privacy of citizens. And third, the threat of full and immediate transparency in many cases might lead administrators and

policy makers to avoid putting things in writing that really should be put in writing. Many public officials may begin to write an e-mail or memorandum on a politically or institutionally sensitive topic, only to reconsider at the last moment and opt instead for a telephone call or conversation behind closed doors. Although the fleeting character and impermanence of a telephone call or face-to-face conversation is acceptable in many cases, in other cases, the points covered and the agreements made during the conversation may have warranted permanent documentation. Still, all of these considerations notwithstanding, in a representative democracy, most commentators insist that public officials should respect the *prima facie* case against secrecy and in favor of transparency, while trying to identify special cases where secrecy may be ethically justifiable.^[5]

Lying

Lying, a third key type of deception, obviously carries substantial risk for the public official who uses it. If found out after such a wrongful *commission*, it is significantly more difficult for the official to find a credible hiding place than it is after the ethically questionable omission represented by propaganda or secrecy. Moreover, because lying is the most extreme form of disinformation, it clearly requires the most stringent ethical limitations on its use. Despite all of this, lying can be tempting to public officials, in part because lies can be so attractive to listeners. Because a liar may have a good idea as to what listeners would like to believe, whatever lies the listeners will be readily susceptible to can be strategically offered. Augment these temptations with the cross-cutting ethical ambiguities of “noble lies” and the oftentimes compelling arguments that can be offered on their behalf, and one is left with a situation in which public officials may be tempted to avail themselves of such lies with scant critical reflection.

In spite of the dim view she takes in general toward deception by public officials, Sissela Bok—whose catalog of common justifications for deception was summarized near the beginning of this article—is willing to accommodate deception by public officials under limited circumstances. For example, she argued that public officials should enjoy greater latitude when it comes to information about their private lives, but that secrecy or a mere withholding of information in such cases—at the very least, a terse “no comment”—is the most appropriate response, not outright lying. She also argues that temporary disinformation may be ethically warranted when the “truth” at a particular stage clearly might be damaging to the public good—for example, in the face of an upcoming interest-rate

decision by the Federal Reserve Board, premature notice of which could lend itself to currency and bond market manipulations (here again, however, it is instructive that Bok favors “secrecy” over “lying”). And finally, Bok accepts the deceit inherent in certain types of necessarily covert activities, such as undercover police “stings.” Nonetheless, she agrees powerfully with the idea that there should be a *prima facie* injunction against deception by public officials, as well as that idea that the burden of proof should be borne by any public official contemplating deception. In doing so, she insists on the following general standard: “only those deceptive practices which can be openly debated and consented to in advance are *justifiable* in a democracy.”^[3] In other words, the possible exceptions cited earlier in this paragraph are justifiable precisely because the public can openly discuss and consent to these within appropriate limits.

Such a strong *prima facie* injunction is warranted for several reasons. First, public officials contemplating “noble lies” can easily confuse “bad” motives and “good” motives when judging themselves, especially in cases where an initially justifiable deception gets followed by unjustifiable lies in order to cover up the original act of deception (this same problem can readily afflict propagandists and advocates of strict secrecy, rather than being unique to liars). Second, public officials often suffer from a tendency to overrate their own ability to make rational choices while underrating the comprehension of average citizens. Third and perhaps most important to critics such as Bok,^[3] Arthur Applbaum,^[6] and John Burke,^[7] however, is the possibility that public officials’ acts of political deception will have damaging cumulative effects, to such a degree that it is inadequate for public officials contemplating deception to evaluate their cases strictly on an individual basis. Each act of deceit, through the power of prior example and the competitive pressures it generates in a pluralistic environment, contributes to the increasing ease with which other officials will come to use deception, even in far less justifiable instances. Ultimately, this can have a long-term corrosive effect on the public’s trust in government.

Destruction of Information

Whatever one’s stance toward these initial three types of political deception and “noble lies,” the type of political deception most difficult to exonerate is a fourth one represented by the destruction of information, i.e., falsifying or destroying pertinent government documents and records. Doing so eliminates any possibility of a retrospective assessment of one’s actions by others, which is necessary to determine whether the conditions cited as justification for the

deception were, in fact, sufficiently compelling. By extension, this would also undermine the evaluation and enforcement of retrospective accountability, needed as a bare minimum for ensuring that democratic principles have not been ignored. In the absence of such a process, one could merely destroy the documents needed by others to assess the legitimacy of one's allegedly "principled" deceptions, insist to others that one's reasons were legitimate ones, and arguably short-circuit any further process of ethical judgment. If a public official truly believes that a situation is so serious as to warrant the deception of other public officials and the public, then that official should be willing to submit to a subsequent review of this deception. This subsequent review cannot be carried out effectively if the relevant documentation has been destroyed or falsified beyond repair, and it cannot be carried out if a public official in general fails to respect the right and ability of others to gather and present relevant information that is contradictory to his or her objectives. This imperative—analyzed tellingly by Dennis Thompson under the rubric of "retrospection,"^[8] and by Donald Warwick under that of "veracity,"^[9]—is perhaps the key one invalidating the "noble lie" justification offered by Oliver ("Ollie") North in the aftermath of the so-called "Iran-Contra" scandal, when he "shredded" or otherwise destroyed a substantial amount of documentary evidence once word was received that an investigation was underway.

SYSTEMATIC DECEPTION, STRUCTURAL SECRECY, AND BUREAUCRACY

The analysis thus far has focused on the political exigencies and ethical implications of *individual* acts of deception, which conforms to the approach most commonly found in deliberations on truthfulness and lying in public life. In contemplating such issues, most observers tend to adopt an approach centered on individuals, one that focuses on a desire to identify wrongdoers and to apportion blame among the various individuals implicated in an ethical transgression. As carried out in many cases, this approach appears to be premised on a vision of administrators and other public officials as rational, calculating, and largely prescient actors able to consciously weigh the choices before them. However, focusing attention on individuals without understanding the organizational environment in which the dilemmas of truthfulness and lying manifest themselves can blind one to the significant influences constituted by an organization's structure, culture, and external political environment. In particular, inattention to these broader environmental factors can obstruct an understanding of the *systematic deception* and *structural secrecy* contributing to incidents such as the Space Shuttle

Challenger disaster. An organization's structure, culture, and external political environment can lead to a systematic distortion or occlusion of important information in ways largely resistant to any individual's control, sometimes to such a degree as to generate organizational pathologies.

Organizations, Resource Scarcity, and Systematic Deception

According to writers who approach issues of truthfulness and lying through this second perspective, organizations should be viewed as networks of "fealty" and "patronage" between superiors and subordinates,^[10] and as arenas of horizontal competition between units or subdivisions striving to preserve or increase their resources in an environment of endemic resource scarcity.^[11] Such conditions can create powerful incentives—and often even unconscious pressures—to present strategically distorted information upward and outward, in order to create the image of one's organizational unit safest for protecting resources and for deflecting blame should problems arise.^[12] This is especially true of mid-level managers whose intermediate location within the organizational hierarchy, in combination with their aspirations to move upward, imposes an unusual mix of burdens and uncertainties.

In some cases, this problem may manifest in a pronounced emphasis on the dexterous but potentially misleading use of tentative, vague, or heavily euphemistic language in order to obscure discomfiting facts, preserve personal maneuverability, defuse criticism, and avoid being viewed as a bearer of bad news.^[10] In organizations with units that are segmented horizontally and vertically into divergent pockets of specialized expertise, this linguistic phenomenon may be reinforced through a reliance on technical references and terminology within reports that are resistant to complete understanding by anyone outside the group, as well as through a tendency for personnel occupying strikingly different roles within the organization to largely talk past one another and to project strikingly different "truths" on the same underlying body of data (in the Space Shuttle Challenger case, the often stark divide between "engineering" and "management" was most notable).^[11]

Structural Secrecy

In other cases, such problems may manifest in the form of upward or outward communications from which significant but potentially disquieting facts have been winnowed in the process of summarizing a diverse body of information. Ironically, recent trends toward more stringent reporting standards may inadvertently

exacerbate rather than ameliorate this problem. Frequent and standardized reporting all but require that information be condensed and generalized in a manner that lends to such winnowing. It may also eventuate in “information hoarding” by units that feel trapped in a competitive struggle within a capricious organizational and political environment, and that regard such tightly held information as a valuable hedge against future difficulties. The secrecy embedded in these phenomena is anchored in the structure, procedures, and protocol of bureaucratic organizations.^[11]

CONCLUSION

In such an organizational environment, even individuals keenly aware of the problems likely to result from systematic deception and structural secrecy may find it difficult to avoid putting on the “team face” and basing their judgments on situational rules of thumb, rather than face the potential dangers inherent in adopting a strong and consistent commitment to truth. Wearing the face of an unwaveringly ethical person risks alienating team members, because such ethical “rigidity” may be viewed as imposing limits on other team members’ maneuverability in the event of trouble.^[10] When contemplated through this organizational perspective, reducing the likelihood of deception in an organization depends far less on whether one scrutinizes and apports blame to specific individuals, and far more on whether one carefully dissects the organization’s underlying structure and reshapes it in ways less hostile to healthy information flows and to individual ethical reflection.

REFERENCES

1. Plato. *The Republic of Plato*; Basic Books: New York, 1968. Translated with Notes and an Interpretive Essay by Allan Bloom.
2. Machiavelli, N. The prince. In *The Prince and Other Works*; Hendricks House, Inc.: New York, 1964; 93–180. Translation, Introduction and Notes by Allan H. Gilbert.
3. Bok, S. *Lying: Moral Choice in Public and Private Life*; Pantheon Books: New York, NY, 1978.
4. Goodin, R. *Manipulatory Politics*; Yale University Press: New Haven, 1980.
5. Bok, S. *Secrecy: On the Ethics of Concealment and Revelation*; Pantheon Books: New York, NY, 1982.
6. Applbaum, A.I. *Ethics for Adversaries: The Morality of Roles in Public and Professional Life*; Princeton University Press: Princeton, NJ, 1999.
7. Burke, J. *Bureaucratic Responsibility*; The Johns Hopkins University Press: Baltimore, 1986.
8. Thompson, D. *Political Ethics and Public Office*; Harvard University Press: Cambridge, 1987.
9. Warwick, D.P. The ethics of administrative discretion. In *Public Duties: The Moral Obligations of Government Officials*; Fleishman, J., Liebman, L., Moore, M., Eds.; Harvard University Press: Cambridge, 1981; 93–127.
10. Jackall, R. *Moral Mazes: The World of Corporate Managers*; Oxford University Press: New York, 1988.
11. Vaughan, D. *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*; The University of Chicago Press: Chicago, 1996.
12. Gabris, G. Beyond conventional management practices: shifting organizational values. In *Ethical Frontiers in Public Management*; Bowman, J., Ed.; Jossey-Bass Publishers: San Francisco, 1991; 205–224.

Tuskegee Study

Holona L. Ochs

Department of Political Science, University of Kansas, Lawrence, Kansas, U.S.A.

Andrew B. Whitford

Public Administration and Policy, University of Georgia, Athens, Georgia, U.S.A.

INTRODUCTION

The Tuskegee Syphilis Study, which was conducted by the U.S. Public Health Service (PHS) through the Tuskegee Institute between 1932 and 1972, sought to examine the “effects of untreated syphilis on the [N]egro male.”^[1] The study began with at least one good purpose: to consider ways to improve the health of African–American males in the south. It ended with the broad recognition that its means violated the individual rights of study participants, that researchers in the study had been failed by their individual integrity, and that the organization conducting the study did not contain sufficient checks on professional practice to restrict unethical action (described below). Perhaps more importantly, it ended with the recognition that many professionals and decision makers throughout society who knew of this study did not take appropriate protective action (an official history of this study is located at: <http://www.cdc.gov/nchstp/od/tuskegee/>). This history, along with changes to federal regulations for research projects involving human subjects, document another lasting effect of the Tuskegee Syphilis Study: it changed federal research establishment, especially organizations such as the PHS and the Centers for Disease Control (CDC).

The Julius Rosenwald Fund, which was established in 1928 to promote the health, education, and welfare of African–Americans, provided many of the initial charitable contributions that funded research and treatment for African–American syphilitic subjects in the south.^[2] By 1931, the Depression had nearly depleted the Rosenwald fund, and attempts at actual treatment of patients ended. Following this, the PHS decided to use a group of 600 low-income African–American men in Macon County, AL, 399 of whom had syphilis, to determine if syphilis differently affected African–Americans. This began the longest, continued, non-therapeutic medical experiments on human subjects in history.^[3]

The policy significance of the Tuskegee study is the fact that the publicity it generated strongly contributed to the development of systematic standards for evaluating research programs involving human subjects. The administrative significance is that, using government authority, officials undertook unethical behavior

and implemented unethical policy. The long-term effect of this experiment has been to undermine trust among American minorities in both government and medicine. Of course, because this study was carried out using African–American doctors, nurses, and institutions, repairing this damage is even more complicated. The Tuskegee study remains a potent symbol of the government’s abuse of African–Americans in the United States, racism in science, arrogance of the medical community, and misconduct in research involving human participants.

A BRIEF HISTORY OF HUMAN SUBJECTS RESEARCH

Historically, non-consensual experiments have been regularly performed on captive populations in an institution, particularly when groups of people are seen as “less than human.” Examples of unethical research similar to the Tuskegee study include the following:^[4]

- Experiments conducted by Nazi physicians during World War II (WWII)
- Experimentation in 1952 on Harold Blauer at the New York State Psychiatric Institute in conjunction with the U.S. Army Chemical Corps
- An experiment conducted in 1953 on a premature infant at the Brooklyn Doctors Hospital without any attempt to obtain informed consent
- The Jewish Chronic Disease Hospital case of 1963 in which 22 chronically ill non-cancer patients were injected with cancer cells
- The injection of hepatitis in severely retarded children at the Willowbrook State Hospital in New York
- The Cincinnati radiation experiments on an entirely non-consensual group of cancer patients of below-average intelligence who were primarily African–American, resulting in acute radiation sickness, severe burns, and premature death
- The study conducted at the Veteran’s Administration West Los Angeles Medical Center, which was shut down in 1999 for experimenting both on subjects who had not consented and those who expressly refused consent.

The recurring theme is that latent tendencies toward dehumanization exist when any process becomes routinized.^[5] The scientific method is a routine, as is much of policy implementation through the application of standard operating procedures.

It has often been argued in the medical community that professional expertise and “compassion” of physicians for human subjects are a better way to protect the rights of patients than external constraints on their behavior, despite evidence that these individual qualities are often ineffective.^[3–6] Even with physicians, the effectiveness of an individual’s integrity depends on, and interacts with, professional ethical standards and institutional reward systems, as well as formal and informal positions of power and autonomy; such concern-producing conditions sufficient to protect the subject occur rather infrequently.^[7] The “compassion” or desire to make good decisions is often misguided or misplaced, particularly when the subjects are regarded by society as “less than human.” Dr. John R. Heller, the Director of Venereal Disease for the PHS from 1943 to 1948, claimed, even after penicillin was being used to treat syphilis in 1943 that, “The men’s status did not warrant ethical debate. They were subjects, not patients; clinical material, not sick people.”^[2]

BACKGROUND AND DEVELOPMENT OF THE TUSKEGEE STUDY

The purpose of the study funded by the Rosenwald Memorial Fund, a Chicago-based philanthropic foundation, was to track the prevalence of syphilis among African-Americans in the 1930s as a way of determining the practicability and potential effectiveness of mass control of syphilis. The project began in 1929 as a syphilis control program using the standard therapy of the time—heavy metals treatment.^[3] Conducted in six rural counties with diverse social and economic characteristics, this was the first attempt to systematically control venereal disease in rural areas.^[8] The study’s findings, which were reported at the 1932 Annual Convention of the National Medical Association at Howard University, were that socioeconomic characteristics had a greater impact on the prevalence of the disease than race, and that syphilis is not a racial disease.

Public health officials were concerned with remaining technical issues, such as the evaluation of blood tests for latent syphilis, which required 10–20 years before resulting in an outcome. The Rosenwald study did not resolve questions about the effectiveness of serological evaluations vs. autopsy examinations, and the general belief that syphilis was a “disease of small consequence to the Negro” persisted.^[9]

This is the environment within which the Tuskegee study took place. Officials’ low confidence in blood

tests for latent syphilis and the broad acceptance of racial hygiene theories contributed to pressures for a study centering on the observation of untreated syphilis in the latent stage through autopsy as a way of verifying the presence of syphilitic destructive lesions.

For the Tuskegee study, PHS selected Macon County, AL—the poorest of the six counties included in the Rosenwald study and the county with the highest syphilis prevalence rate (approximately 40%).^[9] The study involved 399 African-American men diagnosed with syphilis and a control group of 201 African-American men who did not have the disease. The men were told by the researchers that they had “bad blood”—a local term describing various common ailments. Informed neither of their actual diagnosis nor the nature of the illness, they were not provided proper treatment.^[3] The Tuskegee study disregarded virtually all of the principles of the Nuremberg Code. A copy of the Nuremberg Code is posted at the U.S. Holocaust Memorial Museum’s web site. Of particular importance are the unnecessary physical and mental sufferings and injuries, and the continuation of the experiments despite reason to believe that death or disabling injury will occur. The problem of syphilis was already solved, and the risk of death in the study was 100%. Participants were subjected to a painful spinal tap in the initial phase of the study, and researchers documented the untreated pain resulting from a curable disease in 399 men for 40 years.

Dr. Vonderlehr of the PHS began the experiment in 1932 by offering a painful lumbar puncture as a “special free treatment” that was actually a diagnostic procedure performed strictly for the benefit of the researchers.^[3] Following this examination, a letter was sent to each of the participants offering special free treatment for “bad blood,” and an African-American nurse from the area, Eunice Rivers, was used to convince the men to participate in the study. The participants consisted of rural farmers who owned their homes, renters considered permanent residents, and day laborers. Five hundred twenty of the original 600 men received consistent follow-ups to the point of autopsy.^[1] In exchange for participating in the study, the men were offered free medical exams, free meals on the days of the exams, free rides to and from the clinics along with an opportunity to shop or visit friends in town on the return trip, free treatment for minor non-syphilitic ailments, and burial insurance.^[1]

The program involved annual examinations by a “government doctor” in which the men were given a physical examination to document the effects and progression of the disease.^[1] They were also assessed for non-syphilitic conditions and provided with advice and medication (when it was available) for other ailments.^[1] The men were given mercurial ointment (which was known to be completely ineffective against

syphilis) and dosages of neoarsphenamine or arsenic in amounts short of what would be necessary to stop the infection under the false pretense that the drugs would cure the men of their “bad blood.”^[3]

The PHS began using penicillin to treat syphilis in 1943, but despite the safe and effective treatment history and general availability by 1953, penicillin was deliberately withheld from the participants.^[3] In fact, the PHS went to the extremes to prevent the study participants from receiving any form of treatment for syphilis. PHS distributed lists of names of participants to local physicians instructing them not to provide penicillin to the men involved in the study, and when about 50 of the men were drafted during WWII, PHS requested that the draft board not test or treat the men for syphilis; the board complied.^[3] In 1943, when the PHS established “Rapid Treatment Centers” across the country to systematically treat syphilis, no clinic was established in Macon County so as not to disrupt the study.^[3]

By not treating the men in the study and by not informing them of the nature of the disease, the PHS contributed to the spread of the disease to many women and infants who acquired congenital syphilis.^[3] As such, the PHS, the draft board, the Macon County Medical Society, the Alabama State and Macon County Boards of Health, the Tuskegee Institute, and all of the physicians involved violated Alabama law by failing to treat a communicable disease.^[3] As many as 100 of 399 participants are believed to have died from syphilis over the course of the study.^[10] In 1950, Dr. Wenger of the PHS noted, “. . . we have contributed to their ailments and shortened their lives.”^[3]

In 1968, Peter Bauxum, a PHS venereal disease interviewer and investigator, and others raised concern over the ethics of the Tuskegee study. The CDC responded by reaffirming the need for the study, and gained the official support of the local American Medical Association and the National Medical Association for the study’s continuation.^[3] The first news articles condemning the study appeared in 1970, and, in 1971, Congress held hearings to investigate the scandal. In 1972, the Assistant Secretary for Health and Scientific Affairs (ASH) appointed an Ad Hoc Advisory Panel to review the study, which concluded that the Tuskegee study was “ethically unjustified.” In November 1972, the ASH announced the end of the Tuskegee study.^[3]

IMPLICATIONS FOR PUBLIC POLICY AND PUBLIC ADMINISTRATION

The legacy of the study at Tuskegee has reached far and deep in ways that hurt our progress and divide

our nation. We cannot be one America when a whole segment of our nation has no trust in America.^a

The Tuskegee study remains a primary and widely cited example of research that lacked adequate protection for the participants. The study violated a number of ethical principles that are now regularly applied to human subjects research. The study used disadvantaged, rural black men to study the untreated course of a disease that is not confined to that population, placing an undue burden on that population and placing at risk a broad population. The study not only failed to minimize the risks to the participants, but increased those risks by depriving them of treatment long after it became available. The 600 men involved in the study were not afforded informed consent, even after the Nuremberg Code was written in 1946, codifying the need for voluntary consent. It may be that the lack of education of the participants made informed consent difficult; most were illiterate, many did not even know their last name, and the best educated participants had only completed 8 years of school. However, it is widely accepted that this disadvantaged position only accentuates the need for the protection that administrators have an obligation to serve.

Along with others, this study provided a broad impetus for federal regulations that guide the practice of research. The resulting policy requires research proposals to be reviewed by a human subjects committee; only research meeting the standards for the treatment of human subjects are approved by the committee and capable of being carried out. The Department of Health and Human Services now implements the federal policy for the protection of human subjects, and the policy is integrated into the procedural process of every public and private organization conducting research involving human subjects in the United States. Further, information about the federal human subjects research guidelines can be found at the United States Office for Human Research Protections (<http://ohrp.osophs.dhhs.gov/index.htm>).

The Tuskegee study also has implications for a lasting difficulty in the evaluation and analysis of medical policy and interventions: the concern among African-Americans that the scientific establishment does not fully address their concerns about the protection federal regulations provide them. “Distrust of the white-dominated medical community, either because of Tuskegee or the long history of subordination of blacks in this country at the hands of whites, obviously

^aThe formal apology issued by President Clinton on May 16, 1997, taken from Ref.^[11].

plays some role in the reluctance of black people to participate in clinical research . . . It is critical for black people to become involved in this type of research in order for skilled clinicians to determine if any measures can be taken to reduce the incidence of a number of diseases that befall black people at a far greater rate than whites.’’^[12] The hidden effect of Tuskegee is the inability of evaluators and analysts often to pull together the type of diverse and rich human information about the real effects of health policy across various segments of the population that are necessary to change policy outcomes in the long run.

CONCLUSION

Individual integrity without proper institutional constraints often results in misguided decisions. At least 16 reports on the Tuskegee experiments were published in reputable journals such as *Public Health Reports*, *Archives of Internal Medicine*, *Journal of Chronic Diseases*, *Milbank Fund Memorial Quarterly*, *American Journal of Public Health*, *Journal of the American Medical Association*, and *New England Journal of Medicine*.^[13] Yet, even though individuals with integrity who read these studies or participated in these experiments may have had compassion and concern for human subjects, that compassion did not result in the ending of the study for 40 years. However, even with external constraints provided by federal regulations for the protection of human subjects, a lack of individual integrity on the part of the researcher can result in unethical practice. Neither compliance nor integrity alone prompts ethical practice, but instead, serving the public interest requires that external institutional accountability and individual integrity reinforce one another.^[14]

REFERENCES

1. Rivers, E.; Schuman, S.H.; Simpson, L.; Olansky, S. Twenty years of follow-up experience in a long-range medical study. *Public Health Rep.* **1953**, 8 (4), 391–395.
2. Thomas, S.B.; Quinn, S.C. The Tuskegee Syphilis Study, 1932 to 1972: Implications for HIV education and AIDS risk education programs in the black community. *Am. J. Public Health* **1991**, 81, 1498–1506.
3. Jones, J.H. *Bad Blood: The Tuskegee Syphilis Experiment*; Collier Macmillan, 1981; 182 pp.
4. *In the Wake of Terror: Medicine and Morality in a Time of Crisis*; Moreno, J.D., Ed.; MIT Press: Cambridge, 2003.
5. Adams, G.B.; Balfour, D.L. *Unmasking Administrative Evil*; Sage Publications: London, 1998.
6. Barber, B.; Lally, J.J.; Makarushka, J.L.; Sullivan, D. *Research on Human Subjects*; Russell Sage Foundation: New York, 1973.
7. Lally, J.J.; Barber, B. The compassionate physician: frequency and social determinants of physician–investigator concern for human subjects. *Soc. Forces* **1974**, 53 (2), 289–296.
8. Public health service. In *Milestones in Venereal Disease Control: Highlights of Half a Century*; Publication PHS 515; U.S. Department of Health, Education, and Welfare: Washington, DC, 1957.
9. American Medical Association. Unraveling the tuskegee study of untreated syphilis. *Arch. Intern. Med.* March **2000**, 160, 585–598.
10. Jones, J.H. *Bad Blood: The Tuskegee Syphilis Experiment, New and Expanded Edition*; New York: Free Press, 1993.
11. Gamble, V.N. Under the shadow of Tuskegee: African Americans and health care. *Am. J. Public Health* **1997**, 87 (11), pp. 1176, 1773.
12. The reluctance of black people to participate in clinical medical research. *J. Blacks Higher Educ.* **1997**, 17, 33–34.
13. Yankauer, A. The neglected lesson of the Tuskegee study. *Am. J. Public Health* **1998**, 89 (9), 1406
14. Lewis, C.W. *The Ethics Challenge in Public Service: A Problem-Solving Guide*; Jossey-Bass: San Francisco, 1991.

U.S. Office of Personnel Management

Teva J. Scheer

University of Colorado at Denver, Highlands Ranch, Colorado, U.S.A.

INTRODUCTION

The U.S. Office of Personnel Management (OPM) is the central personnel agency for the United States' federal government. It was established in 1978 as a result of the civil service reform measures promulgated by the administration of President Jimmy Carter. At its founding, the primary goals of OPM were to serve as the President's primary advisor in personnel matters, to establish government-wide personnel policies, and to assist federal executives and managers by simplifying the personnel rules and regulations. Assessments of OPM's effectiveness after almost 30 years of existence are, at best, mixed.

WHY OPM?

In 1977, shortly after President Carter appointed him Chairman of the U.S. Civil Service Commission (CSC), Alan K. Campbell instituted an in-depth study, called the Federal Personnel Project (PMP), of the charges made by federal executives, managers, and employees that the complicated body of personnel rules and regulations were hurting morale and—even worse—sapping public confidence in the federal civil service system. The original civil service system had been established under the Civil Service Act of 1883, which had been designed to curb undue political influence over the civil service. The body of laws and regulations that were the focus of the PMP's analysis had been developed to support that goal and to create a neutral, merit-based system. However, after several months of study and consultation with some 800 organizations and individuals, the PMP concluded that the problems with the existing civil service structure were “so pervasive and fundamental that merely patching the existing system [would] not solve them.”^[1]

Among the most critical issues that the PMP identified were:

1. Managers were confused about what they could do without violating merit principles; the rules were so complicated that even good-faith compliance was difficult.
2. Excessive centralization of personnel hindered the decision-making prerogatives of line managers and required managers to prepare

voluminous justifications to obtain CSC approval of minor actions such as extensions of temporary appointments.

3. The “confusing array of regulations and procedures” made it difficult for managers to reward good performance, discipline employees, and remove employees who failed to meet performance or conduct expectations.^[1]

As a result of the PMP findings, President Carter submitted Reorganization Plan No. 2 of 1978 to the U.S. Congress. The Plan proposed to create two new agencies to replace the CSC. The OPM would be the President's principal advisor on personnel matters, and it would be responsible for government-wide personnel policy as well as for administering the CSC programs for recruitment, evaluation of agencies' personnel programs, training, health and retirement benefits, pay systems, background investigations, and training. Its overriding goal would be to simplify the system, decentralize authority to agencies, and give federal managers the tools they needed to manage. The CSC responsibility for adjudication of employee appeals would be given to the Merit Systems Protection Board (MSPB) so that OPM could concentrate on its mission as a management advocate. In response to Congressional concerns that the new OPM would be subjected to undue political influence, the PMP director testified that he was confident the new OPM director would be under no more pressure than the former CSC Chairman, he stated: “I think we ought not to be afraid of the President being associated with management, because he is the one who is being held accountable by the voters of this country for carrying out programs.”^[2]

On March 2, 1978, President Carter followed up on Reorganization Plan No. 2 by submitting to Congress a bill to reform civil service laws, which was entitled the Civil Service Reform Act of 1978 (CSRA). After lengthy hearings and numerous amendments to the original bill, Congress passed the CSRA. President Carter signed the Act on October 13, 1978. On January 1, 1979, the OPM became a reality.

ASSESSMENTS OF OPM'S PERFORMANCE

Many scholars and practitioners would say that OPM has failed to live up to the goals set forth in the CSRA.

The first OPM director, Alan K. Campbell, had a clear leadership vision for the new agency as well as the strong support of President Carter. However, the second director, Donald Devine, was appointed by President Reagan, who excoriated the government during his campaign. Devine replaced “OPM’s traditional management orientation with an unswerving emphasis on responsiveness...solely to political direction,” according to one observer (L. Lane, as quoted in Ref.^[3]). During Devine’s tenure, OPM began a long, probably irreversible, decline in agency budget, staff, and even mission; for example, during the 1990s, OPM jettisoned its role as the government’s primary training provider. Subsequent directors to Devine, while well-meaning, each changed OPM’s goals and priorities in response to their presidents’ individual policy directions.

In addition to introducing new personnel policies, each President had his own ideas about how to reform government. When Vice President Al Gore instituted the National Performance Review to “reinvent government,” OPM director James King led the way by further downsizing his agency and abolishing the *Federal Personnel Manual*. In sum, it appears that the director of President Carter’s PMP erred in his 1978 assessment that OPM would not be unduly affected by partisan policies.^[4,5]

By 1988, the National Academy of Public Administration had declared that there was a “quiet crisis” in the federal service. The government was failing to attract, develop, and retain the quality work force that it would need to meet the challenges of the next century. The Office of Personnel Management must lead the charge for solutions; but it could not do so unless the agency had a “clearly defined mission and the capacity to perform that mission.... The position of OPM director must be recognized as one of great importance to the long-term effectiveness of the new administration.”^[6,7] But a 2002 assessment pointed out that OPM’s influence had become so negligible that the Office of Management and Budget was leading President Bush’s competitive sourcing initiative, under which about 850,000 government positions will be subject to competition with private companies who can demonstrate their ability to perform the same work more cost-effectively.^[8,9]

On another critical front, numerous agencies have given up on the government-wide regulatory umbrella administered by OPM, and have successfully sought Congressional approval to opt out of key provisions of Title 5 of the United States Code. Key among these agencies are the Internal Revenue Service, the Department of Homeland Security, and the Department of Defense. Although OPM officially supports these initiatives, analysts are left wondering what role will be left for the central personnel agency if most federal

departments are granted the authority to design their own systems.^[10]

AND THE FUTURE?

In 2003, one scholar, who was himself a key member of the PMP staff that laid out the framework for civil service reform, regretfully described OPM as a “toothless Chihuahua” (Lee,^[10] quotation from Chester A. Newland). The agency is positioning itself to become a consultant to agencies rather than their regulator, but OPM’s loss of staff and expertise has left it competing with an array of private contractors, many of whom developed their own expertise while working for OPM. The question is whether the federal government still needs an agency to provide the leadership of a central personnel agency. The original goals of the Civil Service Reform Act of 1973 were to decentralize authority to managers and to reduce the burden of regulations and paperwork, so perhaps OPM’s mission is complete and the individual agencies can carry on, tailoring personnel systems to their own unique missions and needs. In sum, is there any reason for the continued existence of OPM?

The central question to be answered by Congress and policymakers in the executive branch is whether they perceive continued value in the existence of a government-wide personnel system. Should each agency have the flexibility of an independent corporation to manage its workforce? Or are there standards of selection, pay policy, and performance management that should apply to all federal employees—and for that matter, to the thousands of federal contractors who now perform work formerly limited to the “federal service”? Should employees be accorded due process before they can be separated, or would the government benefit more from a flexible, at-will termination system? Should federal employees continue to enjoy protection from political pressure? Should they all be held to the same standards of conduct and ethical behavior?

In sum, do policymakers support the concept of an American civil service? If the answer is affirmative, then there remains a critical role for OPM. The challenge for OPM will be to carve out for itself the strong leadership role envisioned in the Civil Service Reform Act of 1978—a role that many believe the agency has lost.

REFERENCES

1. Campbell, A.K. Statement before the Committee on Governmental Affairs, United States Senate, April 6, 1978. In *Civil Service Reform Act of 1978 and*

- Reorganization Plan No. 2 of 1978, Hearings Before the Committee on Governmental Affairs*; United States Senate, 95th Congress, 2nd session; U.S. Government Printing Office: Washington, DC, 1978; 85, 87–88.
2. Ink, D.A. Testimony before the Committee on Governmental Affairs, United States Senate, May 3, 1978. In *Civil Service Reform Act of 1978 and Reorganization Plan No. 2 of 1978, Hearings Before the Committee on Governmental Affairs*; United States Senate, 95th Congress, 2nd session; U.S. Government Printing Office: Washington, DC, 1978; 1117–1119.
 3. Ingraham, P.W. *The Foundation of Merit: Public Service in American Democracy*; Johns Hopkins University Press: Baltimore, 1995; 90 pp.
 4. Brook, D. Merit and the Civil Service Reform Act. In *The Future of Merit: Twenty Years after the Civil Service Reform Act*; Pffifner, J.P., Brook, D.A., Eds.; Woodrow Wilson Center Press: Washington, DC, 2000; 6–7.
 5. Ink, D. What was behind the 1978 civil service reform? In *The Future of Merit: Twenty Years after the Civil Service Reform Act*; Pffifner, J.P., Brook, D.A., Eds.; Woodrow Wilson Center Press: Washington, DC, 2000; 52–55.
 6. Zuck, A.M. The future role of OPM. *Bureaucrat* **1989**, 18 (1), 20–22.
 7. Kettl, D.F.; Ingraham, P.W.; Sanders, R.P.; Horner, C. *Civil Service Reform: Building a Government That Works*; Brookings Institution: Washington, DC, 1996; 10–11. (The authors state that the new authorities and flexibilities granted to individual agencies during the late 1990s and early 2000s have “often been granted without central supervision or a means to ensure that agencies pursue the central principles that have guided the system for more than a century. Neither has the Office of Personnel Management (OPM) redefined what its new role ought to be in a more flexible system.”)
 8. Strobe, L. 850 federal jobs to go to bid. *Washington Post* **2003**, May 30.
 9. Friel, B. The personnel touch. *Government Executive* **2002**, January, 47–51.
 10. Lee, C. Defense pushes bill to free itself from OPM personnel rules. *Washington Post* **2003**, April 29.

Uncertainty in Decision Making

William J. Mead

Consultant, White Plains, New York, U.S.A.

INTRODUCTION

In decision-making and problem-solving processes, it is obviously preferable to have complete, consistent, and accurate information. However, unfortunately, all too often, the information available to us is necessarily less than perfect. There may be missing or unknown elements that are of significant importance. The “facts” we have may or may not be fully correct, or may be ambiguous. Measurements and data may be less than precise. Some decisions involve complex systems, where what happens in one part impacts other parts, which can introduce complications. All of this points to the unavoidable reality that *uncertainty*, the lack of being sure we know all we need to know to make a good decision, is often a confounding issue.

OVERVIEW

Blau and Sinclair^[1] defined uncertainty somewhat differently, saying that it is the randomness of the external environment that cannot be changed but only managed. They likened this to the weather, where people may be aware of the prediction that rain is probable, over which they have no control, but they can manage their exposure by carrying an umbrella.

As Merry^[2] said, we feel a need for regularity and predictability in all we do, and we strongly seek order and constancy even in the face of unrelenting change. So, dealing concurrently with partial predictability and partial uncertainty tends to pose difficulties for all of us, on a psychological level and on a practical level.

The advent of computer technology has greatly improved our ability to collect and store data, but this has not necessarily enhanced the accuracy, precision, or completeness of the available information needed for decision making. Moreover, with today’s vast databases, it is frequently difficult to sort and find the desired information in the format that would be most useful. Therefore, today’s information technology is only a partial step toward reducing uncertainty.

March^[3] said that decision makers choose among the alternatives open to them on the basis of their expected consequences, but often, those consequences are not known with certainty. He used the example

of a food vendor at a football game deciding how much of each item he should take to the stadium. His actual sales of each item depend on many variables, such as the weather and the spectators’ interest in the particular game. His decisions must be made under uncertainty, and the results may or may not be favorable. The outcome may be pleasant or unpleasant for the vendor, and he is likely to suffer regret in any event at being left with unsold goods or having lost possible sales by having been too conservative. This same type of phenomenon often occurs in the more significant decisions required in public administration.

Kiel^[4] pointed out that uncertainty is inevitable in our private lives and in public management. Because of chaos and complexity in our nonlinear world, it is impossible to completely avoid unintended and unwanted outcomes of our decisions and actions. Clearly, this impacts policy making and the regulatory process. Therefore, it is essential to realize the constant pressure caused by uncertainty. The role it plays cannot simply be ignored. Instead, it must be understood, appreciated, and dealt with as best as possible. Good decisions *can* be made even in the face of uncertainty, but it is important to remember that some measure of doubt always lurks in our deliberations and plans.

Risk and uncertainty are related but not identical topics. Risk presupposes we can identify and quantify all of the factors in our decision making, so that we can assign probabilities, and we know the consequences of all of the possible outcomes, which enables us to take appropriate actions to minimize the possible losses. In contrast, under uncertainty, some of the outcomes and probabilities of their occurrence are unknown; therefore, it is much more difficult to arrive at a suitable decision.^[5]

Many of the major decisions faced in public administration are complex and difficult, including those involving changes in the natural environment, such as global warming, the fragility of the ozone layer, correction of flood zones, the control of acid rain, whether genetically modified crops are sufficiently safe, and many others. Issues of how best to control traffic flow in congested areas, and zoning to prevent urban sprawl, also involve uncertainty. In addition, there are public health decisions laced with uncertainty, such as whether to approve or ban certain pesticides, issues on control of air quality for outdoor and for indoor

air, whether or not electromagnetic fields from high-tension power lines or cell phones pose substantial risk, and whether specific new drugs should or should not be approved for use. These, and many similar matters, not only involve high levels of uncertainty, but they also tend to be politically loaded and may have high visibility to the public, often resulting in major disagreements between groups holding opposing views. It should then be obvious that decisions involving considerable uncertainty are difficult and tricky at best, and it is important to be familiar with some of the methods and techniques of approaching resolution of the issues.

MEANS OF COPING WITH LESS-THAN-ADEQUATE INFORMATION

Mathematical Computer Models

An important aid in decision making under uncertainty, in those instances where appropriate, is the use of mathematical computer models. This approach is relevant for highly complex topics, where information is still being gathered. Various modeling techniques and programming skills are used to simulate conditions. Models, by their nature, are incomplete descriptions of reality but are simplified approximations. But, even with this caveat, they are extremely useful in studying complicated interactions by varying certain parameters. For example, models are widely used in studying global warming, to assist in decision making regarding the steps necessary to curtail this phenomenon if it needs to be curtailed.

The use of models in this way is, of course, an outgrowth of advancements in information technology in recent years. Prior to the advent of high-speed computers, it was generally not feasible to make extensive use of models.

One practical problem with using complex models of social, economic, and environmental systems is that most people have difficulty understanding them, which tends to make policy formulation somewhat difficult.^[6]

Expert Judgment

In the absence of extensive and reliable data, it is not unusual to solicit judgement from qualified experts in the specific field. In some instances, multiple judgments are combined by using groups of experts. While subjective as opposed to quantitative, this approach is a logical effort to get useful input to help the decision makers with their tasks.

One approach is to split the subject matter into definitive variables and parameters, and then ask the

experts to indicate their views on the likelihood of events occurring and the probable consequences of these events if they do occur. This was one of the tools used in a report on climate change, from an international panel, for example.^[7]

When expert judgment is used, the methods of getting the opinions should be well documented by identifying the experts and why they were considered to be qualified, and by detailing how their judgments were obtained. If there is any significant disagreement among experts, this should be spelled out. Having these methods clearly documented allows a review of the process used and helps make it possible to repeat the exercise to show that the results are reproducible.

Although experts may be well qualified technically, they are subject to inadvertent human biases. Psychological research has identified certain inferential rules, known as *heuristics*, people tend to use to assimilate what they know about a given topic and to judge how likely specific events are to occur and the probable consequences if they happen. These “internal rules” can lead to biases. Therefore, it is important to compensate for these cognitive biases in judgment.^[8,9]

Statistical Decision Theory

There are a number of statistical tools that can be helpful in attempting to quantify a situation, even in the face of uncertain information. Collectively, the use of such tools is called *statistical decision theory*, which has become a separate discipline. Advanced degrees are now offered in this field, and a considerable body of literature has evolved.^[10–12]

The tools are based on the laws of probability, which can express the likelihood of an event occurring under specified conditions. In simple cases, such as flipping a coin or rolling dice, the likelihood of any given outcome can be known with precision. But, in more complicated scenarios, such likelihoods are not clear and can only be estimated in a subjective way. The mathematics involved can be quite daunting. But, regardless of the difficulties involved with these procedures, they provide information useful to decision makers, plus they provide a useful means of communication.

In addition to mathematical methods based on classical statistics, there is another useful variation on the theme, called Bayesian statistics. This approach is named after the founder of this school of thought, a Presbyterian minister named Thomas Bayes who first proposed the concept in the 1700s. This approach has gained prominence and rather wide acceptance in recent years, for dealing with uncertainty. In essence, in addition to the laws of probability, the Bayesian approach also considers the degree of belief the

decision maker has that a given event will or will not occur, based on knowledge of the matter as well as on any new information that has been acquired. Obviously, Bayesian statistics involves a certain amount of subjectivity, but this fits nicely with the concept of uncertainty. With the rebirth of interest in the Bayesian approaches, this method is now taught in various colleges and universities, and a body of literature has developed around it.^[6,13–15]

Decision Analysis

Another useful tool now rather widely accepted for dealing with uncertainty is called *decision analysis*. This uses a number of techniques for achieving its goal of focusing on the relative merits of the various alternatives open to the decision-maker.^[16] Skinner defined decision analysis as being a methodology based on a set of probabilistic frameworks that facilitate high-quality, logical discussions, leading to clear and compelling action by the decision maker.^[17]

There is now a Decision Analysis Society, with a web site^[18] offering bibliographies and reprints of papers on decision analysis and its techniques.

Monte Carlo Analysis

In applying probability theory to observed data, it is important to determine whether chance alone can account for the results of observations. If not, other factors, including system bias, may be involved. A probability distribution describes the range and frequency of data related to events or specific measurements obtained. Often, but not always, such distributions when graphed appear in the form of the familiar bell-shaped “normal distribution” curve. However, other collections of data result in nonnormal distributions, complicating the application of the rules of probability. Both where bias is suspected, and where multiple factors are involved (some of which may have non-normal distributions), a technique called Monte Carlo analysis may be used.^[19] The name comes from the famous gambling casino in Monaco, where games of chance exhibit purely random but predictable behavior. In rolling a pair of dice, all possible outcomes are known, but the results of any given roll are not. If the data on the outcomes of a large number of rolls show that the frequency of specific values fails to match what would be predicted by probability theory, it is obvious that some sort of bias exists. For example, one die, or both dice, may be inadvertently or deliberately imperfect in dimensions or weight distribution. This is one kind of application of the so-called Monte Carlo method.

Somewhat similarly, in complex studies where there are several variables interacting, it can be of importance to take into account the impact of each source of variability and uncertainty as part of a decision-making process. This requires identification of each of the important parameters involved, and then attempt to determine the probability distribution of each. Sometimes, this requires using some human judgment, when mathematical approaches are not feasible. In doing so, variability (true inherent heterogeneity in the data) is separated as best as possible from uncertainty (lack of full knowledge of the situation). Once these factors have been established, the next step is to calculate many (often tens of thousands) scenarios, selecting differing input values for each of the probability distribution curves. This generates one overall distribution curve for the output of the model. In other words, this is a method of analyzing the effects on the final state of the system of varying inputs, in the face of uncertainty. An example of the kind of situation where such methods are useful would be in studying the potential hazard of radon in tap water used in household showers, where the amount of radon in the water, the range of water temperatures, the frequency of showering, and the typical duration of each shower are all factors requiring consideration. Fortunately, software programs are available to handle the immense number of calculations involved with this sort of application of Monte Carlo analysis.^[20–22]

Decision Trees

Another of the tools is what is known as the *decision tree*. This is a technique of graphically expressing the various factors involved in a decision-making process, showing the various alternatives and expected consequences in a logical and relatively easy-to-understand format. It lays out the issues to facilitate seeing the choices, the possible outcomes, and the probabilities of each outcome. In other words, it is an effective way of organizing a summary of a decision-making situation. It is called a “tree” because the graphics resemble a tree, with major branches subdividing into smaller branches and then into twigs.^[23–25]

CONCLUSION

Uncertainty is unfortunately a very real factor in arriving at decisions in many instances. It is essentially unavoidable in some situations, particularly those involving complex and unclear issues, such as environmental protection, climate change, and health protection. It is of great importance for those in decision-making positions in public administration to recognize the role of uncertainty and to be familiar

with the methodologies available to help make good decisions even when uncertainty is a significant factor.

REFERENCES

1. Blau, G.; Sinclair, G. Account for uncertainty in new product development. *Chem. Eng. Prog.* June **2001**, 80.
2. Merry, U. *Coping with Uncertainty*; Praeger Publishers: Westport, CT, 1995; 16–18.
3. March, J.G. *A Primer on Decision Making*; The Free Press: New York, 1994; 6 pp.
4. Kiel, L.D. *Managing Chaos and Complexity in Government*; Jossey-Bass Publishers: San Francisco, 1994; 210–211.
5. <http://people.morehead-st.edu/fs/t.creaha/risk.htm> (accessed February 2001).
6. Morgan, M.G.; Henrion, M. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*; Cambridge University Press: New York, 1990; 83–85, 295–297.
7. Houghton, J.T., Ed.; *Climate Change 2001: The Scientific Basis*; Cambridge University Press: New York, 2001.
8. Kahneman, D.; Slovic, P.; Tversky, A. *Judgement Under Uncertainty: Heuristics and Biases*; Cambridge University Press: New York, 1999.
9. <http://www.uwinnipeg.ca/~epritch1/jdm.htm> (accessed July 2001).
10. <http://darwin.eeb.uconn.edu/eeb310/lecture-notes/decision/decision.html> (accessed July 2001).
11. <http://cobweb.creighton.edu/knudsen/decisiona/tsld002.htm> (accessed August 2001).
12. <http://www.labs.bt.com/library/papers/PAMIpaper/node12.html> (accessed August 2001).
13. Malakoff, D. Bayes offers a 'new' way to make sense of numbers. *Science* **1999**, 286, 1460–1464.
14. Rasenick, M.M. The Bayesian way. *Science* **1999**, 286, 2449–2550.
15. <http://www.bayesian.org/openpage.html> (accessed August 2001).
16. <http://web.usfca.edu/~villegas/classes/003-608/DecAnalysis/sld001.htm> (accessed February 2001).
17. Skinner, D.C. *Introduction to Decision Analysis*, 2nd Ed.; Probabilistic Publishing: Gainesville, FL, 1999; 13 pp.
18. http://www.bus.utexas.edu/~dyerj/DA_WP/index.html (accessed August 2001).
19. Kammen, D.M.; Hassenthal, D.M. Chapter 4, uncertainty, Monte Carlo methods, and Bayesian analysis. In *Should We Risk It?* Princeton University Press: Princeton, NJ, 1999.
20. Sobel, I.; Sobol, A.; Sobol, I.M. *A Primer for the Monte Carlo Method*; CRC Press: Boca Raton, FL, 1994.
21. Anon. *Guiding Principles for Monte Carlo Analysis*; EPA/630/R-97/001; U.S. Environmental Protection Agency: Washington, DC, 1997.
22. Rubenstein, R.Y. *Simulation and the Monte Carlo Method*; John Wiley & Sons: New York, 1981.
23. <http://mscmga.ms.ic.ac.uk/jeb/or/dectree.html> (accessed August 2001).
24. <http://www.dbmsmag.com/9807m05.html> (accessed August 2001).
25. <http://www.mindtools.com/dectree.html> (accessed August 2001).

Uncertainty, Concept of

James McBain

*Department of Social Sciences, Pittsburg State University,
Pittsburg, Kansas, U.S.A.*

INTRODUCTION

Uncertainty refers to a psychological state of agents when they do not possess the property of having sufficient epistemic justification for a belief. There are two basic ways to analyze the concept of uncertainty—epistemologically and decision theoretically. The former concerns the epistemic status of a belief and the latter concerns the knowledge of the probabilities of an outcome for a decision. Both analyses are important in public policy in that^[1] uncertainty is a feature of countless policies, and^[2] the way of thinking of this concept changes depending on whether we are asking questions concerning the future, present, or the past, theory or application.

PROPOSITIONS AND POSSIBLE WORLDS

While uncertainty is a psychological state of agents, it should be clear that it is not a property that agents have, rather it is a property that is ascribed to a belief that an agent has. Uncertainty attaches to the propositional content of a belief the agent has. In this light, it is more precise to say that propositions are uncertain (or certain), not agents.

While propositions are usually expressed as declarative sentences, which can bear truth and falsity, they should be conceived of as sets of possible worlds. Possible worlds can be understood as descriptions of the way the world might be or has been. In conceptualizing uncertainty, we must understand that uncertainty represents some set of possible worlds where the proposition does not hold. The proposition “Stella is going to vote today” corresponds to the set of worlds where Stella heads off to the voting booth to vote. If we were uncertain of whether Stella would be voting today, there would be some set of possible worlds in which Stella does not go to the voting booth or, once there, cast a vote.

The introduction of possible worlds raises numerous questions. Are possible worlds ontologically real? How are we to reason with possible worlds? Are the worlds finite or infinite? Although such questions are

important and answers to them necessary, we should not dismiss the necessity of using possible worlds in understanding the concept of uncertainty.

AN EPISTEMOLOGICAL ANALYSIS

In its common locution, the phrase “I am uncertain that p ” (where p is any proposition) refers to an agent lacking the level of justification needed to be certain in an absolute sense of the truth of the proposition. To say “Stella is uncertain who will win,” is to say that Stella does not have the warrant needed to be certain of the truth of who will win. More precisely, one is uncertain when one is justified in believing the proposition, but there is a possibility of the proposition’s being false. This reading of uncertainty is derived from the Cartesian tradition. According to this tradition, an agent is certain that p , roughly, when she is justified in believing p and there are no possible grounds to doubt p . While it is a matter of epistemological debate as to how to characterize those grounds, we can take it that if the grounds reduce the level of warrant one has for the proposition, she is no longer certain. Hence, to be uncertain is to be justified in believing p , but for there to be some ground the agent is warranted in believing, which added to the agent’s other beliefs, reduces or removes the warrant for p .

Such an analysis is important in understanding the concept of uncertainty in everyday situations as well as appealing to evidence concerning the past or the present (we will reserve discussing of the future for the decision theoretic analysis). Suppose an administrator needs evidence as to whether there is some statute, which would inhibit a certain course of action. Normally, such evidence would come from those who do research for the administrator. As such, the evidence would be testimonial in nature. The administrator would need a way of determining whether the testimonial evidence she is receiving is certain or uncertain. One such way would be to assess whether there is some possibility of the testimonial claim being false. Is it the case that the researcher reads the literature and

case studies accurately? Is there some possible precedence that was set at a later time than what the researcher was using as evidence? Essentially, the administrator would be asking whether there is a possible world in which the researcher was mistaken in her acquiring and reporting the information. If there was, then the administrator's belief is uncertain.

While many may dismiss an epistemological analysis as not being practical, in that such analyses invites skepticism about ever achieving certainty about a belief, we must be careful to recognize that many daily judgments concerning policies require answers to questions of uncertainty, which pertain to evidence that one has or is researching. When appealing to testimonial evidence, administrators should be careful to recognize possible situations (possible worlds) in which they should not take their beliefs to be warranted and hence would require further warrant.

A DECISION THEORETIC ANALYSIS

In decision theory, a judgment under uncertainty (sometimes referred to as a "judgment under ignorance") is a choice by an individual or a group between two and more alternatives (acts), which will produce an outcome, but (some or perhaps all of) the probabilities of the outcomes are unknown or meaningless. This is contrasted with a judgment under certainty, which is a choice between acts of which one is certain of the specific outcomes. Judgments under certainty are fairly easy to assess, in that all an agent needs to do is to determine the most desirable outcome and act accordingly.

Judgments under uncertainty should further be contrasted with judgments under risk, which are choices between alternative acts where the probabilities of the outcomes *are known* or treated as known. Choices between options of parlor games such as roulette, craps, and card games are common examples of judgments under risk. In such games, the agents know the probability of the possible outcomes and bet according to what they believe is best in light of their desires and/or preferences. It is sometimes also useful to treat certain decisions as judgments under risk when it is of pragmatic value to assign a definite probability to the outcome. Examples of such behavior are investment decisions, sports gambling, agricultural decisions, and marketing decisions. So, where it is possible to assign probabilities to all the outcomes, it is a decision under risk. If it is impossible to assign the probabilities to all the outcomes or it is meaningless to do so, it is a decision under uncertainty.

Hence, "uncertainty" in decision theory refers to the realm of judgments where the probability of one

or more of an act's outcomes is completely unknown or meaningless. To be more precise, a judgment under uncertainty is a choice among a set of acts ($A_1, A_2, \dots, A_n$), where the desirability of the act is contingent upon which possible outcome ($O_1, O_2, \dots, O_n$) prevails, but the probability of $O_1, O_2, \dots, O_n$ is unknown or meaningless.

Policy decisions are often made under uncertainty. The outcomes of decisions made by administrators may depend on a state of affairs whose actuality may not be known at the time the decision is or needs to be made. One such controversial example is the Precautionary Principle. The Precautionary Principle has been applied to domains such as the production and consumption of genetically modified foods, aeronautics, pollution, species extinction, public health, and others. The Precautionary Principle states that whenever some action increases the threat of harm to (or decrease in the safety of) human health or the environment, precautionary actions ought to be taken even if some causal relationships of the threat are not scientifically established. The Precautionary Principle advises to err on the side of caution when facing a decision under uncertainty, in which the outcome is possible harm to human and/or the environment. Given that the probabilities of the outcomes of decisions are unknown (which in this case means scientifically unjustified), whether to appeal to this principle qualifies as a decision under uncertainty. As this is a decision under uncertainty, questions as to whether to accept this principle would turn to questions of the value of human welfare, the immediacy of the possible threat, what technologies are available to cope with the possible threat, etc.

An analysis of the moral status of the Precautionary Principle is beyond the scope of the analysis, but what is important to note is the commonality of policies being decisions under uncertainty. For many (if not most) policy decisions, the concept of uncertainty is present as many policies are based upon scientific certainties or guaranteed future states of affairs. Policy makers hence must recognize when policies involve uncertainty and when the probabilities of the outcomes are known or are not meaningless.

CONCLUSIONS

The basis of this explication of uncertainty has been conceptual. To apply the concept of uncertainty to various beliefs and decisions, we must first have a more precise concept. From this foundation, we can venture to ask other questions concerning the use of uncertainty in policy decisions. What is the best way to represent uncertainty formally? What rules

are best to apply in decisions under uncertainty (MaxiMin, MiniMax Regret, etc.)? Once we understand the concept of uncertainty, it is possible to address such theoretical and practical questions more efficiently.

REFERENCES

1. Klein, P. *Certainty: A Refutation of Skepticism*; University of Minnesota Press, 1981.

2. Luce, R.D.; Raiffa, H. *Games and Decision: Introduction and Critical Survey*; John Wiley and Sons, Inc., 1957.

BIBLIOGRAPHY

1. Resnik, M.D. *Choices: An Introduction to Decision Theory*; University of Minnesota Press, 1987.
2. Weirich, P. *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances*; Oxford University Press, 2004.

Understanding the Basics of Refunding in the Municipal Bond Market

Jun Peng

*School of Public Administration and Public Policy, University of Arizona,
Tucson, Arizona, U.S.A.*

INTRODUCTION

Refunding of outstanding debt is probably the most important management activity after the debt has been first issued in the primary municipal bond market, as long as the issuer's debt-paying capacity does not deteriorate. Refunding, in a nutshell, means issuing new debt to retire outstanding debt. The primary purpose of refunding is to lower the borrowing cost.

Although there are key differences, the best analogy of refunding in the municipal bond market is mortgage refinancing in the residential housing market. When mortgage rate goes down, homeowners will take out a new mortgage loan to retire the old mortgage loan, which typically carries a higher mortgage rate, to reduce their monthly mortgage payment. This is, however, where the similarity between municipal bond refunding and mortgage refinancing ends. It is the differences between these two that make municipal bond refunding more complicated. To understand the major differences between the two, it is important to understand the concept of option first.

CALL OPTION AND ITS ROLE IN REFUNDING

In the financial market, option gives the seller of the option the right but not the obligation to engage in certain financial transactions. There are two basic options, call option and put option. Call option gives the issuer the right but not the obligation to purchase some financial products at a certain price and at a certain date, and it will be exercised only when the actual market price of the product is higher than the purchase price agreed upon in the contract. The put option gives the issuer the right but not the obligation to sell some financial products at a certain price and at a certain date, and it will be exercised only when the selling price agreed upon is higher than the actual market price. It goes without saying that such options do not come free of cost. Both issuers of call and put options have to pay a premium to have such kind of right.

It is the call option that is relevant in the case of municipal bond refunding. The vast majority of the municipal bonds are callable bonds. If a bond is callable, it is written in the official statement that the issuer can (although he does not have to) call back (or purchase back or redeem) the bond from the investors at a certain date in the future before the final maturity of the bond. The issuer, of course, has to pay a premium for having such a call option. Such premium is not paid up front when the bond is first sold but rather paid when the bond is purchased back from the investors in the form of call price. The call price is usually expressed as a percentage of the par value of the bond, such as 103% of the par value of the bonds. Therefore an investor with \$1000 worth of the bond can obtain \$1030 from the issuer when it is called back. The size of this call premium varies positively with the time period between call date and final maturity. The closer the call date gets to the final maturity date, the smaller the call premium will be. As a matter of fact, after some time period, the bond will be called back at par value. For example, if the bond has a final maturity of 30 years, it may be called back at 103% of the par value 10 years after the bond is first sold, whereas it will be called back at par 20 years after the bond is sold.

As mentioned earlier, the primary reason for refunding is interest cost savings. Therefore issuers will buy back old bond with proceeds from the new refunding bond only when the interest rate on the new bond falls below that on the old bond. This is where the call protection period comes in. The call option builds uncertainty into the financial planning of the investors who purchase the bonds. When the bonds are called back, investors are left with proceeds that will be invested at a lower interest rate, thus reducing their future investment income. To reduce that uncertainty, investors need some protection from very early redemption. Call protection stipulates how early the bond can be called (or redeemed). The call protection period usually has a 10-year span, which means that the bond cannot be called within the first 10 years. Call protection has an important impact on how refunding

is performed. Whether the refunding bond is sold inside or outside the call protection period determines how the refunding transaction is performed.

CURRENT AND ADVANCE REFUNDING

There are two kinds of refunding transactions: current refunding and advance refunding. In a current refunding, the proceeds of the refunding bond need to be used to redeem the old bond within 3 months after the refunding bond is issued. That means that current refunding should happen no later than 3 months before the last day of the call protection. In most cases, current refunding happens when the call protection period is over. When the interest rate drops, the issuer can issue refunding bonds at a lower rate to immediately retire the old bond. Therefore there is no need for investment of the refunding bond proceeds.

What happens when the interest rate drops and there are still several years before the call protection period is over? The issuer cannot do a current refunding because they cannot call back the bond within 3 months. To lock in the low interest rate even when they cannot call back the bond, the issuer needs to do an advance refunding. It is called advance refunding because the refunding bond is issued way in advance of the redemption of the old bond. It is this time lag that makes advance refunding a more complicated transaction because it involves the investment of the advance refunding bond proceeds and the coexistence of two bonds: the new refunding bond and the old bond (now also called refunded bond).

ESCROW ACCOUNT AND FEDERAL REGULATIONS

In an advance refunding transaction, the proceeds of the new bond are invested in an escrow account. The investment income is then used to pay the debt service on the old outstanding bond. At the end of the call protection period, the proceeds in the escrow account will be used to retire the old bond. Because the bond proceeds in the escrow account and the investment income will exactly match the debt service on the old bond till its redemption, the old bond is also considered escrowed to defeasance. Once the old bond is escrowed to defeasance, it is no longer considered a legal obligation of the issuer and will not appear in any financial statement of the issuer. Although there seems to be two bonds, there really is one bond, the new refunding bond, for which the issuer is responsible in terms of debt service payment.

Two aspects of the investment of escrow proceeds make advance refunding unique. One is the risk profile

of the securities to be invested in, and the second is the investment yield restriction. In order for the old bond to be fully considered escrowed to defeasance, the proceeds in the escrow account have to be invested in risk-free financial securities so that there is no doubt whatsoever about its ability to meet the debt service on the old bond. That usually means Treasury securities, which have no risk of default as they are backed by the full faith and credit of the U.S. government or other triple-A-rated fixed-income securities. This is where the yield restriction issue comes in. The yield restriction issue arises because of the tax exemption of municipal bond interest. As the interest on municipal bond is tax exemption, everything else being equal, the yield on municipal bond is lower than those on other financial securities including Treasury securities. Therefore the refunding proceeds in the escrow account will be invested at a yield higher than the yield at which the refunding bond is issued, leading to arbitrage profit for the issuer. Internal Revenue Service thus requires that the yield on the investment should be equal to, or restricted to, the yield on the refunding bond. Such yield restriction makes it very difficult to find the proper investment securities in the open capital market because such securities have to meet a certain profile regarding size, maturity, and yield. To accommodate the specific investment needs of state and local government issuers of advance refunding bonds, the Treasury Department created the State and Local Government Series (SLGS). These are Treasury securities tailored to the specific investment needs of each advance refunding transaction. Both the federal and local governments benefit from the SLGS, as the federal government can issue federal debt at a lower yield, and the local governments can easily meet their investment requirement without worrying about arbitrage.

Such convenience is not without its drawback. The dependence on SLGS for investment purpose renders advance refunding vulnerable to the vicissitudes of the financial market and the financial condition of the federal government. Whereas the yield on Treasury securities is slightly higher than that on the municipal bonds most of the time, it can actually fall below once a while. In 1998, when Russia defaulted on its debt, which led to a world financial crisis, investors sought safety in the Treasury securities. Such “flight-to-quality” substantially increased the demand for the Treasury securities, thus increasing the price and lowering the yield. The yield on Treasury securities fell below that of the municipal bonds, which did not enjoy the benefit of “flight-to-quality.” Advance refunding transactions came to a halt because the issuers would be forced to invest the proceeds at a yield lower than that on the refunding bonds. The federal financial condition matters because SLGS is part of the federal government debt and thus counted toward the federal

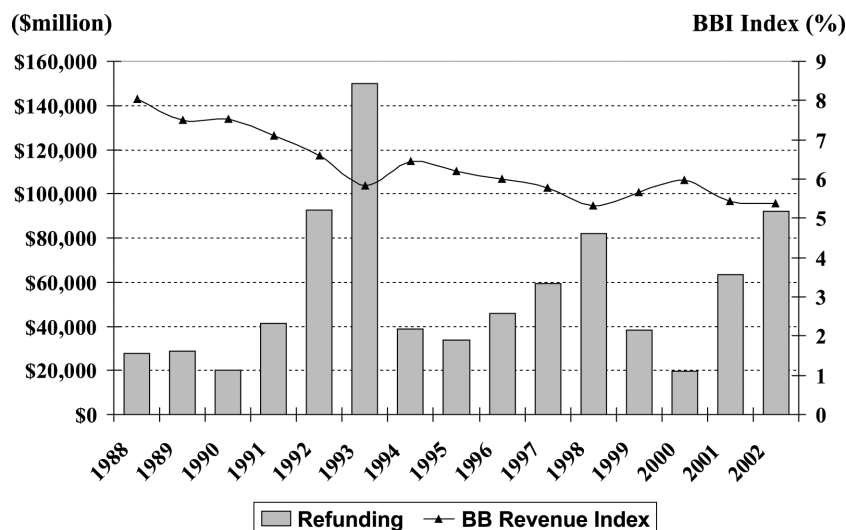


Fig. 1 Refunding volume and interest rate.

debt limit. When the federal debt limit is reached, which happens when the federal deficit worsens, as it did in fiscal year 2003 and again in 2004, the Treasury Department can no longer issue any debt. The Congress has to pass a bill to increase the debt limit before the Treasury can issue debt (including SLGS) again. Although the Congress never fails to do that, it takes time, and local government advance refunding becomes the victim during the waiting period.

While yield restriction is one federal regulation on refunding, the second federal regulation is that for each municipal bond, the issuer can advance refund only once because of the federal government's fear of overissuance of tax-exempt bonds if state and local governments are allowed unlimited advance refunding opportunities. This is another major difference between municipal bond refunding and mortgage refinancing. A mortgage loan comes with unlimited call options, which means that the homeowner can refinance his mortgage loan as many times as he wants. If the mortgage rate falls after he refinance for the first time, he can refinance again to take advantage of the even lower rate. State and local governments, however, cannot do that. Once they have refunded their bonds, they cannot refund them again even if the interest rate keeps falling. This happens between 2000 and 2003, when the Federal Reserve reduced interest rate many times to a modern-day low. Many issuers refunded their bonds when the interest rate fell early on in this period. When the rates kept falling, they could not do anything to further lowering their borrowing cost. Because of this, U.S. Senators Gordon Smith (Republican, Oregon) and Jon Corzine (Democrat, New Jersey) introduced S.271, the "Municipal Debt Refinancing Act," in January 2003 to permit states and local governments one additional opportunity to advance refund existing tax-exempt bonds within a 2-year window.

This bill is still being debated in Congress when this paper is submitted for publication.

OTHER PURPOSES OF REFUNDING

It is mentioned earlier that the primary reason for refunding is to lower the borrowing cost. This purpose can be seen in Fig. 1. The left vertical axis indicates the total national annual refunding volume between 1988 and 2002, whereas the right vertical axis is an annual average municipal bond interest rate index collected by the bond buyer.^a It is quite obvious that there is a very clear inverse relationship between refunding volume and interest rate.

Whereas lowering borrowing cost is the dominant reason, there are at least two other reasons that can also come into play. One important reason is to relieve pressure on government budget. Although lower borrowing cost is one way for immediate budget savings, restructuring debt service through refunding can also lead to immediate budgetary relief. In this case, the interest rate on the refunding bond may not necessarily be lower than that on the old debt. The primary objective here is to reschedule the debt service by lengthening the final maturity of the bond to lower the debt service payment right away. For example, after 10 years into a 30-year bond, the government decides to refund the old bond that has only 20 years left with a new 30-year bond. As long as the interest rates on the refunding and old bonds are fairly close, such restructuring does not lead to more debt service from a present value point of view, although the nominal amount of debt service will be higher for the refunding

^aData are collected from the Bond Buyer Yearbook.

bond.^b The worst-case scenario, however, is that the issuer is so desperate for immediate budget relief from debt service that it refunds the old bond with new bond that carries higher interest rate and longer maturity. Although such a move will lead to smaller debt service payment in the meantime, the issuer will pay substantially more in debt service over the life of the bond.

Another reason for refunding is to remove the restrictive covenants on the old bond. For various reasons, such as low credit quality, when the issuer first issued a bond, it put in place many restrictive covenants, such as requirement for debt reserve fund. Many years later, the issuer might be in a different financial situation and would like to remove these restrictive bond covenants. The only way to remove these restrictive bond covenants, however, is to retire the old bond, along with the restrictive covenants, with a new refunding bond that does not contain any of these restrictive covenants.

AN EXAMPLE

Finally, a basic example is used to illustrate the basic mechanics of an advance refunding transaction. City ABC issued a \$1 million 30-year bond at an interest rate of 6%. For reason of simplicity, the debt service is assumed to be level and paid annually. There is a call protection of 10 years. After 5 years, the interest rate dropped to 4%, and the city wanted to refund the bond at 4%. The annual debt service on the old bond was \$72,649, and the principal remaining at the end of the 10th year will be \$833,277.^c Therefore the total debt service on the old bond before its redemption will be five annual payments of \$72,649 and the final principal payment of \$833,277. The present value of this debt service payment stream discounted at 4% will be \$1,008,313, which also becomes the par amount of

the refunding bonds.^d This amount will be invested in an escrow account earning 4% to pay the debt service on the old bond before its redemption in another 5 years. The city will then only have to pay debt service on the new refunding bond. When the new bond is amortized over the next 25 years at the rate of 4%, the annual debt service on the new bond comes out to be \$64,544, resulting in an annual debt service saving of \$8105 for the next 25 years. The present value of total savings discounted at 4% is \$126,617.^e

One factor not considered in this example is issuance cost, which can cost up to 2% of the refunding amount, although it does not change the example in any significant way. Any reduction in interest rate does not immediately translate into an opportunity for refunding. There is cost involved every time when debt is issued, such as underwriting fee, legal fee, and credit rating fee, among others. It is like homeowners incurring closing cost when they refinance their mortgage. The interest rate has to fall to such an extent that at least the present value of future savings should cover all the cost associated with the issuance. Otherwise, it will be financially a net loss. It is probably caused by the concern that refunding should be approached cautiously and not be manipulated for budgetary purposes that some governments require that the refunding transaction can go ahead only if the present value of savings is at least 3% of the refunded bond. In the above example, this 3% rule is easily met even after deducting the issuance cost.

BIBLIOGRAPHY

1. "Advance Refunding" at <http://www.munibondadvisor.com/refunding.htm>.
2. Jordan, L. Understanding current and advance refundings. *Gov. Finance Rev.* **1992**, 8 (2), 13–15.
3. "Refinancing Municipal Debt" at <http://www.bergencapital.com/research/files/RefinancingMunicipalDebt.htm>.
4. Vogt, J. *Capital Budgeting and Finance: A Guide for Local Government*; International City/County Management Association (ICMA), 2004. Chapter 9.

^bThis is because the present value of future all debt service payments is equal to the face value of the bond, which is the same for the refunded and refunding bonds.

^cFinding out the annual debt service is a present value of annuity problem, using the formulae

$$PV = A \left[\frac{1 - \frac{1}{(1+i)^N}}{i} \right]$$

where PV is present value (or the par amount of the bond in this case), A is the annuity (or annual debt service in this case), *i* is the interest rate, and *N* is the number of years to maturity. Because we know PV (\$1 million), *i* (6%), and *N* (30 years), we can solve the problem and find A to be \$72,649.

^dIt should be noted that the refunding amount is always greater than the amount to be refunded. This has to be the case because the refunding proceeds in the escrow account are earning at a rate that is lower than that paid on the old bond. To make up for the difference, the principal of the refunding bond has to be greater.

^eFinding out the present value of total savings is also a present value of annuity problem, as explained in footnote³. The difference is that now we know A (which is the annual saving of \$8105), *i* (4%), and *N* (20 years). Solving for PV, we find the present value of total savings to be \$126,617.

Uniform Guidelines on Employee Selection Procedures

Teva J. Scheer

University of Colorado at Denver, Highlands Ranch, Colorado, U.S.A.

INTRODUCTION

The Uniform Guidelines on Employee Selection Procedures were issued in 1978 by five agencies of the federal government to assist employers in complying with the employment non-discrimination requirements of the Civil Rights Act of 1964. This study explains the requirements of the Uniform Guidelines and discusses its current status.

BACKGROUND

Following the passage of the Civil Rights Act of 1964, the Equal Employment Opportunity Commission (EEOC) issued two sets of guidelines—the first in 1966 and the second in 1970—to assist employers in complying with the Civil Rights Act of 1964 as it related to Title VII of the Act (i.e., non-discrimination in employment practices). In 1971, the Supreme Court recognized the authority of the EEOC guidelines in its *Griggs v. Duke Power Company* decision when it concluded that the guidelines were an accurate interpretation of Congressional intent in passing the Civil Rights Act.^[1] *Griggs* was a landmark case in the field of employment discrimination. The Court ruled that the Duke Power Company's use of intelligence tests and high school graduation requirements violated Title VII of the Civil Rights Act because these selection tools deprived Black applicants equal opportunity in hiring and promotional opportunities. The Court's majority decision stated, "What is required by Congress [in the Civil Rights Act] is the removal of artificial, arbitrary, and unnecessary barriers to employment when the barriers operate invidiously to discriminate on the basis of racial or other impermissible classification On the record before us, neither the high school completion requirement nor the general intelligence test is shown to bear a demonstrable relationship to successful performance of the jobs for which it was used."^[2]

The *Griggs* decision heartened employees and applicants who believed they had been subjected to employment discrimination. It put employers on notice that they could face similar legal challenges, and it galvanized those federal agencies with responsibility for civil rights' oversight and enforcement. With regard to

compliance with Title VII, however, *Griggs* raised as many questions as it answered. Employers questioned, for example, whether Title VII's coverage extended beyond written tests to other selection methods, how to determine whether a selection tool was discriminatory, and the standard that employers should use to validate their selection processes. By the mid-1970s, three other agencies had issued their own interpretive guidelines that differed in certain technical aspects from those promulgated by the EEOC. This confusing state of affairs was finally resolved in 1978, when five agencies—the EEOC, the Office of Personnel Management, the Department of Justice, the Department of the Treasury, and the Department of Labor's Office of Federal Contract Compliance Programs—jointly issued the Uniform Guidelines on Employee Selection Procedures.^[3]

THE UNIFORM GUIDELINES—SUMMARIZED IN PLAIN ENGLISH

The Uniform Guidelines were written by psychologists, after which they were finalized by enforcement attorneys. Not all the psychologists were happy with the final version. In sum, while they provided generally helpful guidance to psychometricians, they were so difficult to interpret and understand that few human resources specialists have ever read them. Regrettably, most HR staff who work in the equal employment opportunity field, or who regularly develop ranking and selection processes, have only the barest understanding of this important document.

In a nutshell, here is the basic principle of the Uniform Guidelines: "The use of any selection procedure that has an adverse impact on employment decisions for any race, sex, or ethnic group will be considered discriminatory unless the procedure has been validated in accordance with the Uniform Guidelines."

An explanation of each part of this principle follows:

1. "The use of any selection procedure" refers to any single selection device or combination of selection devices that an employer uses for hiring or promotion decisions. These include written tests, performance tests (e.g., typing, driving, physical performance), interviews, evaluations

of experience and education (e.g., application form ratings), qualification standards, certification or licensure requirements, probationary periods, supervisory appraisals if used for selection, and training periods if used to determine future advancement.

Not included in the list of selection devices are bona fide seniority systems or recruitment procedures designed to attract applicants of groups under-represented in the agency or company workforce. The Uniform Guidelines also specifically exempt discrimination cases based on the Age Discrimination in Employment Act of 1967, as amended, or on sections 501, 503, and 504 of the Rehabilitation Act of 1973.^a

In 1982, the Supreme Court clarified in *Connecticut v. Teal* that even if an overall selection process has no adverse impact on a protected group, the process is still discriminatory if any one element of the process (e.g., a test or an interview) does cause adverse impact.

2. “Adverse impact” is defined using the “4/5s” or “80 percent” rule. The rule states that a selection rate for any race, sex, or ethnic group that is less than 4/5s (or 80%) of the rate for the group with the highest rate will generally be regarded as evidence of adverse impact.^b In other words, the Uniform Guidelines hold that selection rates for all race, ethnic, and sex groups in the applicant pool should be generally the same—but the guidelines allow for chance variation of up to 20%.

Table 1 displays two examples of 4/5s-rule computations. The first shows a comparison of male and female applicants. Thirty men, or 50% of all the men who applied, were selected, compared to 10 women, or 25% of all female applicants.

^aOn March 30, 2005, the Supreme Court issued its decision in a case brought under the Age Discrimination in Employment Act, *Smith v. City of Jackson, Mississippi*. The Court held that under this Act, adverse impact based on age is cognizable (i.e., it falls under the jurisdiction of the courts). However, the Court also held that the scope of age-based adverse impact determinations is narrower than under Title VII cases, which govern other types of discrimination. Age-based adverse impact is not unlawful if the employer can demonstrate that the practice or procedure that has caused the adverse impact is “based on reasonable factors other than age discrimination.”

^bHowever, the 4/5s rule is a guideline only. The Uniform Guidelines also state that “smaller differences in selection rate may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms or where a user’s actions have discouraged applicants disproportionately on grounds of race, sex, or ethnic group. Greater differences in selection rate may not constitute adverse impact where the differences are based on small numbers and are not statistically significant, or where special recruiting or other programs cause the pool of minority or female candidates to be atypical of the normal pool of applicants from that group.”

Table 1 Two examples of the 4/5s rule

	No. of applicants	No. of selectees	Selection rates (%)
<i>Example 1</i>			
Males	60	30	50
Females	40	10	25
<i>Example 2</i>			
Blacks	40	10	25
Whites	50	10	20

Because the female selection rate is half that of the males—far less than 4/5s—the selection procedure has adverse impact on women. The second example compares the selection rate for Blacks and for Whites; their selection rates are 25% and 20%, respectively. As the ratio of 20/25 is 4/5, there has been no adverse impact.

3. “Employment decisions” include but are not limited to hiring, promotion, demotion, membership (for example, in a labor organization), referral, retention, and licensing and certification to the extent that the licensing process is covered by federal equal opportunity law. Other employment decisions, such as selection for training and transfer, may also be considered employment decisions if they lead to any of the other decisions listed in this bulletin.
4. The protected groups are specified in Title VII of the Civil Rights Act, which states that employers may not discriminate in employment decisions on the basis of race, color, religion, sex, or national origin.
5. A selection procedure is “valid” under the Uniform Guidelines if it predicts job performance, i.e., it distinguishes between the best and poorer performers. The Uniform Guidelines define three acceptable types of validity studies. If a selection procedure meets the validity requirements of the Uniform Guidelines, it means that the procedure can be used to predict future job success, i.e., to identify the best-qualified candidates.

CURRENT STATUS OF THE UNIFORM GUIDELINES

A 2001 symposium in the *Review of Public Personnel Administration* questions why the 1978 Uniform Guidelines have never been updated, despite advances in psychology, the passage of the Civil Rights Act of 1991, and the evolution of case law during the 1980s and 1990s. The EEOC initiatives to revise the Guidelines in the mid-1980s failed in light of opposition in

Congress and the Reagan administration. One of the symposium authors, Christopher Daniel, argues persuasively that the Guidelines have become a lightning rod in the controversies over affirmative action and preferential treatment for minorities. Although the Guidelines are in critical need of an overhaul, it is almost certain that political considerations would intervene.

Daniel concludes, “Revising the Guidelines would probably require a presidential administration’s active support. Attention would need to be devoted to the matter, controversy would be endured, and expenditure of some “political capital” would be required. It is unusual for presidential administrations to sacrifice simply to make regulations more useful to managers and test developers. Therefore, the Guidelines may never change.”^[4]

ARTICLE OF FURTHER INTEREST

Validation, p. 2024.

REFERENCES

1. 401 U.S. 424 at 433–434.
2. 401 U.S. 424 at 431.

3. *The Uniform Guidelines*; United States Code of Federal Regulations, Title 29, Part 1607.
4. Daniel, C. Separating law and professional politics: the Uniform Guidelines then and now. *Rev. Public Personnel Adm.* **2001**, 21 (3), 175–184.

BIBLIOGRAPHY

1. Civil Rights Act of 1964. Pub. L. No. 88–352, 78 Stat. 241, 1964.
2. Questions and answers to clarify and provide a common interpretation of the Uniform Guidelines on Employee Selection Procedures, Equal Employment Opportunity Commission, 44 Fed. Reg. 11996; U.S. Government Printing Office: Washington, DC, March 2, 1979.
3. The Uniform Guidelines on Employee Selection Procedures: Hearings before the Committee on Education and Labor, Subcommittee on Employment Opportunities, United States House of Representatives, 99th Congress, 1st Session, U.S. Government Printing Office: Washington, DC, 1985.
4. *Uniform Employee Selection Guidelines: Interpretation and Clarification* (Questions and answers), Equal Employment Opportunity Commission, 45 Fed. Reg. 29350; U.S. Government Printing Office: Washington, DC, May 2, 1980.

Unit Determination

Jonathan P. West

Department of Political Science, University of Miami, Coral Gables, Florida, U.S.A.

INTRODUCTION

A bargaining unit decision necessarily precedes labor-management collective bargaining, because it helps to determine with whom the employer must deal in negotiations. Specifically, it is the process of grouping certain jobs and employees for the purpose of representation. Statutory requirements, administrative agency decisions, union and employer agreement, an arbitrator, or, less frequently, the courts can determine bargaining units. At the federal level, where 2218 labor union bargaining units exist, the Federal Labor Relations Authority (FLRA) makes such determinations. In the private sector, the National Labor Relations Board (NLRB) sets standards for determination of the bargaining unit. At the state level, it is often the Public Employee Relations Boards (PERBs) that make these decisions.

The process of unit determination and certification of an exclusive bargaining agent typically involves five steps. These include filing a petition for a representation election, “showing” by the union that a certain percent (e.g., 30%) of workers in the unit prefer representation, validating the petition at a hearing, designating the unit by administrative action, and certifying the unit.^[1]

Unit determination is important, because the composition and size of the unit may affect bargaining success. The composition of the bargaining unit influences not only the parties who engage in negotiations, but also the types of demands and concessions made in negotiations and the rules affecting employee work and supervisory actions. Depending on the nature of the bargaining unit, it might be easy or difficult to gain majority support, recognition, or certification as the exclusive bargaining agent for employees. Unit determination also affects the scope of topics considered in negotiations as well as the probability of successful dispute resolution, contract ratification, and contract administration. If the bargaining unit is of limited size and scope, the range of topics subject to negotiation will be similarly restricted. If the unit lacks cohesion, it will reduce its leverage in negotiations and its ability to implement the contract.^[2,3]

SIZE AND COMPOSITION OF THE BARGAINING UNIT

It is not uncommon for employers and unions to differ over the size and nature of the bargaining unit. From the employer’s perspective, large units are often preferred. This is because it facilitates cost efficiencies when a single agreement is negotiated with a standard package rather than multiple separate bargaining contracts, each with its own unique provisions. A proliferation of units involves more negotiating sessions, heightens the probability of disruption, and adds complexity of multiple working rules and personnel practices. Employers may believe that smaller units increase the likelihood that perverse union tactics may be used, including end runs (circumventing management negotiators by going over their heads to elected officials or outside parties), leapfrogging (unions seeking bargaining gains greater than those achieved by other unions), and whipsaw bargaining (gains by one union used to justify benefits for another).^[4]

Employees, by contrast, usually favor small units. Smaller, more homogeneous units maximize opportunities for employee participation, may better reflect the needs and objectives of union members, amplify their voting power, foster greater solidarity, and are easier to organize. Larger units with more diverse constituencies add to the difficulty of fairly representing all employees’ interests (specialized interests might be ignored in larger units) and can become fractionalized, further complicating ratification and implementation of agreements. However, in certain situations, these concerns may be overridden by an employee organization’s desire to increase its leverage with management and maximize the numbers of eligible employees for the unit.^[1–3,5]

There are strategic considerations that influence the calculations of employers and employees regarding unit size. The union will often support a unit that it feels most capable of organizing. Those employers who are resistant to collective bargaining would likely prefer a unit that the union would have difficulty winning. If collective bargaining is inevitable, management may support larger units, because it can capitalize on

disagreements within the employee group (divide and conquer) to gain a bargaining advantage. When competing unions are simultaneously seeking recognition for a group of workers, they will each seek to define the unit—a narrow versus broad configuration—in ways that reflect their organization's strength.^[2] In other words, all parties are seeking their strategic advantage in the crucial decision of unit determination.

A unit determination decision is not only potentially contentious but also complicated. This is because the criteria or standards used in making unit determinations are vague and subject to multiple interpretations, despite considerable guidance from diverse sources. Employees, employers, and the public may differ in their perceptions of what constitutes an “appropriate” bargaining unit. The issue can confront a labor relations agency on different procedural matters, including certification or initial organization (no previous history of collective negotiations), decertification or severance (a preexisting unit with a splinter group wishing to bargain separately), accretion (a preexisting unit absorbing a new group of employees into the employer), or unit clarification (inclusion of a new or changed job in a preexisting unit).^[2]

CRITERIA FOR UNIT DETERMINATION

The criteria used to determine the composition of the bargaining unit usually conforms to NLRB guidelines. While there is variation in criteria among the states, they typically include the following:

- **Community of interest**—This is determined by several factors, including common duties, conditions of employment, skills, kinds of work, job classifications, career ladders, geographic location, and supervision.
- **Bargaining history**—This alludes to prior labor-management relations in the jurisdiction, including patterns of cooperation, negotiation, and representation. If there is a prior history of bargaining, it is likely that the preexisting structure will continue.
- **Unit size**—The unit needs to be large enough to facilitate effective bargaining and to avoid proliferation of small units that are costly for government, distracting to negotiating parties, and time consuming.
- **Effective dealings and efficiency of operations**—This seeks to avoid creation of units that are incompatible with administrative groupings and disruptive to existing human resource policies and practices.
- **Exclusion of supervisory or confidential employees**—This factor assumes that supervisors are more appropriately grouped with management

and should not be allowed to bargain with rank-and-file employees.^[6]

The most common standard used is “community of interest.” But, this term is ambiguous and subject to different interpretations. It can refer to commonalities within occupational groupings (similar work rules), within departments (same personnel practices and work environment), within functional units (common wage and benefit structure), and within the competitive service (uniform regulations). Community of interest can also be used to group employees with a common status or skill (e.g., profession), with distinctive characteristics identified by public policy (e.g., supervisor exclusion), and with a common pattern or history of bargaining (e.g., a workable relationship).^[2]

The community of interest criterion to be considered in unit determination is often specified in statutes with terminology such as “geographical location,” “avoidance of fragmentation,” “need for efficient administration,” “recommendations of the parties,” “requirements of standardization,” “desires of the employees,” “extent of employee organization,” and so forth. Some of these terms would likely lead to units that cross departmental lines (e.g., requirements of standardization), others suggest grouping workers at the same work site (geographic location), and still others allude to units attuned to the needs of employers (avoidance of fragmentation) or employees (extent of employee organization).^[7]

In forming bargaining units, different standards may apply. For example, a distinction is sometimes made between an “appropriate unit” and “the most appropriate unit.” The NLRB originated the “appropriate unit” test, and certain states use it as well. The “appropriate unit” standard can result in unit proliferation, but it leaves more discretion to employees in forming their own units based on community of interest and ease of organization. Where the “most appropriate unit” standard to represent community of interest is applied, it leads to a smaller number of units for employers to deal with, while reducing discretion of employees to choose groupings and complicating unions’ efforts to organize and win representation elections.^[7]

UNIT DETERMINATION AND SUPERVISORS

In the public sector, unlike the private sector, the decision of whether to exclude supervisors from a bargaining unit can be confusing. In the private sector, a federal statute and government agency defines “supervisor” and makes determinations about whether a job qualifies as supervisory. Supervisory jobs are then excluded from the bargaining unit. In the public sector,

where such distinctions are less clear cut, different jurisdictions handle the issue in different ways. Thus, the issue of whether to include supervisors in the same unit with non-supervisory personnel is a recurring one in the public sector. As Nigro and Nigro pointed out, three separate questions are involved regarding the status of supervisors and unit determination: Is the position in question really supervisory in nature? Is it desirable to create units composed of supervisors and non-supervisors? Should supervisors be allowed to form their own units?^[8] Regarding the first question, certain state laws make the distinction between bona fide supervisors (excluded from the bargaining unit) and less-than-bona fide supervisors (permitted in the employee unit), but even here, these distinctions are difficult to make, and there is considerable variation among the states on whom to include or exclude. At the federal level, supervisors who exercise considerable discretion in hiring, directing, rewarding, or disciplining employees are excluded.^[9]

The desirability of grouping supervisors and non-supervisors in the same bargaining unit is a subject of dispute as well. Generally, unions support a policy of inclusion, because they are interested in expanding the scope of union membership; employers, by contrast, may want supervisory personnel excluded because of their managerial responsibilities.^[5] An example highlights the problem: In a public university, should a department chair be included in the bargaining unit with the faculty? This issue has been hotly debated, because on the one hand, the department chair is a faculty colleague; on the other, the chair is a representative of management. This issue has typically been resolved in favor of including the department head in the faculty unit, while excluding other administrators such as president, vice president, or dean, but there is variation among the states on this matter.^[10]

This issue arises in other educational and departmental settings as well. For example, in secondary schools, should the principal be placed in a unit with the teaching faculty? In police and fire departments, it is not uncommon to find bargaining units that include all personnel except the chief. This can lead to conflicts of interest when grievances are filed, because the grieving employee and the supervisor may be part of the same bargaining unit. This puts the steward in a difficult position, because the steward's role is to represent everyone in the unit, in this case, both parties to the dispute.^[11] The supervisor

also faces a dilemma when belonging to the same unions as a subordinate and has to exercise supervisory responsibility (e.g., discipline) over these subordinates in the bargaining unit.^[12]

Disagreements also surface over whether supervisors should be permitted to form their own bargaining units. Absent such permission, supervisors may feel disadvantaged vis-à-vis others who have bargaining rights and, thus, greater leverage to pursue their interests. However, those who oppose separate units argue that supervisors should identify more with management, and they fear that supervisors would instead align with organized labor in workplace disputes.^[8]

REFERENCES

1. Kearney, R.C.; Carnevale, D.G. *Labor Relations in the Public Sector*; Marcel Dekker: New York, 1998.
2. Kurtz, J.P.; Barr, M.L.; Denaco, P. Bargaining units in the public sector. In *The Evolving Process—Collective Negotiations in the Public Sector*; Helsby, R.D., Tener, J., Lefkowitz, J., Eds.; Labor Relations Press: Fort Washington, PA, 1985; 96–122.
3. Coleman, C.J. *Managing Labor Relations in the Public Sector*; Jossey-Bass: San Francisco, 1990.
4. Spero, S.D.; Capazzola, J.M. *The Union Community and Its Unionized Bureaucracies*; Dunellen: New York, 1973.
5. ICMA. *Human Resource Management in Local Government: An Essential Guide*; International City/County Management Association: Washington, DC, 1999.
6. Berman, E.M.; Bowman, J.S.; West, J.P.; Van Wart, M. *Human Resource Management in Public Service*; Sage: Thousand Oaks, CA, 2001.
7. Balfour, A. Appropriate bargaining units for state employment. In *Public Personnel Management: Readings, Cases, and Contingency Plans*; Levine, M.J., Ed.; Brighton Publishing Co.: Salt Lake City, UT, 1980.
8. Nigro, L.G.; Nigro, F.A. *The New Public Personnel Administration*; F.E. Peacock: Itasca, IL, 2000.
9. Tomkins, J. *Human Resource Management in Government*; Harper Collins: New York, 1995.
10. Robinson, M. Middle managers in the bargaining unit. *Rev. Public Pers. Adm.* **1993**, *13* (3), 29–57.
11. Cayer, J. *Public Personnel Administration in the United States*; St. Martin's Press: New York, 1996.
12. Pynes, J.E.; Lafferty, J.M. *Local Government Labor Relations: A Guide for Public Administrators*; Quorum Books: Westport, CT, 1993.

United States Treasury Securities

Theo Edwin Maloy

*Department of History and Political Science, West Texas A&M University,
Canyon, Texas, U.S.A.*

INTRODUCTION

Among the many duties of the United States Department of Treasury is managing government accounts and the public debt.^[1] The department's Bureau of the Public Debt borrows the money needed to operate the federal government by issuing and servicing marketable, savings, and special securities.

The Bureau of the Public Debt evolved from the Register of the Treasury, becoming the Public Debt Service in 1919. It was designated as a bureau in 1940. With the passage of the Government Securities Act in 1986, the bureau assumed the rulemaking authority for the government securities market. The bureau's functions include the following:

- Borrowing the money necessary to operate the federal government, and accounting for the resulting public debt;
- Receiving, storing, issuing, and redeeming government securities;
- Servicing registered accounts and paying interest when due;
- Maintaining accounting and audit control over public debt transactions and publishing statements;
- Processing claims for securities that are lost, stolen, or destroyed;
- Promoting the sale and retention of United States Savings Bonds.^[2]

DEBT INSTRUMENTS

Treasury securities are debt instruments. They are issued to raise the money needed to operate the federal government and to pay off maturing debt obligations. Currently, individuals, corporations, state or local governments, foreign governments, and other entities outside of the United States Government hold just less than US\$4 trillion in United States Government debt. Government trust funds, revolving funds, and special funds hold just less than US\$3 trillion in United States Government debt, including Federal Financing Bank securities. Treasury securities are a safe and secure investment because the full faith and credit of the United States Government guarantee that interest and

principal payments will be paid on time. Interest on Treasury securities is exempt from state and local income taxes.^[3–5]

Most Treasury securities are “liquid,” which means that they can easily be sold for cash at the prevailing price in a “secondary market”—a public double-auction market where buyers continually offer securities for sale and sellers continually offer to buy those securities. As a group, Treasury securities are often called “marketable Treasury securities,” “Treasury securities,” or “Treasuries.” Individually, they are sometimes called “T-bills,” “T-notes,” and “T-bonds.”^[6]

OWNERSHIP OF DEBT INSTRUMENTS

The algebraic sum of all accumulated United States deficits and surpluses is popularly known as the “federal debt.” Beginning in 1918, Congress has set a limit on the total dollar amount of debt securities that the Treasury can have outstanding. The current “debt ceiling” is US\$7.38 trillion, raised in May 2003 from US\$6.4 trillion. Congress may have to increase the debt limit when the federal budget is in surplus on a unified budget accounting basis; the unified budget surplus or deficit only reflects the government's transactions with the public. The debt limit includes Treasury borrowing by government trust funds as well as borrowing by the public.^[7]

The current federal debt is approaching US\$7 trillion at a rate of increase of approximately US\$1.5 billion/day. Foreign and international individuals and institutions hold in excess of US\$1.5 trillion of the US\$4 trillion in government securities held by entities outside of the United States Government. As the supply of government securities increases, foreign and international holdings of government securities increase at an approximately 22% year⁻¹ rate.^[8] Asians hold 55% of foreign investments in Treasury securities, and the buying continues. Since the financial crisis of 1997 and 1998, Asian economies have rebounded, in part, by using cheap currencies to boost exports to the United States. In a search for stable investments, the region's central banks have invested trade surpluses in dollar-based securities, primarily United States bonds. In addition, the central banks have sold their currencies for dollars

to maintain the weak currency advantage. Many of the dollars, too, are invested in United States securities. Although the buying supports the United States securities markets, international selling—and rumors of selling—has a detrimental effect on American markets. However, the foreign central banks' large holdings of United States Treasury securities are difficult to sell. If large amounts of securities are sold, the value of the banks' remaining securities also declines.^[9]

TYPES OF SECURITIES AND THEIR OWNERSHIP

The Bureau of the Public Debt sells Treasury bills, notes, and savings bonds to individual investors. The buyer lends money to the United States Government, and receives the payment of interest in exchange for the loan. Treasury bills and notes are sold in increments of US\$1000, with a minimum purchase of US\$1000.

All Treasury bills and notes are issued and held by an entry in an electronic ledger, called the “book-entry form.” An investor can hold Treasury securities in one of two systems: Treasury Direct or the commercial book-entry system. The security's owner has a direct relationship with the Treasury in the Treasury Direct system. The commercial book-entry system is an indirect holding system where a financial institution, government securities broker, or dealer holds the record of the security for the owner. The commercial book-entry system is a multilevel arrangement that involves the Treasury, the Federal Reserve System (acting as the Treasury's agent), banks, brokers, dealers, and other financial institutions.

Treasury Direct makes principal and interest payments by direct deposit to the security holder's bank account and sends statements to the holder. The Treasury charges no fees for opening an account or buying securities, but it charges a maintenance fee for accounts holding securities with par values in excess of US\$100,000. Treasury Direct allows security holders to automatically reinvest the proceeds from most maturing securities.

In the commercial book-entry system, the security investor maintains a relationship with a financial institution, broker, or dealer, and potentially pays fees for its services. The institution will receive the security owner's principal and interest payments from the government. The commercial book-entry system allows an investor to easily buy and sell securities, as well as—unlike Treasury Direct—to use them for collateral. An investor can also hold Treasury securities in stripped form—known as STRIPS or zero-coupon treasuries—in the commercial book-entry system.

STRIPS are Treasury securities that do not make periodic interest payments. Market participants create

STRIPS by separating the interest and principal parts of a Treasury note or bond. For example, a 10-year Treasury note consists of 20 interest payments—one every 6 months for 10 years—and a principal payment payable at maturity. When the security is “stripped,” each of the 20 interest payments and the principal payment becomes separate securities and can be held and transferred separately. STRIPS can only be bought and sold through a financial institution, broker, or dealer, and held in the commercial book-entry system.

To sell a security held in the commercial book-entry system, the holder arranges a sale through a financial institution, government securities dealer, broker, or investment advisor. Normally, there is a fee for this service. If the investor holds a security in Treasury Direct, it can be transferred to an account in the commercial book-entry system, or it can be sold through the Treasury's Sell Direct program for a fee. Securities can be transferred between Treasury Direct and the commercial book-entry system.^[5]

The Treasury Department has not offered a Treasury bond because of the decision in October 2001 to suspend the issuance of 30-year bonds. Bonds issued with coupon rates above current market interest rates are periodically called for early redemption.^[10] The Treasury sold non-callable 30-year bonds on February 15 and August 15, and may have subsequently issued an additional amount of a bond at a “reopening.” The 30-year bond was the “benchmark” for all debt instruments. Until February 1985, the Treasury issued fixed-principal bonds that are callable 5 years prior to maturity. After providing 4 months' notice, the Treasury can call any callable bond on its first call date, or on any semiannual interest payment date thereafter. The callable bond with the longest remaining period to maturity is the 11.75% bond of November 15, 2009–2014, which was issued in November 1984.^[11]

TREASURY BILLS

Treasury bills, or T-bills, are securities that mature in 1 year or less from their issue date. T-bills are bought for a price less than their face value, called the “par value.” T-bills can be issued for any period of time less than 1 year, but 28-day, 91-day, and 182-day T-bills are most prevalent. When the T-bill matures, the Treasury pays the purchaser the par value. The interest is the difference between the security's purchase price and what the Treasury pays the owner at maturity. If a T-bill were sold before it matures, a buyer in the open market would pay the seller an amount including the accrued interest. There is an active secondary market for T-bills.

The Treasury usually auctions 13-week and 26-week bills every Monday and a 4-week bill on Tuesdays, for

issue on Thursday of that week. It also issues cash management bills from time to time for periods as short as 1 day to maintain the Treasury cash balance, especially immediately before regular tax payment dates.^[12]

Although T-bills pay interest, they are sold at a “discount rate.” If a 91-day US\$1000 Treasury bill were bought at a 0.840% discount, it would be purchased for US\$99.788 per US\$100 of face value. The discount for a 91-day US\$1000 T-bill with a 0.840% discount can be approximated at:

$$2.123 = \text{US } 1000(\text{discount rate}) \\ (\text{days to maturity}/360)$$

The discount rate is not the investment rate, sometimes called the “bond equivalent yield.” For a 91-day US\$1000 Treasury bill bought at a 0.840% discount, bond equivalent yield can be approximated at:

$$0.008535 = (365 \times \text{discount rate})/[360 \\ - (\text{discount rate} \times \text{days to maturity})]$$

The Treasury’s Recent Treasury Bill Auction Results lists term, issue date, maturity date, discount rate, investment rate, price per hundred, and the T-bill’s Committee on Uniform Securities Identification Procedures (CUSIP) number. A 91-day T-bill with a 0.840 discount would be listed with a 0.854 investment rate and a US\$99.788 price per US\$100 of face value.^[13]

TREASURY NOTES AND BONDS

Treasury notes and bonds are securities that pay a fixed rate of interest every 6 months until the security matures, when the holder receives the par (or face) value. The only difference between a Treasury note and a Treasury bond is their time until maturity. Treasury notes mature in more than a year, but not more than 10 years from their issue date. Bonds, when they are issued, mature in more than 10 years from their issue date. Notes and bonds are usually originally offered at a price close to their par value.

Currently, the Treasury sells fixed-principal and inflation-indexed notes. Both pay interests twice a year, but the principal value of inflation-indexed securities is adjusted to reflect inflation as measured by the Consumer Price Index (the Bureau of Labor Statistics Consumer Price Index for All Urban Consumers). With inflation-indexed notes and bonds, the semiannual interest payments and maturity payment are based on the inflation-adjusted principal value of the security.

The Treasury issues 2-year notes at the end of each month, and 3-year, 5-year, and 10-year notes are issued on February 15, May 15, August 15, and November 15

of each year. The Treasury reopens 5-year notes on the 15th day of the month after each original issue. An offering is usually announced 1 week before the auction.^[14]

BUYING MARKETABLE TREASURY SECURITIES

Treasury bills, notes, and bonds can be purchased in a Treasury auction, or in the securities market. If an investor wants to buy a Treasury security at an auction, the buyer contacts the Treasury, a Federal Reserve Bank, a financial institution, or a government securities broker or dealer. If the buyer wants to buy a Treasury security in the securities market, the investor contacts a financial institution, broker, or dealer.

Each Treasury bill, note, or bond is sold at a public auction. In the Treasury’s auctions, all successful bidders are awarded securities at the same price, which is the price equal to the highest rate or yield of the competitive bids that are accepted. A complete explanation of the auction process is published in the Uniform Offering Circular, which is in the Code of Federal Regulations at 31 CFR Part 356. About 1 week before each auction, the Treasury issues a press release announcing the security being sold, the amount being sold, the auction date, and other pertinent information.

For each security that the investor wants to buy, a bid is submitted. The investor can either bid non-competitively or competitively, but not both in the same auction. A non-competitive bid receives the full amount of the security that the buyer wants at the return determined at that auction. An individual cannot bid for more than US\$1 million in a bill auction, or US\$5 million in a note auction. Most individual investors bid non-competitively. A competitive bid specifies a “rate” for bills, or the “yield” for notes. If the return that the bidder specifies is too high, the investor might not receive any securities, or just a portion of what is sought. However, competitive bids can be much larger than non-competitive bids. Bids can be submitted directly to the Treasury, to a Federal Reserve Bank, or through a financial institution, broker, or dealer. Bids are accepted by mail. Current customers can submit bids via the Internet, or by touch-tone telephone. A financial institution, government securities broker, or dealer can also submit bids on an individual’s behalf, probably for a fee.

UNITED STATES SAVINGS BONDS

Savings bonds are Treasury securities that are payable only to the person to whom they are registered. They are not transferable in a secondary market. Savings bonds can earn interest for up to 30 years, but they

can be cashed after 6 months if purchased before February 1, 2003; or after 12 months if purchased on or after February 1, 2003.

Series EE and Series E bonds and savings notes are “accrual securities” where the interest is periodically added to the amount the buyer originally paid, to establish their current redemption value. When a buyer cashes a Series EE bond, a Series E bond, or a savings note, the owner receives the redemption value, which is the return of the original investment plus the interest that was earned as the bond was held. The Series EE savings bond increases in value until it is cashed or reaches final maturity in 30 years. The “double E” is the successor to the Series E bond that was issued from May 1941 to June 1980. Investors can purchase up to US\$30,000 in Series EE bonds each calendar year.

The amount printed on a Series EE is its denomination, also known as its face amount. Eight denominations are available: US\$50, 75, 100, 200, 500, 1000, 5000, and 10,000. The US\$50 and 75 denominations are not available through Payroll Savings Plans and other employer thrift plans. Because bonds accrue interest for up to 30 years, the bond’s redemption value can exceed its face value. Since December 2001, Series EE bonds have been inscribed as “Patriot Bonds” as part of the country’s antiterrorism effort.^[15–17]

Series HH and Series H bonds are current-income securities issued with face values of US\$500, 1000, 5000, or 10,000. The redemption value of the bonds remains constant at the amount the buyer invested, and the interest is paid to the holder every 6 months. When the owner cashes a Series HH or Series H bond, the original investment is returned. There is no limit on how much a person can invest in Series HH bonds.^[17,18]

Series I bonds are a new type of accrual security bond designed for investors seeking to protect the purchasing power of their investment and earn a guaranteed real rate of return. Series I bonds are sold at face values of US\$50, 70, 100, 200, 500, 1000, 5000, and 10,000, and they grow in value with inflation-indexed earnings for up to 30 years. An investor can purchase up to US\$30,000 in Series I bonds during a calendar year. The bonds offer a fixed rate combined with semiannual inflation adjustments to help protect purchasing power. The fixed rate of return is announced by the Treasury Department each May and November. The fixed rate of return announced in May is the same over the entire life of the Series I bonds purchased between May 1 and October 31 of that year. The fixed rate of return announced in November applies to the entire life of Series I bonds purchased between November 1 and April 30 of the following year. The semiannual inflation rate is also announced each May and November by the Treasury Department. The semiannual inflation rate is based on changes in the Consumer Price Index. The semiannual

inflation rate announced in May is a measure of inflation over the preceding October through March. The inflation rate announced in November is a measure of inflation over the preceding April through September. The semiannual inflation rate is combined with the fixed rate of a Series I bond to determine the Series I bond’s earnings rate for the next 6 months. Series I bonds increase in value each month, and interest is compounded semiannually. They earn interest for up to 30 years. In the unlikely event that the Consumer Price Index is negative, the value of Series I bonds would remain the same until the inflation rate becomes greater than zero. In addition to earnings being exempt from state and local income taxes, Series I bonds’ federal income taxes can be deferred for up to 30 years, until redemption or other taxable disposition. Taxpayers who qualify can exclude all or part of Series I bonds’ interest from income as long as the proceeds are used to pay for tuition and fees at eligible post-secondary educational institutions.

Two types of savings bonds can be bought for cash—the Series EE bonds and the Series I bond. Series HH bonds can only be bought in exchange for Series EE, Series E, and savings notes, or when the owner reinvests the proceeds of matured Series H bonds.^[19]

OTHER SECURITIES

Government Account Series securities are issued because various government trust fund statutes provide for interest-earning investments of money that is not needed immediately for the purposes of the fund. Growth in the trust funds is attributable to returns on these investments and current contributions to the funds. Interest rates on par-value Government Account Series securities are prescribed in statutes that provide the investment authority for government trust funds, which include the Social Security trust funds, the Civil Service Retirement and Disability Fund, the Federal Hospital Insurance Trust Fund, and the Federal Retirement Thrift Savings Fund (G Fund). Generally, the Treasury pays interest semiannually on par-value Government Account Series securities. The G Fund invests in overnight par-value securities, and interest is compounded each day. Similar, the Treasury started to issue market-based securities in 1974 to provide an investment mechanism within the government for federal accounts that are authorized to invest in securities that are issued or fully guaranteed as to principal and interest by the United States. These accounts, which number over 200, include retirement fund assets managed by the Pension Benefit Guarantee Corporation, Federal Housing Administration mortgage insurance funds, and the Bank Insurance Fund. The statutes of federal agencies and funds that invest in market-based

securities authorize investments but do not authorize investing in par-value Government Account Series securities.^[20]

The Treasury sells State and Local Government Series securities to issuers of federally tax-exempt securities as an investment for the gross proceeds of a tax-exempt issue, or any other amount that assists the state or local government to comply with provisions of the Internal Revenue Code related to the tax exemption. The book-entry securities are not transferable. The purchaser can designate the interest rate on a time deposit security, but the rate cannot exceed the maximum interest rate. The maximum rates are “five basis points” (five one-hundredths of 1%, or 0.0005) below the current Treasury borrowing rate for a Treasury security of comparable maturity. The interest rate for a time deposit security that matures in more than 30 years, up to and including 40 years, is the maximum rate on a 30-year State and Local Government Series security. The interest rate for demand deposit securities is a variable rate that the Treasury calculates by using the investment yield on 13-week Treasury bills in the most recent weekly auction, adjusted by the estimated marginal tax rate of investors in tax-exempt securities and adjusted for Treasury administrative costs. Interest is added to the principal and is reinvested daily until redemption.^[21]

The Treasury has issued dollar-denominated Foreign Government Series securities directly to foreign governments since 1962. It has also issued foreign currency-denominated foreign series securities on several occasions between 1961 and 1980, but none of them remains outstanding. Special operations, mostly involving United States Government loans to foreign countries for military purchases in the United States, necessitate dollar-denominated securities as temporary investments for the borrowed funds. The Treasury also issued long-term zero-coupon securities to foreign governments on several occasions between 1988 and 1994 to assist them in restructuring their obligations under the Brady Plans. The proceeds of the zero-coupon securities were pledged to pay the principal amount of bonds issued by the foreign governments.^[22]

Domestic Series securities were issued in 1989 as an investment for Refcorp under the Federal Home Loan Bank Act. Proceeds of the 30-year and 40-year zero-coupon bonds were pledged to pay the principal amount of marketable Refcorp bonds. The Treasury has not issued Domestic Series securities to another entity.^[23]

The Federal Financing Bank is a government corporation, created by Congress in 1973, under the general supervision of the Secretary of the United States Treasury, established to centralize and reduce the cost of federal and federally assisted borrowings from the public. The bank aids federal agencies and other borrowers whose debt is guaranteed by the federal

government. The Federal Financing Bank was also established to deal with federal budget management problems that occurred when offbudget financing flooded the government securities market with offers of a variety of government-backed securities that competed with Treasury securities. The bank has the authority to purchase any obligation issued, sold, or guaranteed by a federal agency to ensure that fully guaranteed obligations are financed most efficiently.^[24]

Other miscellaneous federal debts exist, including Adjusted Service Bonds from World War I and Armed Forces Leave Bonds from World War II.^[25]

CONCLUSION

Federal government yearly deficits result from expenditures exceeding revenues. The Treasury borrows money to finance the accumulated deficit. In an effort to increase demand for government debt instrument, the Treasury offers a variety of debt types and maturities. As financial markets develop and mature, the Treasury will create new debt instruments to make U.S. government debt more attractive to domestic and foreign investors and governments.

REFERENCES

1. <http://www.treas.gov/education/duties/index.html>.
2. <http://www.treas.gov/education/duties/bureaus/public-debt.html>.
3. <http://www.publicdebt.treas.gov/opd/opdpdodt.htm>.
4. <http://www.publicdebt.treas.gov/opd/opdfaq.htm#opdfaq32>.
5. <http://www.publicdebt.treas.gov/of/ofbasics.htm>.
6. <http://www.publicdebt.treas.gov/sec/secfaq.htm#secfaq1>.
7. <http://www.publicdebt.treas.gov/com/comlimit.htm>.
8. <http://www.fms.treas.gov/bulletin/>.
9. <http://online.wsj.com/article/0,,SB106000774754160200-email,00.html>.
10. <http://www.publicdebt.treas.gov/com/combills.htm>.
11. <http://www.publicdebt.treas.gov/com/combonds.htm>.
12. <http://www.publicdebt.treas.gov/sec/seccall3.htm>.
13. <http://www.publicdebt.treas.gov/AI/OFBills>.
14. <http://www.publicdebt.treas.gov/com/comnotes.htm>.
15. <http://www.savingsbonds.gov/sav/savsere.htm>.
16. <http://www.savingsbonds.gov/sav/savpayme.htm>.
17. <http://www.publicdebt.treas.gov/sav/savpatriotbond.htm>.
18. <http://www.savingsbonds.gov/sav/sbhglanc.htm>.
19. <http://www.savingsbonds.gov/sav/sbifaq.htm>.
20. <http://www.publicdebt.treas.gov/com/comgas.htm>.
21. <http://www.publicdebt.treas.gov/com/comslg.htm>.
22. <http://www.publicdebt.treas.gov/com/comfor.htm>.
23. <http://www.publicdebt.treas.gov/com/comdom.htm>.
24. <http://www.ustreas.gov/ffb/>.
25. <http://www.publicdebt.treas.gov/>.

Urban Planning and Ethics

Carlos Nunes Silva

Department of Geography, Faculdade de Letras, University of Lisbon, Lisbon, Portugal

INTRODUCTION

Urban and regional planning faces complex ethical dilemmas related to public interest; professional integrity; obligations toward the community, clients, employers, and colleagues; intellectual property; as well as the environment, other forms of life, and future generations, both in strategic and daily decisions. Therefore, besides the fact that they have to obey and implement the law, the mission of planning professionals includes complex judgements over conflicting land uses, which makes this one of the areas in public administration and public policy where decisions are embedded in highly complex, multilayer, and multi-actor decision-making processes. A planner often has to weigh personal values against those established in one's professional organization, or in the society. In spite of this, the discussion of ethical principles in urban and regional planning is a relatively recent fact, even in countries where the planning profession has been organized for a long time.

THE RISE OF ETHICS IN THE PLANNING PROFESSION

The pioneering studies of Howe and Kaufman^[1,2] on planners' values and ethics showed that planners acted differently according to their values and ethical principles, but, contrary to what could be expected from these results, the issue was not on top of the agenda in planning theory discussions at that time. Notwithstanding the fact that planners commonly did not acknowledge explicitly that they were considering ethical principles (e.g., when planning new land uses), they were in fact reasoning usually as consequentialists (or utilitarians), weighing the "good" and the "bad" of the outcomes of their decisions through planning techniques such as cost-benefit analysis. That planning has been guided somehow by values seems to be a fact throughout the history of urban planning, at least in its utopian proposals.

But only in the last two decades has a clear interest in the ethical dimension of planning been developed as a result of different factors, as numerous authors such as Kaufman,^[3-5] Khakee and Dahlgren,^[6] and Hendler^[7] show. First, it is a consequence of the conviction that

the allocation of different land uses is an ethical problem, in the sense that decisions about land use have economic, social, cultural, and environmental consequences. Second, the emergence of new visions of planning, in reaction to the then-dominant "rational planning model," led to the recognition that decisions about land use are not value-neutral; it is through the political process that some groups or individuals gain and others lose, according to their power resources. This interest in the political dimension of planning was further reinforced by the theory of justice by Rawls^[8] and the justification for the minimal state by Nozick,^[9] which together helped the development of a strong interest in the ethical dimension of planning in the late 1970s onward. In fact, unlike many other professions, planning involves choices on key social facets of collective life. For example, planners have to decide what are good or bad public spaces, or on housing estate characteristics, or on a transport solution. Thus they need to know how to decide among different claims, taking into consideration multiple and contradictory dimensions. Third, the growing number of professionals in this field increased the social relevance of this activity, further augmented by a more demanding citizenship. Fourth, the social impact of incorrect professional behaviors forced professional organizations to make explicit ethical guidelines in the form of codes of conduct. Finally, the increasing relation between the public sector and the private sector, as a consequence of the liberal economic policies after the 1970s, challenged traditional public administration values and legal control procedures. This and the conviction that different people in the same service can behave differently encouraged governments and international organizations dealing with public policy to adopt ethical standards.^a

The adoption of codes of ethics in planning, if centered on human rights and on the environment, is

^aFor example, for a brief review of measures proposed in Organization for Economic Cooperation and Development (OECD) countries, refer to Refs.^[10-12]. For the kind of measures proposed by the United Nations for the public sector, in general, refer to Refs.^[13,14]. For a similar review in the case of local government, refer to Ref.^[15]. See also the Organization of American States for its anti-corruption measures and the World Bank's "Code of Professional Ethics."

commonly seen as having practical relevance. It not only facilitates the resolution of moral problems which professional planners face, but is seen also as a self-regulation reference for professional conduct; it helps to find an ideal for the profession, states the profession core values, and reinforces the profession's status in society; it informs employers, clients, and the population, in general, about the kind of responsibilities and obligations a planner has toward them; it helps to create an environment of fairness in which ethical conduct will be the norm, as the works of Kaufman,^[3-5] Hendler,^[16,17] and Lawton^[18] suggest. Additionally, it is widely believed that professional behavior is conditioned by the organization's ethical environment and the ethical culture of that specific group. Therefore, to have an ethical code of professional conduct, explicit moral norms is seen as a positive factor for the definition and prevention of unacceptable behavior. Indeed, if a certain professional conduct is seen as unacceptable, then it will be less probable that it will be practiced by many members.

Codes of professional ethics in planning are relatively recent, even in countries where the discussion of ethical principles inside professional organizations has a long-standing tradition, as Kaufman,^[3-5] Mahoney,^[19] and Fisher^[20] point out. Notwithstanding, there are several examples of professional organizations related to planning that adopted codes of conduct long ago, although usually with only very general guidelines. For example, in 1914, the American Society of Civil Engineers (ASCE) adopted a code of ethics (www.asce.org), amended several times afterward. In 1924, the International City-County Management Association (ICMA) adopted a Code of Ethics that was subsequently revised on several occasions. The Town Planning Institute in the United Kingdom (www.rtpi.org.uk) adopted in the 1930s a code of professional conduct. The predecessor of the American Institute of Certified Planners (AICP) (www.planning.org/aicp), the American Institute of Planners (AIP), adopted a code of ethics and professional conduct in 1971, which was revised in 1978 by the new AICP and subsequently revised again in 1981. The American Planning Association (APA) (www.planning.org) adopted a "Statement of Ethical Principles" in 1987 as a guide for all its members, replacing the 1962 code of the preceding organization, the American Society of Planning Officials (ASPO). In 1992, the APA adopted a "Statement of Ethical Principles in Planning" to serve as a guide of ethical conduct for all those involved in the planning process.^[3,21] In Canada, in 1994, the Canadian Institute of Planners adopted also a code of conduct for its members (www.cip-icu.ca).

In Europe, the European Council of Town Planners (ECTP) in 1985 adopted guidelines for professional conduct to be developed by each national association

of planners in each member country.^[22] In the European Union (EU), there were additional reasons for the harmonization of professional codes related to the increased labor mobility in the EU market—a situation that affected several professional groups related to urban and regional planning; this led the ECTP to recognize the need for some sort of deontological harmonization within Europe (www.ceu-ectp.org).

A similar process toward the adoption of professional codes of conduct by planners can be found in many other countries around the globe such as, for example, Australia (www.planning.org.au), New Zealand (www.nzplanning.co.nz), India (www.itpindia.org), Brazil (www.arquitetofna.org.br), and South Africa (www.saplanners.org.za).

Other professional groups involved in urban and regional planning also adopted codes of professional conduct in the same period, such as Landscape Architecture (www.asla.org), Architects (www.aia.org/), Civil Engineer (www.asce.org), Geographers (www.aag.org), Sociologists (www.asanet.org), among others. Most of these codes deal with issues of professional conduct but say little about the major aims of planning. A search of North American professional organizations in other fields of activity also shows that the 1990s was a period during which most of them revised their previous code of ethics and professional conduct,^b which were initially adopted, in most cases, in the 1970s or 1980s. This means that we are talking of a practice that—in most professions related to urban and regional planning—has been in effect for only around three decades.

But probably the situation in all these professional organizations in the early years of professional ethics codification was not much different from the one described by Kaufman^[3] and related to the AICP case: "While some American planners were probably aware that their respective planning organizations adopted ethics codes in the 1980s, few probably knew what was in these codes. This is not meant as a criticism of planners because codes tend to be exceedingly dry and uninspiring documents to read." But since then, the situation seems to be changing. Indeed, since the 1980s, the subject of planning ethics has been introduced in several graduate planning programs in the United States, Europe, and other parts of the world, including a discussion on how best to teach planning ethics. At the same time, numerous empirical studies

^bFor example, that was the case of the American Historical Association, in 1990; American Evaluation Association, in 1994; Archaeological Institute of America, in 1994; American Mathematical Society, in 1995; American Psychological Association, in 1997; American Society of Landscape Architecture, in 1998; and American Statistical Association, in 1999.

of the ethics of planners were performed and the results were published, as a consequence of the growing concern with the political and ethical dimensions of planning as Kaufman,^[4,5] Klosterman,^[23] and Hendler^[7] point out. It also became common to find seminars dedicated to these topics in programs of professional planning conferences and congresses, as well as special training sessions organized by planners' professional associations.

Despite planning professionals' general acceptance of ethic codes, they often also are seen as an unsatisfactory set of recommendations—vague and paternalistic, and based on a naive view of planning theory and on a simplistic conception of power relations, as Lucy^[24] and Fisher^[20] argue. As planning is a multidisciplinary field and a multiactor activity, having a code of conduct specific for planners risks ignoring other key stakeholders' concerns and impacts. This view argues for a more comprehensive approach to professional ethics codification. On the other side, the code's effectiveness depends on several conditions not always practiced by all professional organizations. Examples include the implementation of ethics education programs, hotlines for reporting wrongdoings, adoption of a legal framework to support ethical standards, and concomitant appropriate sanctions to deal with misconduct.

ETHICAL PRINCIPLES APPLIED IN PLANNING

There are two main groups of ethical perspectives to be considered in urban planning. The first group includes the consequentialist and deontological perspectives, and the second includes the anthropocentric and non-anthropocentric perspectives.^{[25,26]c} If both dimensions are combined simultaneously, then four major categories of perspectives can be considered.

For the consequentialist perspective, it is the outcome of an action or behavior that determines if it is correct or not, or if it is "good" or "bad." Utilitarianism is an example of the consequentialist perspective. On the contrary, deontological theories see any action or behavior as having a value in itself, independently of its consequences. An ethical urban and regional planning has to consider and contrast ends-oriented ethics (consequentialist ethics) with means-oriented ethical principles (deontological ethics), as both sets

of principles are morally relevant in addressing dilemmas faced by urban and regional planners.

A second dimension is related to the importance given to *Homo sapiens* in the definition of the moral community. For anthropocentric perspectives, only *H. sapiens* matter for that purpose; whereas for non-anthropocentric perspectives, other forms of life and the environment are also valued independently of the importance they have for human beings.

Therefore the professional practices of planners can be rather different from each other depending on the ethical perspective adopted, as research works by Howe and Kaufman,^[1,2] Khakee and Dahlgren,^[6] and others showed.

From a deontological and non-anthropocentric perspective, for example, ethical urban and regional planning seeks to maximize the public interest, without compromising other moral duties in relation to other forms of life, the environment, and future generations. Therefore it should seek to create a more just society and to improve redistributive justice among individuals, groups, and territories, following the theory of justice and communitarian perspectives by Rawls, but at the same time care about the impact on the environment.^[31] This can be performed through the choices made for the provision of housing, education, culture, employment, and health facilities, for example. Thus planning should protect the interests of the disadvantaged (disabled, aged, children, women, and minorities). Planning proposals should respect basic human rights, including the right to a minimum of social benefits available to all, irrespective of economic capacity, and should also prevent harm to people or communities and the environment. Deep Ecology, first proposed by Aarne Naess, is one example of a deontological and non-anthropocentric perspective relevant to the discussion of urban planning ethics.^[32]

Whatever the ethical perspective adopted, planners face three main sets of ethical dilemmas. One is related to individual rights, with the tensions between freedom and autonomy on one side, and governmental regulations on the other. The second is related to the overall society, with the tensions, for example, between economic efficiency and social justice. The third relates to the environment and the conflicts between, for instance, environmental protection and development.

The first one (planning and individual rights) has to do, for example, with the preservation of the characteristics of the locality and the choice of lifestyles, with paternalism in planning decisions and the need to prevent or minimize damages, or with limits imposed on private propriety.^[26,33] In fact, in urban planning decision-making processes, there are frequently controversies regarding the location of certain activities, or about certain local characteristics (landscape beauty, architectural styles, etc.), or about the limits

^cRefer to Ref.^[27] for a review of classical texts (Jeremy Bentham, John Stuart Mill, etc.), as well as contemporary expressions of consequentialism (Richard Brandt and Robert Adams, etc.); for a discussion of utilitarianism, see, among others, Refs.^[28,29]. Refer to Ref.^[30] for a review of classical texts on deontology (Immanuel Kant, W. D. Ross, etc.) and also its contemporary expressions (Robert Nozick, Thomas Nagel, Stephen Darwall, etc.).

imposed on owners of private properties. Urban growth controls, such as building restrictions that stimulate higher prices, can be seen in some cases as an instrument to protect the character of an area by reducing the ability of low-income families to move in. This and community opposition to the location of poor families, homeless, mentally disabled, or ethnic minorities in the neighborhood are two examples of the social forces that a planner has to face and balance, and which can directly affect individual rights. The balance between national security and individual liberties is also becoming an important ethical issue for planners, in part related to the increasing use of electronic surveillance device in public spaces.

Some planners consider it acceptable and desirable to preserve certain characteristics of the community and to restrict land use, new buildings, or reconstruction of old ones; this requires sensitivity to place identity. Other planners maintain that, if there are no strong arguments such as public health or security reasons, people should be left free to carry out their own choices.^d However, the argument for the protection of local characteristics should not imply the creation of a monolithic community; therefore urban planning should facilitate the multiple lifestyles that may exist. Moreover, a fair and equitable public participation procedure should always be put in practice to ensure equal opportunities for all interests to be considered. Such procedures may include the use of a local referendum, but this is an uncommon practice.

The possibility of damages on third parties caused by individual decisions, such as those resulting from conflicting land uses, is another type of situation that justifies public intervention and concomitant restrictions on property rights.^e If some of the negative consequences created by private actions cannot be avoided, it is important, according to this perspective, to seek compensation for such damages from those that caused them.

Restrictions and benefits imposed by a plan zoning on private property need to be balanced fairly. In this case, it is commonly considered correct to apply a system of transfer of development rights from new development areas to compensate owners not being allowed to build, or owners allowed to build less than average, and also to compensate owners of rural land as a form of compensation for the preservation of agricultural and forest lands.

The second group of ethical dilemmas has to do with the social responsibility of planners. In fact, it can be argued from a deontological perspective that planners also have redistributive obligations toward different social groups if we take, for instance, Rawls' contractarian theory of justice. Urban planning should seek to reduce economic and social inequalities through maximization of benefits for the most disadvantaged social groups. From a deontological perspective, ethical urban and regional planning achieves the maximum well being for the majority of citizens, which means that the utilitarian dimension of a plan has to be conditioned by other moral obligations (such as those related with certain human rights) as part of the social contract. Therefore there is a moral obligation for the content of an urban plan to maximize the net benefits for the most discriminated or oppressed in society.^[26,35,36] A planner who values children, the aged, the disabled, or women, for instance, will certainly design urban environments that are different from those who do not, or will not endorse or approve developments that do not meet these criteria.

From this point of view, planners also have ethical obligations in relation to future generations,^f especially due to the fact that few professions have a greater influence on the human habitat and on the quality of people's life. Therefore they should consider the cumulative impact of their decisions and prevent the destruction of built or natural patrimony, and avoid, when possible, all decisions with irreversible consequences. From this perspective, planners have a moral obligation to keep children and young people at the center of planning options, as they will be the generation that will receive the consequences of today's decisions. Nevertheless, there are arguments against the existence of obligations to future generations on the part of planners.

Finally, a third category of ethical dilemmas faced by planners is related to the environment and urban sustainability. In an irreversibly and increasingly urban world, professional planners have a moral obligation to plan for a sustainable city, which cannot be achieved without social equity and social inclusion, and is an additional reason for urban planning to be redistributive.^[37] For this, it is necessary to incorporate new values and a different relationship with nature. Such an aim requires a shift in the dominant urban sprawl model of development toward a more compact city—a condition to achieve sustainability.^g Because part of

^dA classical reference for this point of view is Ref.^[34].

^eThat is the case of the so-called NIMBY (Not In My Back Yard) or LULU (Local Unwanted Land Uses) type of land use conflicts. A socially responsible planner has a moral obligation to overcome these practices and to promote inclusive communities, reversing the NIMBY to a YIMBY attitude (Yes In My Back Yard).

^fSee Rawl's just saving principle; refer to Ref.^[8].

^gWe consider here the following widely used definition of sustainable development: "Development that meets the needs of the present without compromising the ability of future generations to meet their own needs."

biodiversity loss is due to pollution, destruction of natural habitats, and other man-created conditions, it is necessary to integrate development with conservation. But defining sustainable development in that way seems not enough. Indeed, as Cafaro^[38] argues, “massive biodiversity loss is sustainable, provided future generations can meet their self-defined needs.” Therefore, from a non-anthropocentric and deontological perspective, planners have a moral obligation to protect and preserve the environment for *H. sapiens* and also for other forms of life in this generation and in future generations.^[26,39] These contradictory aims that planners have to address between environmental protection and development are but one more example of the ethical dilemmas that these professionals have to deal with in their daily decisions.

From a deontological perspective, it can be argued that communities have a right to live in a clean environment, which means that there is an ethical duty for planners to minimize the ecological footprint of cities and to promote a model of urban development contrary to urban sprawl. In this model, urban development is more compact, with priority for the redevelopment of brownfield zones instead of occupation of new natural areas, which should only be urbanized when there is no alternative. Therefore it is necessary to develop new ethics for the built environment; reduce urban sprawl associated with the use of individual cars; promote more compact urban patterns, through increase of urban densities and mixing of activities; and create a city that is both safer and easily accessible on foot, as the “New Urbanism” movement claims in its statement of principles.^[40] If initially and for most people these ethical obligations for the promotion of an eco-friendly urban and regional development had essentially an anthropocentric justification, they now need to be seen also in relation to other forms of life, thereby enlarging the moral community. Partially related to this is also an ethical duty to coordinate with nearby local or regional authorities and to estimate the impact of local decisions on neighboring administrative areas—a situation that calls for and justifies planning at metropolitan or regional scales.

In planning procedures from a deontological perspective, every citizen has the right to be treated fairly and with equity in the planning process, which means equal treatment by the plan for those in the same condition and the inclusion of those for whom a plan is being prepared. Nonetheless, this sometimes can be an unfair process because of the unequal power relations of those involved in the participation process. In practice, an ethical process of public participation in planning requires that those in a more unfavorable social condition be able to make themselves heard in the decision-making process, as the concept and practice of “advocacy planning,” proposed by Paul

Davidoff in the 1960s, tried to introduce into mainstream planning. Planners, like other professionals, not only should take into consideration the interests of all stakeholders, but also have an obligation of professional integrity; therefore they should not take personal gains from the decision-making process.

CONCLUSION

Whether dealing with inner-city processes or with natural areas, urban and regional planning inevitably faces conflicting interests and values. The idea of a value-neutral and rationalist planner is no longer a consensual description of mainstream professionals in this field. New ethics concerned with human rights and social justice, and with a deeper green consciousness emerged gradually in the planning profession worldwide in the last two decades. This new ethics replaced the traditional utilitarian and anthropocentric type of criteria in land-use decisions. Like other domains of public policy, urban and regional planning faces a wide range of values, often in conflict with each other. As a result, the task of deciding in practice what is wrong and right is a rather difficult enterprise. In recognition of this complexity and diversity of issues, planners’ professional organizations and schools of planning all over the world have been promoting the integration of ethics into their activities and teaching programs since the 1980s. If a driving idea emerges from discussions on these issues, it is certainly that ethics should guide urban and regional planning aimed at restoring the equilibrium of the natural, social, and built environments by reducing the ecological footprints of cities and making them more livable and inclusive places for the present and future generations, with due consideration for human and non-human forms of life and ecosystems.

REFERENCES

1. Howe, E.; Kaufman, J. The ethics of contemporary American planners. *J. Am. Plan. Assoc.* **1979**, *45* (3), 243–255.
2. Howe, E.; Kaufman, J. The values of contemporary American planners. *J. Am. Plan. Assoc.* **1981**, *47* (3), 266–278.
3. Kaufman, J. American codes of planning ethics. Content, development and after-effects. *Plan Can.* **1990**, *30* (5), 29–34.
4. Kaufman, J. Reflections on teaching three versions of a planning ethics course. In *Planning Ethics*; Hendler, S., Ed.; Center for Urban Policy Research: New Brunswick, 1999.
5. Kaufman, J. Is the Increased Output of Planning Ethics Knowledge in North America Affecting Those Who

- Practice Planning? An Exploratory Investigation, Paper Presented at World Planning Schools Congress, Xangai, China, 2001.
6. Khakee, A.; Dahlgren, L. Ethics and values of Swedish Planners: a replication and comparison with an American study. *SHPR* **1990**, 7, 65–81.
 7. Hendler, S., Ed.; *Planning Ethics*; Center for Urban Policy Research: New Brunswick, 1999.
 8. Rawls, J. *Uma Teoria da Justiça*; Editorial Presença: Lisbon, 2001. (“A Theory of Justice,” Harvard University Press, Cambridge, MA, 1971; first published in 1971, based on the author’s papers published in the previous 12 years).
 9. Nozick, R. *Anarquia, Estado e Utopia*; Zahar Editores: Rio de Janeiro, 1991. (“Anarchy, State, and Utopia,” Basic Books, New York, 1974).
 10. OECD. *Principles for Managing Ethics in the Public Service*; OECD: Paris, 1998.
 11. OECD. *Building Public Trust: Ethics Measures in OECD Countries*; OECD: Paris, 2000.
 12. Washington, S. L’étique et le service public. *Obs.-OCDE* **1997**, 204, 15–17.
 13. UN. *Professionalism and Ethics in the Public Service: Issues and Practices in Selected Regions*; United Nations: New York, 2000.
 14. UN. *Promoting Ethics in the Public Service*; United Nations: New York, 2000.
 15. IULA. *Local Government Associations: Promoting Ethics in Local Government*; International Union of Local Authorities: The Hague, 1999.
 16. Hendler, S. Moral theories in professional practice: do they make a difference? *Environments* **1990**, 20 (3), 20–30.
 17. Hendler, S. Do professional codes legitimate planners’ values? In *Dilemmas of Planning Practice. Ethics, Legitimacy and the Validation of Knowledge*; Thomas, H., Healey, P., Eds.; Avebury: Aldershot, 1991.
 18. Lawton, A. Developing and Implementing Codes of Ethics, Paper Presented at the European Group of Public Administration’s Conference, Oeiras, 2003.
 19. Mahoney, J. *Teaching Business Ethics in the UK, Europe and the USA. A Comparative Study*; The Athlone Press: London, 1992.
 20. Fisher, S. How to think about the ethics of architecture. In *Fox, Warwick, Ethics and the Built Environment*; Routledge: London, 2000; 170–182.
 21. Barrett, C. *Everyday Ethics for Practicing Planners*; American Institute of Certified Planners: Washington, 2001.
 22. ECTP. *International Agreement and Declaration by the National Institutes and Associations of Professional Town Planners Within the European Economic Community. Appendix C—Professional Conduct Requirements*; European Council of Town Planners, 1985.
 23. Klosterman, R. Ethical theory and planning education. introduction. In *Planning Ethics*; Hendler, S., Ed.; Center for Urban Policy Research: New Brunswick, 1999.
 24. Lucy, W. APA’s ethical principles include simplistic planning theories. In *Readings in Planning Theory*, 1st Ed.; Campbell, S., Fainstein, S., Eds.; Blackwell: Oxford, 1997; Vol. 54 (2). (JAPA, 1988).
 25. Howe, E. Normative ethics in planning. *J. Plan. Lit.* **1990**, 5 (2), 123–150.
 26. Beatley, T. *Ethical Land Use. Principles of Policy and Planning*; The John Hopkins University Press: Baltimore, 1994.
 27. Darwall, S., Ed.; *Consequentialism*; Blackwell: Oxford, 2003.
 28. Sen, A. Utilitarianism and welfarism. In *Consequentialism*; Darwall, S., Ed.; Blackwell: Oxford, 2003.
 29. Rawls, J. 5. Classical utilitarianism. In *Consequentialism*; Darwall, S., Ed.; Blackwell: Oxford, 2003; Chap. 1.
 30. Darwall, S., Ed.; *Deontology*; Blackwell: Oxford, 2003.
 31. Blanco, H. Community and the four jewels of planning. In *Planning Ethics*; Hendler, S., Ed.; Center for Urban Policy Research: New Brunswick, 1999; 66–82.
 32. Devall, B. The deep, long-range ecology movement: 1960–2000—a review. *Ethics Environ.* **2001**, 6 (1), 18–41.
 33. Brook, I. Can ‘Spirit of Place’ be a guide to ethical building? In *Ethics and the Built Environment*; Fox, W., Ed.; Routledge: London, 2000; 139–151.
 34. Mill, J.S. *On liberty and others writings*; Collini, S., Ed.; Cambridge University Press: Cambridge, 1993. (“On Liberty,” first published in 1859).
 35. Hendler, S. Feminist planning ethics. *J. Plan. Lit.* **1994**, 9 (2), 115–127.
 36. Ritzdorf, M. Feminist contributions to ethics and planning theory. In *Planning Ethics*; Hendler, S., Ed.; Center for Urban Policy Research: New Brunswick, 1999.
 37. Talbot, R.; Magnoli, C. Social inclusion and sustainable city. In *Fox, Warwick, Ethics and the Built Environment*; Routledge: London, 2000; 91–101.
 38. Cafaro, P. For a grounded conception of wilderness and more wilderness on the ground. *Ethics Environ.* **2001**, 6 (1), 1–17.
 39. Girardet, H. Greening urban society. In *Ethics and the Built Environment*; Fox, W., Ed.; Routledge: London, 2000; 15–30.
 40. Silva, C.N. A Carta do Novo Urbanismo e a Nova Carta de Atenas: A Utopia Urbana do Século XXI? (Chart of New Urbanism and the New Chart of Athens: The Urban Utopia of the XXI Century?). *Cad. Munic.-Rev. Ação Reg. Local* **2002**, XVI (79), 35–47.

Use and Management of Information Technology in State Government

Paul M. Chalekian

Department of Political Science, University of Nevada, Reno, Carson City, Nevada, U.S.A.

INTRODUCTION

The effective use and management of information technology (IT) is a major goal of every state. Almost all forms of government use some aspect of IT. Even within the most remote and rudimentary functions, attempts are made to try and improve efficiency. Personal computers are a standard tool and the use of spreadsheets is a common practice. As the numbers of Internet users increase, electronic government (e-government) portals have been developed as a way to reach those users. In terms of management, information resources must be cost effective at a time when calls for fiscal accountability are pervasive. Various forms of governance exist, resulting in debates about whether or not to centralize a state's IT resources. Unless otherwise noted, the references used will be about practices among the 50 American states. However, for many of the following concepts, the wider political term "state" could be interchanged. Although many variations in the use and management of IT exist, some common patterns of governance have been identified.

Max Weber's essay on "Bureaucracy" delineated elements of ideal-type organizations, and among these is a system of files.^[1] But unlike Weber's era, remote and complex organizations access those files, usually over a network. A database is a well-organized collection of files. Because of advances in communication technologies, it is common for an agency's database to reside in another bureau. A data warehouse is a collection of data which may originate from files or databases in one or more locations. Many feel that data warehouses should be managed by an information resource management (IRM) organization and those organizations often take on that role. Before describing more complex organizational structures, a historical perspective is in order.

HISTORY

Because of each bureaucracy's file continuity requirement, the evolution of IT in a state can be understood from a historical perspective. At the end of the 1960s, 90% of all data processing systems were supported in nine major areas: finance, revenue, highways, motor

vehicles, employment security, health and welfare, law enforcement, legislative and institutional functions.^[2] Unless a reorganization of agency functions occurred, the original location of file-related services tended to prevail. This was the case even more so within the largest state agency functions listed above. Recently, however, the portability of files has greatly increased the creation of options.

Similar to centralized purchasing and mail services, a type of "Holding Company" was envisioned by Gulick^[3] and these have often materialized among the states in the form of central data processing bureaus. For control purposes, the greater the role of a top executive in decision making, the greater the degree of centralization.^[4] In agencies late in adopting IT practices, their first system may have been developed as part of a controlled project. If developed within a central data processing bureau, staff are often trained to become highly specialized and their managers were anxious to expand the agency's client base. At the same time, the administrators of the slower adopting agencies may have sought out workload services from a central data processing organization.

Currently, the use of centralized agencies among the states is common; however, many bureaus procured and developed their own IT operations and support staff. Over time, the size of the files became less important, but the factors that influence control of the files became more so. When mini- and microcomputers appeared in increasing numbers, the centralization/decentralization debate radically changed.^[5] Because of lower equipment start-up costs, these systems could be more easily justified and procured; however, the ongoing expenses of having stand-alone systems were not always delineated. Whether or not an agency is highly networked, specialized staff are often needed to maintain their systems. As such, financing and cost recovery factors will be included in the state patterns described below.

SINGLE AUTONOMOUS AGENCIES

Ernest Alexander, in his book entitled *How Organizations Act Together*, identified *Single*, *Lead* and

Coordinating types of agencies.^[6] This is not to say that *Single* agencies are not networked or that other pooled structures cannot exist. Alexander's categories are useful in describing the current types of IT arrangements. In this modern era, few agencies are without interoperability with others; however, in functions involving specialized applications, encoded communications, or proprietary vendor arrangements, agencies often develop, maintain, and enhance their own computer resources.

Specialized applications within a state, such as voting, legislative, judicial or medical systems, are just a few examples of systems that usually retain separateness from other agencies. These autonomous agencies often have their own computing resources due to legislation or the unique nature of their specialized function. In an attempt to provide customer service, most have provided Internet and credit card interfaces. Agencies with more file intensive functions may also make kiosk or interactive voice response systems available to their customers.

Encoded communication requirements often separate law enforcement and other technology specific agencies. These requirements have increased even more in the post-9/11 world. In some instances, extensive connectivity requirements may be in place, utilizing advanced encryption schemes. As a result of these special security requirements, state law enforcement agencies are often the primary host for this integration function. As a primary point of contact for the federal government, the enforcement entities often act as a coordinator for local jurisdictions.

Proprietary vendor arrangements may also restrict whether the files of an agency can be hosted by another. Often developed in the private sector or by outsourcing, these systems, to name a few, may include the functions of corrections, health care, recreational reservations or gaming. Based on a vendor's contractual arrangement, the computer source code for manipulating a system may be restricted or denied. These functions may also have their own revenue sources or accounting systems.

Once procured, agencies that host their own computer systems must continually scrutinize their operating expenses. They must have the ability to obtain adequate funding for their specialized equipment, software, and staff. Whether these applications are specialized, encryption based, or originate from a proprietary vendor, automated interfaces to the state's accounting system may have to be established if the autonomous agency does not have an accounting system of its own. Staffing is a large expenditure. Recruiting IT specialists can be a challenge for many state bureaus. Positions may have an actual IT classification, or the autonomous agency may require their workers to do IT tasks while within a

non-IT professional classification. These later conditions often prevail as many individuals perform crucial IT functions along with their other agency duties.

Within these networked but autonomous agencies, IT governance can be delicate and complex. Usually a high level steering committee is formed containing members of the single agency's administrative staff. Typically a steering committee is chaired by the top IT representative, although the priority-making tasks usually originate from outside the realm of IT with each member of the committee trying to advance his/her functionality. A specialized project may detract from the development of an agency's enterprise framework, and this may prove critical for future compatibility. Meanwhile, the need to implement new legislation prevails.

The development of an entity's IT framework puts the group in a unique situation. For instance, Henry Mintzberg has identified individuals and/or groups as part of a separate support staff, existing outside the basic flow of work. He indicated the presence of a separate technostructure group, one task of which is to help organizations adapt to their environment.^[7] Within a single autonomous agency, it is possible that this technostructure may be developed internally, or private sector and/or outsourcing alternatives may be explored.

CENTRALIZED STORAGE

Diverging slightly from the single autonomous pattern, many organizations are increasingly dependent upon one another for information. Upon closer examination, some additional models have evolved as one agency structure may host the primary files of others. For instance, interfaces are often established to link agencies with similar functions, making common data available to share. Centralization may be caused by a crisis^[7] and a centralized bureau may pick up a client agency if it has lost critical data. Whether or not to centralize the processing of data has been a long standing debate^[5,8] and the pattern of centralization is elaborated upon below.

Rather than having each agency administer its own IT contracts and equipment, the centralized IRM form of organization may be used. In the 1960s and 1970s, its primary function was just to support and host other client agency data. Central data processing services are most likely to be under departments of Administrative Services or General Services departments.^[9] Similar to Gulick's "Holding Company," these agencies are equipped to provide centralized data processing storage and services. Programmers and other highly

specialized individuals are often staffed and controlled by the centralized bureau. They have a limited knowledge of client agency functions, so they often concentrate solely on supplying computer resources.

As opposed to agencies that have their own systems, computing economies of scale can often be obtained with many agencies storing their data on a single large system. Various types of charge-back arrangements are in place whereby the centralized bureau can get revenues from services rendered, as well as recovering the direct costs associated with storing client bureau files. For these centralized entities, the dominant organizational units were found to involve data processing operations, telecommunications, and policy and planning.^[10] With regard to the first two, they represent direct costs and can often be charged back directly to the client agencies. Operations are usually staffed for 24-hr operations, and specialized systems programmers are available in an emergency. With the third, policy and planning aspects are often considered indirect costs and are associated more with discretionary expenses.

Previously, James Thompson has described technologies as being *Long-linked*, *Intensive* or *Mediating* varieties. The last of these can serve as a broad example for the remaining descriptions. A *Mediating* technology is a form in which clients or customers "...are or wish to be interdependent."^[11] From a central bureau' perspective, all client agencies are anticipated to be serviced equally; however, contingencies and priorities soon become apparent and well known to even entry level computer operators. Among Alexander' categories, centralized storage agencies represent *Lead* organizations, charged with, or assuming "...the responsibility of coordinating the other organizations in the network."^[6] Because of these interorganizational issues and the computing risks associated with down time, more elaborate forms of computing governance have evolved.

CENTRALIZED COORDINATION

For those who utilize centralized IT services, complex dependencies can develop. Thus the hosting agency becomes a *Mediating* organization in accordance with Thompson^[11] and a *Coordinating* type of agency based on the descriptions of Alexander.^[6] The first and most likely place to find a policy and planning entity for IRM is under the state comptroller, finance, budget or treasury departments, with the second most likely place being under the chief executive' office.^[9] The authority of these agencies may start out with simply the oversight of standards, but it may progress into

project management services or large scale software development.

As in any state entity, the success of planning and project management is crucial. With less than optimal planning, the computing resource requirements of a single agency may slow the transaction processing statewide. Project management experience is usually more commensurate among the staff of a centralized agency. Typically they have participated in many more projects and utilized a greater variety of computing software and resources. Expertise in project management methodologies, such as the system development lifecycle, may be gained from the central IRM staff. However, if the resources between the agencies become scarce, a form of mediation among the agencies involved needs to occur while their priorities are deliberated and established.

Some states hold the files in an agency as a way of storage management, some develop and coordinate policy and standards for IT, and some do both.^[12] As such, the funding structures become more complex, the more the centralized agency takes on. A proprietary internal service fund type of accounting may be used. Direct computing costs, as well as the less quantifiable indirect costs, are often elements of client agency bills. Due to federal uniformity requirements, all agencies need to be billed in a consistent way. For an agency less dependent upon federal funding, this could fuel a desire to break away from the storage agreement with the centralized IT agency and become autonomous.

As an example of a cost-saving measure, the centralized IRM bureau may have approval authority for statewide IT hardware acquisitions. Being so empowered, the staff attempt to standardize procurement to increase the likelihood that purchases are more compatible with the state' computing framework. These standards are often set based upon what the centralized bureau' staff can adequately support. For instance, the amount of available expertise in a given programming language may influence the decision to endorse one software provider over another. Further, these decisions may be imposed upon client agencies. The licensing for software is often costly and so statewide discounts are often sought. The overall goal of these IRM agencies is to save costs and to develop and maintain employees who are specialized and can maintain systems and associated files.

To administer these complex bureaus, a cabinet level position is often required. A Chief Information Officer (CIO) is the person with the responsibility, authority and accountability for IRM in a state.^[12] These individuals are usually a senior executive and have a high degree of technical expertise. States with a CIO structure have performed well,^[13] but

sometimes this authority and responsibility is deemed too great for a single individual. As such, an Information Resource Management Commission may have authority. An IRM Commission refers to formal boards, commissions, committees or authorities established for the purpose of "...directing or managing the planning and implementation of information processing resources, policies, standards and services within the state."^[12] With such a wide scope, various user committees may be involved with the management of client agencies. As such, the setting of standards and priorities may be part of a committee's responsibilities. Committees may also be appointed by administrative officials from outside the centralized IT bureau. Some common committee functions may include those pertaining to security, strategic planning, project oversight, networking and communications, law enforcement, enterprise architecture, or employee development.

The coordination of Internet activities has extended the scope of many centralized bureaus. All of the American states route their Internet services through a single portal. In this way, one or more firewalls or other security measures can be invoked. This also forces state agencies to standardize, knowing that their clients' inquiries may originate from within the same site. Thus, a centralized pattern of frameworks, from state to state, has prevailed. The majority of constituents are able to access these sites for simple agency forms and information. More advanced applications are becoming increasingly available for the more file intensive functions.

CONCLUSIONS

Unlike Max Weber's era, 50 different configurations are used in the structure of American state government. In one study, it was determined that IRM cannot and should not be limited to a single model.^[10] It has been said that even in uncentralized systems, such as the Internet, at least some mutually agreed upon standards are needed.^[14] In their political environments, IT services must be efficient and effective so that agency clients find their services easy to use. From their state home pages, links to the executive branch agencies and other political subdivisions are often standardized and available. As indicated earlier, the financial struggles continue as client agencies seek to reduce their costs while the centralized bureau seeks to retain control. For every agency that pulls its files and services away from the hosting bureau, the costs may increase for those remaining.

Some close to practitioners of IT have said that centralization and decentralization is cyclical.^[15]

The need for economies of scale in state government often leads to a centralization debate. Once away from centralized support, the specialization of IT employees becomes increasingly necessary. The posting of static web pages may be simple, but the development of more sophisticated services, involving disparate files, is more complex. Only by planning and a careful analysis of requirements can the systems of a state be successful. The ability for any agency to train and retain project managers and specialized staff is an ongoing challenge. Adequate funding and fiscal control may allow for specific functions of a state's IT to continue, and the services provided must be continually evaluated. The importance of files and their integrity will continue to be a vital and vibrant issue for years to come.

REFERENCES

1. Gerth, H.H.; Mills, C.W. *From Max Weber: Essays in Sociology*; Oxford: New York, 1946.
2. National Association for State Information Systems. In *Information Systems Technology in State Government*; National Association for State Information Systems: Lexington, KY, 1970.
3. Gulick, L. Notes on the theory of organization. In *Papers on the Science of Administration*; Gulick, L., Urwick, L., Eds.; Institute of Public Administration: New York, 1937.
4. Simon, H.; Kozmetsky, G.; Guetzkow, H.; Tyndall, G. *Centralization vs. Decentralization in Organizing the Controller's Department*; Controllership Foundation: New York, 1954.
5. King, J.L. Centralized versus decentralized computing: organizational considerations and management options. *Comput. Surv.* **1983**, *15* (4), 319–349.
6. Alexander, E.R. *How Organizations Act Together: Interorganizational Coordination in Theory and Practice*; Gordon and Breach: Amsterdam, Netherlands, 1995.
7. Mintzberg, H. *The Structuring of Organizations: A Synthesis of the Research*; Prentice-Hall: Englewood Cliffs, NJ, 1979.
8. George, J.F.; King, J.L. Examining the computing centralization debate. *Commun. ACM* **1991**, *34* (7), 63–72.
9. Caudle, S.; Marchand, D.; Bretschneider, S.; Fletcher, P. *Managing Information Resources: New Directions in State Government*; School of Information Studies, Syracuse University: Syracuse, NY, 1989.
10. Caudle, S. Managing information resources in state government. *Publ. Admin. Rev.* **1990**, *50* (5), 515–524.
11. Thompson, J.D. *Organizations in Action: Social Science Bases of Administrative Theory*; McGraw-Hill: New York, 1967.

12. National Association of State Information Resource Executives. In *State Information Resource Management Organizational Structures: 1996 NASIRE Biennial Report*; National Association of State Information Resource Executives: Lexington, KY, 1996.
13. Lee, G.; Perry, J. Are computers boosting productivity? A test of the paradox in state governments. *J. Publ. Adm. Res. Theor.* **2002**, *12* (1), 77–102.
14. Cleveland, H. The future is uncentralized. *Publ. Admin. Rev.* **2000**, *60* (4), 293–297.
15. Sujitparapitaya, S.; Janz, B.; Gillenson, M. The contribution of IT governance solutions to the implementation of data warehouse practice. *J. Database Manage.* **2003**, *14* (2), 52–69.

BIBLIOGRAPHY

1. Cats-Baril, W.; Thompson, R. Managing information technology projects in the public sector. *Publ. Admin. Rev.* **1995**, *55* (6), 559–566.
2. Dawes, S.S.; Prefontaine, L. Understanding new models of collaboration for delivering government services. *Commun. ACM* **2003**, *46* (1), 40–42.
3. Fulk, J.; DeSanctis, G. Articulation of communication technology and organizational form. In *Shaping Organization Form: Communication, Connection and Community*; DeSanctis, G., Fulk, J., Eds.; Sage: Thousand Oaks, CA, 1999; 5–32.
4. Watson, H.J.; Carte, T.A. Executive information systems in government organizations. *Publ. Prod. Manage. Rev.* **2000**, *23* (3), 371–382.

Using Econometric Models for Public-Sector Decision Making

Richard G. Sims

*Director of Tax Policy, Institute of Taxation and Economic Policy,
Washington, District of Columbia, U.S.A.*

INTRODUCTION

The analysis of public-sector decisions involves consideration of the consequences that are likely to result from the decisions. The point of decision making, after all, is to affect the future. The more informed decision makers are regarding the consequences of their actions, the more likely they are to make decisions that achieve their intended objectives. Because policy decisions at the state and local levels frequently involve economic matters, analysts often turn to economic models for assistance. Models can help predict the likely path of the economy, forecast revenues and expenditures, or depict the regional economy as it might appear if a proposed policy is enacted.

An economic model is, by definition, a simplified description of economic reality. Simplification focuses attention only on the economic relationships that produce meaningful observations about the policy question at hand. This article addresses economic models that are useful when the issue of concern relates to state and local public policy decisions. Such models may be developed in-house or be acquired “off-the-shelf,” and may be relatively simple or highly complex. Policy models often start as input–output models but are enhanced to take into consideration an array of factors that improve their forecasting ability and their usefulness in depicting the results of policy changes. Models that address geographic or political areas smaller than the national economy are referred to as regional models. The selection of a particular model and its enhancements will depend on the unique requirements of the respective decision-making environment; the user’s taste, preferences, and abilities; and the organization’s resource constraints.

In any case, there are certain attributes and relationships that the model should contain and operational characteristics that the user should bear in mind. Models that incorporate interindustry economic linkages and fundamental economic principles (rising prices lead to lower quantity demanded, etc.) are particularly useful and are sometimes referred to as “well-described” economic models. Well-described economic models perform two important functions: 1) they produce results that are useful in making decisions; and

2) they provide a framework for thinking about and understanding the economy. A well-described model may be viewed as a tool that depicts many of the complex interactions in an economy, although it may not include all of the aspects that an informed observer might be aware of. In this sense, the model serves as an organizing tool that provides a structure for the user to generate a clear, logical, and internally consistent picture of the regional economy and of various alternative futures.

Among the primary uses of economic models for state and local public policy decision making involve forecasting economic activity; analyzing tax and spending measures; and analyzing economic development proposals. Other uses include analyzing changes in business, environmental, or employment regulations; the local consequences of changes in national policy; and changes in energy policy or prices. A model can be used to aid policy makers in developing and evaluating alternative futures for the state or local economy. This article is intended to provide public-policy professionals with a better understanding of how economic models work and an appreciation how such models can be used to improve public decision making. What follows is an overview of some of the main functions of economic models when used to address some of the more common state and local government economic issues.

FORECASTING

Policy questions are, by their nature, about influencing the future, and consequently, policy practitioners are often asked to provide forecasts of economic activity. Such forecasts may be accomplished by simply projecting recent trends into the future using, for example, simple projections or linear regressions. This approach may produce satisfactory results over a relatively short time frame in a relatively stable economic environment. However, this simplistic one-variable model assumes nothing other than population matters and explains little. A simple mathematical expression that states that the demand for public services (p.s.) is

\$1000 for each person in the population,

$$\text{p.s.} = \$1,000 \times (\text{POP})$$

constitutes a model and could be useful in giving a rough approximation of the costs of providing services. The opportunity costs of using this economical approach are as follows: 1) reduced forecast reliability, particularly over long periods of time or in dynamic periods; and 2) the loss of an opportunity to increase the understanding of the local economy. Neither cost is trivial. A small improvement in forecast accuracy can make the difference between a budget surplus and having to raise taxes and cut public programs. An increased understanding of the regional economy can lead to improved decision making with resulting improvements in the efficiency of public policies and their effectiveness in achieving the intended policy goals.

A first step in improving forecast accuracy beyond projecting the trends in the variables themselves involves adding consideration of outside influences on the regional^a economy. Because the economies of even the largest states or even multistate areas are influenced heavily by trends in the broad general economy, forecasters typically start with forecasts about the larger national economy developed “exogenously,” i.e., outside the regional model. This forecast of the national economy becomes the “driver” for the regional model.^b The regional economic activity associated with changes or trends in the broader economy is then estimated in the model. For example, forecasts of changes in the gross domestic product, national employment, and income trends are viewed as among the forces that propel regional output, employment, and income.

To be more reliable as a predictive and explanatory tool, a model should capture the dynamic structure of the real economy. That is, it should incorporate feedback responses depicting the reactions that occur throughout the economy when a policy or economic variable changes. For example, an increase in demand for locally manufactured auto parts 1) pushes up regional income and wages, 2) attracts new workers into the area, 3) increases the demand for local housing, and 4) raises the cost of production of other regionally manufactured goods. Each of these affected sectors, in turn, has their own linkages and produces their own impacts. The forecast thus becomes a statistical projection based on historic regional trends *plus* explicit consideration of major outside trends affecting

the local economy *plus* consideration of the dynamic influences of basic economic fundamentals.

Regional models may be constructed in ways that the variables they address are regional (recognizing, of course, that regional variables implicitly incorporate national trends), or they may present a full array of regional and national variables. A regional model is usually sufficient when comparing local policy options, considering local or regional initiatives, or evaluating exogenous events that affect only the region. A more robust model that allows the user to directly control regional and national economic variables can address the effects of a national policy on the region or can explore the regional effects of alternative national forecasts.

Though similar in structure, regional and national models have certain intrinsic differences. Consider, for example, interest rates, which are one of the most significant factors affecting economic activity. At the national level, the monetary system is endogenous. The reactions of the Federal Reserve to changes in the unemployment rate and the resulting consequences for inflation are directly related to the national economy and should be explicitly considered in a national model. The model containing multi-year features could allow an interest-rate response to be incorporated into the model to return the economy to its control path after a period of adjustment to a policy stimulus. Regional economies, on the other hand, are open systems, where money flows freely from one area to another at a common interest rate. There is no regional monetary policy.

TAX AND EXPENDITURE ANALYSIS

State and local governments regularly confront decisions that have direct and indirect economic consequences for the region and individuals they represent. Decisions relating to the tax system and to government spending have the potential to influence economic activity. Taxes, such as income taxes, sales taxes, business taxes, and property taxes, impact the economy, though each has distinct and operationally different effects. The ability to distinguish these differences enhances a model's accuracy and utility. Models can be used to address additional tax and expenditure issues such as distributional impacts and balanced budget requirements.

Individual Income Taxes

Individual income taxes reduce purchasing power by directly reducing a consumer's disposable income. An increase in income tax rates will slow economic activity

^aThe term “regional” is used herein to mean any subnational economic unit, whether city, county, or state, or other units.

^bA national forecast, in turn, will have incorporated the international economy within its findings.

and cause the collections of consumer spending-related taxes, such as sales taxes, to decline. Over the long term, an income tax increase will slow inward population migration by reducing workers' disposable income. In addition, because the economic base has declined, each percentage point of nominal tax will produce less revenue as tax rates rise. Some income tax proposals, such as those that apply only to the top marginal rates, for example, may require an independent estimate of the tax base to which the rate change is applied. Significantly, the ultimate burden (the incidence) of an individual income tax increase falls almost exclusively on income earned within the region (i.e., there is no burden shifting).

Sales Taxes

A sales tax increase affects the economy as if it were an increase in the price of locally purchased goods and, consequently, a reduction in income available for spending on other goods. A significant feature of the sales tax is that for some regions, such as tourist destinations or regional trade centers, a substantial portion of the total sales tax burden may shift to visitors to the area, therefore, raising revenue for area services which is not paid by area residents. Alternatively, for less fortunate tax jurisdictions, particularly those that may be located near other jurisdictions with significant trade competition, an increase in a sales tax rate may produce a substantial shift in consumer spending to those surrounding jurisdictions. Over the longer term, sales tax increases deter inward migration by widening the gap between nominal personal and real disposable income. A highly specified economic model that takes the price elasticity of sales within the specific region into consideration can be extremely useful in accessing sales tax impacts.

Business Taxes

The effect of a change in a business tax is to change the implicit rental cost of capital. Raising a business tax makes capital relatively more expensive, lowering the number of potential projects that would meet a given investment criteria and reducing investment in the region. In addition, for new and existing projects, business taxes raise their local production cost relative to other regions. To appropriately depict these features requires consideration of various macroeconomic features, such as capital and labor substitution and import and export substitution, as well as multiyear considerations to address business tax features, such as accelerated depreciation or capital investment credits, which change the timing and size of business decisions. Businesses also pay a portion of the general taxes of

a region, such as sales taxes and property taxes, and business employment cost may rise when an increase in individual income taxes cause firms to have to pay more to attract or retain workers. The returns to ownership of business are typically assumed to be widely distributed, and therefore, regional personal income is not directly reduced by business taxes. These considerations can be built into the model to improve its ability to describe the real economy.

Property Taxes

Local governments, in particular, are often concerned about the effect of property taxes. In most regions of the United States, property taxes are left to local governments to fund local services, especially public schools. Property taxes affect the cost of capital and consumer purchasing power. Capital-intensive businesses, such as auto dealerships or petroleum refineries, are particularly sensitive to property tax rate changes. For individuals, the cost of home ownership rises, leaving less money available for spending on non-housing items. Modeling a property tax change requires converting the exogenously obtained estimate of the dollar amount of the property tax change into a capital cost change. An estimate of the share of the tax change that is directly paid by homeowners is used to reduce household disposable income and, therefore, consumer spending on other items. The dollar amount of the change and the share paid by households is generally obtained from the official property tax authorities, because their data will reflect the actual base to which the rate change will be applied and might differ from the base obtained from secondary sources. The burden of property taxes falls initially on property owners, though part of the burden is ultimately shifted to renters and consumers, depending on the respective elasticities.

DISTRIBUTIONAL IMPACTS

Policy makers are often concerned with the distributional impacts of tax changes. The effect of a tax change on households of various income levels might be estimated by applying the results of a tax revenue simulation, which typically presents results in a fairly broad aggregate form (i.e., a change in total personal income), to a distribution model that shows household spending at various levels of income, for example. Satellite models of household spending can be developed from Consumer Expenditure Survey (CES) data, for example, to analyze the impact of sales tax changes by household income. Income tax and other tax distributional models can be developed using data from

sources such as state revenue departments, the Internal Revenue Service (particularly, the *Statistics on Income* bulletin), the U.S. Bureau of Census, the Bureau of Economic Analysis, Bureau of Labor Statistics, and County Business Patterns. The CES data present a matrix of household spending by income category. This matrix can then be reviewed to identify the specific items that would be subject to the sales tax.

THE TAX-AND-SPEND AXIOM

This axiom is the state and local government finance equivalent of Sir Isaac Newton's famous axiom that "for every action there is an equal and opposite reaction." Namely, that for every tax dollar collected, there is a dollar of public spending; and its flip-side corollary, that for every dollar increase in government spending, there is a dollar increase in taxes. This axiom follows directly from the fact that state and local governments generally require some form of balanced budgets, implying that they must spend what they take in and may only spend what they take in.^c Consequently, a dollar taken out of consumer spending in the form of taxes does not simply disappear into a desk drawer in the treasury or into an economic black hole, where money goes in but nothing comes out. Rather, a dollar of tax is transformed into a dollar of government spending. Similarly, each dollar of government spending is not a gratuitous event but must ultimately be paid for by taxes, which should be accounted for in a comprehensive analysis.

Considering only the effects of changes in taxes would show, unsurprisingly, that tax increases have negative economic impacts, because, by themselves, taxes represent a cost increase and a reduction of disposable income. Conversely, all proposed public expenditure projects would show positive economic impacts, if only the benefits of the projects are considered, and their costs in terms of taxes are ignored. Ignoring either side of the tax-and-spend equation would be comparable to a business evaluating the benefits of a new piece of equipment without asking the cost, or deciding to raise the price of its product by a large amount without considering the broader consequences of the decision. In the real world of public-policy analysis, it is not unusual for a policy proposal, such as a plan to cut taxes or a program involving public spending, to refer to only one side or the other

of the tax-and-spend equation. A tax cut may be proposed with no specific reference to the spending measures to be reduced as a consequence, or a government project will be proposed with no reference to the source of its funding. Even when the full ramifications are not specified in the proposal, the analyst should bear in mind that these consequences exist and should be addressed in some appropriate manner. To ignore either side of the tax-and-spend equation is to, in effect, advocate for one side or the other of the proposal. If a proposed tax cut is proposed and specifies no particular expenditure cuts, an analyst might, for example, use an "across-the-board" reduction in existing programs. A proposed new program with no specific funding might be matched with an "across-the-board" tax increase, or a proportionate reduction in all other programs.

The caveat is that the tax-and-spend axiom may appear to be avoided, at least temporarily, by the issuance of debt. Debt allows current spending to exceed current revenues by allowing the expenditure payments to be spread over future time periods. Debt financing is accounted for in a model by considering the discounted net present value of the stream of debt service payments. The full implications of debt financing should be specified in the analysis.

A final point on expenditures is that the *composition* of government spending is economically important and should be specified as clearly as possible. Some components of government spending, such as elementary education, are relatively labor intensive and regionally acquired, producing relatively high local employment impact from government effect. Other components, such as technology and scientific equipment, tend to be capital intensive and to be produced outside the region. Often, a government spending program will involve more than one facet to consider. Highway spending, for example, may contain a construction-related component, producing construction jobs and income, and a transportation cost reduction component, lowering regional business costs and increasing transportation productivity in the region. Similarly, education spending is a source of jobs for teachers, administrators, cooks, and other support personnel; a source of increased labor productivity in the region; and an attractive amenity luring business and people to the community. Spending on corrections, the fastest-growing component of most state government budgets, is a producer of local jobs but does little to improve the economic climate of the region.

ECONOMIC DEVELOPMENT PROPOSALS

Policy analysts are often called upon to proclaim on the impact of a proposed new manufacturing plant,

^cThis is radically different from the federal government, where spending and revenues seldom balance. In fact, the difference between the two is the basis for national fiscal policy, where spending intentionally exceeds taxes in times of slow economic growth, and taxes may be set at a level above spending when the economy is "overheated."

sports coliseum, or tourist attraction. Economic models are often used to predict the direct and indirect effects of a new activity on the economy of the region and are sometimes used in the development or evaluation of “incentive packages” used to attract targeted new projects. The direct effects of a project are typically straightforward and readily identifiable. In general, the project will be known to entail X millions of dollars in construction spending over a certain period of years and directly employ Y number of workers when completed, with an annual payroll of Z . A project will also have indirect effects, or linkages, to the local economy, i.e., the “multiplier effect.” To obtain these indirect effects, the model estimates the share of demand for the particular project inputs that will be supplied by existing local producers. The direct and indirect effects add new household income that is spent locally and is referred to as the “induced effect.” Each of the effects adds to local economic activity and the local tax base and represents the benefits of the project.

Local multiplier effects are typically quite small relative to state-level effects, which in turn, are quite small relative to national-level effects. The larger the economic unit, the more likely that inputs will be supplied from firms located within the unit. The locally supplied input share, sometimes referred to as the Regional Purchase Coefficient or Locally Supplied Inputs, is obtained by looking first at the input matrix for the new project and then at the existing array of local suppliers to identify potential local suppliers. Determinations must then be made regarding the likely proportion of the new project’s total demand for inputs that will be satisfied by these local suppliers. An algorithm to address industry-specific considerations, such as supplier capacity relative to project demand and historical shipping patterns, can be used for this purpose. For relatively small economic units, the estimation of the expected locally supplied inputs can be exceptionally difficult. In a relatively large economy with a large number of firms in the potential supplier pool, the likelihood of one or more of them supplying the required inputs is fairly high. At the local (say, county) level, it is not unusual to have only one, if any, supplier of certain required inputs, and that potential supplier may face capacity constraints, be a dedicated supplier to an existing firm, or not produce the specific type or quality product demanded by the new facility. While a large supplier pool would tend to average these firm-specific difficulties, in local-level analysis, firm-specific knowledge is extremely beneficial.

In some cases, growth in one industry may, as part of its indirect effects, come at the expense of other existing industries. A new casino, for example, may reduce the demand for existing local restaurants and other non-casino recreation activities. A new

manufacturing plant, in another case, might drive up prevailing wages, crowding out some existing industries and making investment in other industries less profitable. In addition, economic development projects generally necessitate some (often considerable) new public expenditures for roads, sewer and water lines, public protection, education, and so forth. To accurately assess a new activity, it is essential to carefully identify all aspects of the proposal, explicit and implied. New, large projects are often given incentives in the form of grants, tax credits, and exemptions, which directly reduce the net benefits that would otherwise be attributable to the project. If the public expenditures for the facility plus the general support cost for new project-related residents exceed new project-related revenues, then other taxes must be increased or other government spending must be reduced.

A new project coming into an area typically meets its employment needs by paying wages that are at least marginally above the prevailing wages being paid in the area for similar labor. This increase in demand stimulates an in-migration to the area of job-seekers and their families, which, in turn, places new demand on public services—schools, streets, police and fire protection, and so forth. The costs of providing these added public services offsets some of the public revenue gains from new development.

BUSINESS, EMPLOYMENT, AND ENVIRONMENTAL REGULATIONS

State and local governments frequently become involved with issues of business regulation, employment regulation, and environmental protection. Business regulations might take the form of limits on the types and size of new industry coming into the area, the imposition of fees and user charges for certain activities, or certain mandatory investments in order to engage in particular businesses. Environmental protection regulations may involve requirements such as land set-asides, the provision of certain amenities, or the compulsory installation of pollution-abating equipment or techniques. Employment regulations may involve limits on hours individual employees may work, the payment of a statutory prevailing wage that may exceed the actual community wage, and minimum wage requirements for low-income workers.

Much of the required information for evaluating regulations, such as the direct cost of pollution-abatement equipment or the share of the workforce that would be affected by changes in the local minimum wage, would be obtained outside the model. The effects of environmental regulations, in particular, are often subtle and require careful analysis of the specific regulation under consideration. The economic impacts of regulatory

changes can usually be determined by careful and explicit consideration of direct impacts in terms of their economic fundamentals (increase in cost of capital, increased employment costs, lowered demand for labor, etc.) then following the economic implications of these changes.

REGIONAL IMPACTS OF CHANGES IN NATIONAL POLICY OR CONDITIONS

Changes in national policy or economic conditions can influence the regional economy as part of their effect on the national economy as a whole or through their relative and direct effects on the particular area being considered. To analyze the first type of impact, it is always necessary to perform a national simulation as part of carrying out a regional analysis. In the second case, the analysis can be conducted using a model with only regional variables and data.

An example of a nation policy change having significant, but diverse, impacts on local regions and governments would be the national health care proposals that were considered in the early years of the Clinton Administration. The principle health care proposals put forth involved massive regional spending shifts, changes in the composition of health care spending, and substantial increases in tax revenues coming from employees (in the form of an employee payroll tax) and businesses (in the form of a tax on their total payroll). Detailed analyses of the health care reform proposal were conducted at the state level. These studies indicated that in addition to national changes, there were substantial shifts in employment and income within industries and among states. Interindustry shifts occurred as labor-intensive industries became less competitive, and as many industries substituted capital for now more expensive labor. State impacts depended on their taxable income relative to their health care requirements, comparative industrial composition and labor intensity, the relative portion of the population previously uninsured, and comparative costs of delivering health care services.

ENERGY POLICY OR PRICES

To appreciate the significance of energy prices in the U.S. economy, one only has to recall that sharp rises in petroleum prices led directly to the two major recessions affecting this country since World War II. Energy costs affect the economies and governments of every region of the United States. Energy-producing regions may generally benefit from higher prices, which leads to higher production, higher employment, and higher wages. The governments in producer regions reap

increased revenues from production taxes as well as revenues from the higher level of activity in the region. Even energy-producing regions, however, face higher prices for this basic consumer item, which is also a major input into other industries.

Because energy use is ubiquitous, fuel costs significant, and fuel prices volatile, energy costs are an important aspect in all parts of the economy, even though direct fuel costs represent only a small portion of total production costs for most industries. An additional consideration in addressing energy issues is that fuel price changes often result from changes in government policies (national, international, or regional) as well as market forces. An economic model can be used to aid in viewing the effect of energy price changes as they relate to electric rates, natural gas rates, or transportation costs, and the resulting implications for the local economy. For example, a new state severance tax on natural gas increases consumer costs and business costs in the affected region. This leaves energy consumers with less to spend on other goods and makes the business in the region less competitive. The higher price for natural gas leads to a shift toward other energy sources, which may be imported from outside the region. If a substantial portion of the natural gas production is shipped outside the region, part of the tax burden can also be exported. The extent of the tax exporting will depend largely on the price elasticity of sales to the export markets.

A fuel price increase simulation is accomplished by increasing the appropriate fuel price variable in the model and by decreasing consumer purchasing power to reflect the extra amount that consumers will have to pay for energy. To complete any simulation, the effects on the production side must also be considered. Because fuels are often imported, the effect of changing energy strategies may be to substitute local production for imports or to replace local production with an external energy supply. Thus, even when the direct regional effects of fuel price changes are in one direction, the choice of energy source may have direct effects that will change the sign of the original effects. For example, higher petroleum prices lead to lower petroleum use and lower direct petroleum-related employment. They may also lead, however, to substitution to locally produced natural gas, which might have greater direct employment effects.

SOME SUGGESTIONS FOR USERS OF ECONOMIC POLICY MODELS

There are certain “tricks of the trade” in using economic models. Practicing these techniques should

allow the user to improve their analysis and to avoid time-consuming difficulties:

1. Check to see if the results make intuitive sense. Look carefully at the “big numbers,” such as total personal income, total employment to develop a perspective of the analysis. It is often easier to spot changes that seem disproportionate or counter to experienced expectations when looking at the big numbers rather than the more specific details.
2. When simulating a change that requires using several variables, run each variable as a separate scenario, and check the results carefully to develop an understanding of the specific effects of each. The results of using each variable should make sense on its own. Remember, however, that when using two or more variables (in a model with built-in feedback features), the results from using the variables together may, in fact, be different than the sum of the variables used separately.
3. When comparing two or more scenarios, make certain that the same methodology was used in developing each. Different methodologies can produce outcome differences that are purely a function of the approach, not the policy. This can be particularly significant when considering multiperiod simulations, where methodological differences are compounded with each period.
4. Graphing the results often reveals patterns that may not be otherwise apparent. Graphs are also often easier for audiences to quickly comprehend.
5. Economic models cannot read legislation. It is the model users responsibility to insure that the analysis is conducted accurately and reflects the actual policy proposal. Subtle nuances in statutory language can produce substantial differences in economic outcomes.

APPENDIX

There are numerous sources for regional economic policy models. Regional Economic Models, Inc., of Amherst, Massachusetts (www.remi.com), and the Minnesota IMPLAN Group of Stillwater, Minnesota (www.implan.com), are among the leading commercial sources of vendor-supplied models, while the Bureau of Economic Analysis will provide tables of regional

multipliers that can be used to develop independent models. In addition, many universities and research institutions offer various forms of regional policy models, some of which may be highly specialized (focusing on an industry such as agriculture, tourism, or the utility sector, for example), relatively simplistic (and relatively inexpensive), or are primarily forecast (as opposed to policy analysis) oriented.

BIBLIOGRAPHY

1. Almon, C. *The Craft of Economic Modeling*, 4th Ed.; University of Maryland: College Park, MD, 1999.
2. Bradford, D.F., Ed.; *Distributional Analysis of Tax Policy*; American Enterprise Institute: Washington, DC, 1995.
3. DRI/McGraw-Hill. *An Assessment of Input–Output Models*; 1994, report prepared for U.S. Department of Transportation by; 62 pages plus appendix.
4. Lahr, M.L. *Comparison of Ready Made Regional Economic Models in the U.S.*, Paper Given at Economic Impacts of Historic Preservation Conference; Rutgers University: Rutgers, NJ, 1996.
5. Maki, W.; Olson, D.; Lindall, S.; Senf, D.; Schallau, C. *IMPLAN Modeling Applications in State and Regional Development*; Department of Agricultural and Applied Economics, University of Minnesota: St. Paul, MN, 1989.
6. National Research Council. *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling, Volume I: Review and Recommendations*; National Academy Press: Washington, DC, 1991.
7. National Research Council. *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling, Volume II: Technical Papers*; National Academy Press: Washington, DC, 1991.
8. Rickman, D.S.; Schwer, R.K. A comparison of the multipliers of IMPLAN, REMI, and RIMS II: Benchmarking ready-made models for comparison. *Ann. Reg. Sci.* **1995**, 29, 363–374.
9. Stevens, B.H. *PC I-O Version 7*; Regional Science Research Corporation: Hightstown, NJ, 1996.
10. Treyz, G.I.; Norwell *Regional Economic Modeling: A Systematic Approach to Economic Forecasting and Policy Analysis*; Kluwer Academic Publishers, 1993.
11. Treyz, G.I.; Rickman, D.S.; Shao, G. The REMI economic-demographic forecasting and simulation model. *Int. Reg. Sci. Rev.* **1992**, 14 (3), 221–253.
12. U.S. Department of Commerce, Bureau of Economic Analysis. *Regional Input–Output Modeling System (RIMS II): Estimation, Evaluation, and Application of a Disaggregated Regional Impact Model*; U.S. Government Printing Office: Washington, DC, 1981.

Validation

Teva J. Scheer

University of Colorado at Denver, Highlands Ranch, Colorado, U.S.A.

INTRODUCTION

Validity refers to the degree to which a test or other selection devices measure a candidate's possession of specified characteristics. The Uniform Guidelines on Employee Selection Procedures, which provide companies and government agencies guidance on ensuring that their selection processes meet acceptable standards in discrimination cases, specify three types of validation strategies that employers may use for their selection devices. This study explains these methods and provides general guidance on when each is appropriate.

WHAT IS VALIDITY?

Besides the definition provided for validity earlier, one eminent statistician describes *validation* as “the process by which a test developer . . . collects evidence to support the types of inferences that are to be drawn from test scores.”^[1] Selection devices include written tests, qualification standards, assessment centers, performance tests, and employment interviews. Examples of characteristics that a selection device might measure range from innate ability (such as IQ), to interpersonal characteristics (like dependability), to job qualifications (knowledge of accounting principles being an example). The three strategies specified in the Uniform Guidelines are criterion-related, content, and construct.

CRITERION-RELATED VALIDATION

Criterion-related validation is a statistical method in which the results of a selection device (e.g., test scores) are compared to a measure of on-the-job performance criteria (e.g., performance appraisal scores, quality rates, or productivity statistics). The relationship between the scores and the performance measures is then computed using a validity coefficient. The higher the validity coefficient, the greater the validity of the selection device.

Some of the issues associated with criterion-related validation studies are: (a) selection of meaningful and measurable performance criteria; (b) insufficient

sample size to meet statistical standards; (c) in-house possession of the necessary statistical expertise to perform a criterion-related study or the funds to contract outside expertise; and (d) restriction of score range. The last of these relates to the need to ensure that the sample reflects the entire range of test scores, from high to failure. This is a particular problem for most government agencies that are restricted to selecting from among the highest-scoring candidates.

CONTENT VALIDATION

Content validation studies typically rely on subject matter experts or SMEs, i.e., supervisors and workers who are experienced and knowledgeable about the occupation, and who work under the guidance of a human resources specialist trained in job analysis techniques in developing the selection instrument. In a content validation study, the SMEs complete a structured job analysis to identify and organize the content elements (e.g., tasks or major job duties) of the job to be filled. The SMEs then identify examples of experience or education that would be acceptable evidence of the content elements. These are bundled into a scorable rating guide. For example, if SMEs in a customer call center determine that problem resolution is a major job duty for their customer service representatives, they would need to identify job experiences, duties, or training classes that are evidence a candidate can perform this important task. Each candidate is then scored, based on the relevance of his or her experience and education to the evidences provided by the SMEs.

Content validation relies on logic rather than statistics.^[2] It is an appropriate method where the limited number of selections and/or applicants precludes the use of a statistical method—for example, when a local government agency needs to advertise for and select a Budget Analyst. While a rigorous job analysis is required for each type of validation, it is particularly critical for logical methods of validation. The written documentation of the job analysis establishes the link between the job content and the evidences that were identified by the SMEs and formed the basis for ranking the candidates. Content validation relies on the following syllogism: If $A = B$ and $B = C$, then

$A = C$. For example, if the [SMEs' rating guide] measures [skill in writing technical reports] ($A = B$) and [skill in writing technical reports] is critical for [success on the job] ($B = C$), then a [high rating guide score] will result in [success on the job] ($A = C$).

CONSTRUCT VALIDATION

Construct validation relies on the same logical syllogism as content validation. However, construct validity job analyses focus on identification of psychological constructs rather than job content elements. Constructs, unlike content elements, are not directly observable. In a construct validation study, a psychologist analyzes the constructs necessary for success on the job and then identifies tests that measure these constructs. For example, if an identified construct is high intelligence, then the Wechsler tests could be used to assess candidates' IQs. If an identified construct is high extroversion, then the Myers Briggs Type Indicator could be utilized to assess candidates on the extroversion–introversion scale.^[1]

SELECTING THE APPROPRIATE TYPE OF VALIDATION STRATEGY

In 1979, the Equal Employment Opportunity Commission (EEOC) issued the first set of questions and answers designed to clarify the Uniform Guidelines' requirements for employers. In response to a question about how to determine when each of the validation strategies is appropriate, the EEOC stated:

“A user should select a validation strategy or strategies which are: (1) appropriate for the type of selection procedure, the job, and the employment situation and (2) technically and administratively feasible. Whatever method of validation is used, the basic logic is one of prediction; that is, the presumption that level of performance on the selection procedure will, on the average, be indicative of level of performance on the job after selection. Thus, a criterion-related study, particularly a predictive one, is often regarded as the closest to such an ideal.... Key conditions for a criterion-related study are a substantial number of individuals for inclusion in the study, and a considerable range of performance on the selection and criterion measures. In addition, reliable and valid measures of job performance should be available, or capable of being developed.... Where such circumstances exist, a user should consider use of the criterion-related strategy.

“Content validity is appropriate where it is technically and administratively feasible to develop work samples or measures of operationally defined skills,

knowledges, or abilities, which are a necessary prerequisite to observable work behaviors. Content validity is not appropriate for demonstrating the validity of tests of mental processes or aptitudes or characteristics; and is not appropriate for knowledges, skills or abilities which an employee will be expected to learn on the job....

“The application of a construct validity strategy to support employee selection procedures is newer and less developed than criterion-related or content validity strategies. Continuing research may result in construct validity becoming more widely used. Because construct validity represents a generalization of findings, one situation in which construct validity might hold particular promise is that where it is desirable to use the same selection procedures for a variety of jobs. An overriding consideration in whether or not to consider construct validation is the availability of an individual with a high level of expertise in this field.”^[3]

ARTICLE OF FURTHER INTEREST

Uniform Guidelines on Employee Selection Procedure, p. 1995.

REFERENCES

1. Crocker, L.; Algina, J. *Introduction to Classical and Modern Test Theory*; Harcourt Brace Jovanovich College Publishers: Orlando, FL, 1986; 217, 230–235.
2. Lyman, H.B. *Test Scores and What They Mean*, 6th Ed.; Allyn & Bacon: Boston, 1998; 10 pp.
3. Questions and answers to clarify and provide a common interpretation of the Uniform Guidelines on Employee Selection Procedures, Equal Employment Opportunity Commission, 44 Fed. Reg. 11996. U.S. Government Printing Office: Washington, DC, 1979; Question No. 54.

BIBLIOGRAPHY

1. Cronbach, L.J. Test validation. In *Educational Measurement*, 2nd Ed.; Thorndike, R.L., Ed.; American Council on Education: Washington, DC, 1971.
2. Edenborough, R. *Using Psychometrics: A Practical Guide to Testing and Assessment*, 2nd Ed.; Kogan Page Limited: London, 1999.
3. Hair, J.F., Jr.; Anderson, R.E.; Tatham, R.L.; Black, W.C. *Multivariate Data Analysis*, 5th Ed.; Prentice Hall: Upper Saddle River, NJ, 1998.
4. *Uniform Guidelines on Employee Selection Procedures*, Title 29, Part 1607 of the United States Code of Federal Regulations; Government Printing Office: Washington, DC, 2004.

Values and Policy Analysis

Steven A. Peterson

*School of Public Affairs, Pennsylvania State University at Harrisburg,
Middletown, Pennsylvania, U.S.A.*

INTRODUCTION

Policy analysts cannot avoid the importance of values in their work. Policy analysis is grounded in values. This article focuses on a number of issues: the role of values held by the policy analyst, by policy-makers, and by the larger society. Values at each of these levels affect the nature of policy analysis and the conclusions that emerge from such analysis. Value conflicts can arise from this situation, and the variety of responses available to policy analysts when conflicts occur are described. In a very real sense, there can be no “value free” policy analysis.

INTRODUCTORY COMMENTS

Values are an integral part of the process of policy analysis—from selecting questions to be asked or problems to be addressed, to selecting items for the policy agenda, to determining which policy option appears most likely to meet problems identified as worthy of government action.

Indeed, in one of the seminal essays regarding policy analysis, Harold Lasswell was quite explicit regarding the role of values. For instance, he notes that “The policy-science approach not only puts the emphasis upon basic problems and complex models, but also calls forth a very considerable clarification of the value goals involved in the policy.”^[1] In a larger sense, the value of democracy is central for Lasswell’s policy orientation. As he says, “It is probable that the policy-science orientation in the United States will be directed toward providing the knowledge needed to improve the practice of democracy. In a word, the special emphasis is upon the policy sciences of democracy, in which the ultimate goal is the realization of human dignity in theory and fact.”^[1]

INDIVIDUAL VALUES AND THE POLICY ANALYST

Paris and Reynolds observe that each person has values that are central to himself or herself. They refer to these as “Individual Overriding Values” (or

IORVs). These tend to be stable and they rank very highly in a person’s hierarchy of values. Paris and Reynolds further note that “They are called overriding because owing to their features, they can be used to override other lesser values and to resolve conflicts between other values when these occur.”^[2] When we examine the world of the policy analyst, IORVs become valuable to understand, as there may be times when the analyst’s fundamental values come into conflict with those of policy-makers. At such times, value conflict can prove to be a challenge for the analyst.

What is apt to be the outcome of such conflicts? Weimer and Vining^[3] use the work of the economist A.O. Hirschman^[4] to outline the options available under such conditions. They assert that the analyst has three choices—exit, voice, and disloyalty. For example, if key political figures demand “cooked results” (i.e., that the analyst skew results to fit the politicians’ desires) or if political figures misrepresent results of policy analysis in public statements, what is to be done?

Exit would be the situation where the analyst feels that he or she has been so compromised that it is necessary to leave the organization. Voice would represent those circumstances where the analyst would make the effort of working from within an organization to change its way of doing business. Finally, and perhaps most dramatically, disloyalty. For instance, if the agency were trying to suppress a policy analysis that questioned the choice actually made by the organizational leaders, then the individual might leak the original results to the media. Thus if an individual’s values were violated by the organization’s behavior, he or she could exercise exit, voice, or disloyalty.^[3]

VALUES OF POLICY-MAKERS AND POLICIES

Traditionally, a small array of values has been described as most salient for policy-makers’ selection of policy choices. Among the most important: equity, efficiency, and effectiveness. Equity is normally defined rather straightforwardly as “treating likes alike.”^[5] Another view is that equity is to be defined in terms of “fairness.”^[3] Or, as Peters puts it, “. . . it is argued that government has the legal and economic capacity

to redress inequities in the distribution of goods and services that result from the operation of the marketplace.”^[6] Unfortunately, once one gets past these generic definitions, the actual means by which one applies equity to policy choices is not always clear. Stone has used the example of dividing up a piece of pie “fairly” in her college class. Should everyone get the same sized piece? What about those who do not attend class regularly? If those people get the same sized piece of pie, is that fair to the regular attendees? What if a student is allergic to chocolate? That person is being deprived of a good that others are receiving. Should some compensatory treat be provided to that person?^[5]

Efficiency refers to producing the most units of service for the best price, or, as Stone puts it, “getting the most output for a given input.”^[5] Market-oriented policy choices are often advanced as solutions because they may be more efficient. However, efficiency may not always translate into fairness or effectiveness (i.e., policies that have a desired effect on the original problem). Thus the concept of tradeoff is introduced. Emphasizing one value may lead to a diminution of other values. We may have to trade off some effectiveness or equity in return for enhancing efficiency.

Effectiveness would be defined as the extent to which policy choices meet goals and address problems. Effective programs are those that solve problems. However, as noted before, this may produce tradeoffs. We might be able to assure decent housing for all, by giving everyone who has inadequate housing a voucher for, let us say, \$50,000 to purchase adequate housing. This would create a market for housing and, theoretically, address the problem. However, this might be considered to raise issues of equity (What about those who are not so well off who have been able to purchase decent housing while making other sacrifices?) or of efficiency (Is this a cost-effective approach to housing?).

SOCIETAL VALUES

Paris and Reynolds refer to the salience of Societal Overriding Values (SORVs, to use their acronym). These are values held by a large majority of citizens and represent basic societal orientations. As the authors suggest, “For a value to count as an SORV there must be a consensus, or something close to it, on just exactly what counts as satisfying the value.”^[2]

In political scientist E.E. Schattschneider’s view, biases are built into the political system. He has pointed out that “All forms of political organization

have a bias in favor of some kinds of conflict and the suppression of others because organization is the mobilization of bias. Some issues are organized into politics, while others are organized out.”^[7] In a sense, these biases are SORVs, in that they are broadly diffused throughout society and structure discourse (i.e., people use the biases to make decisions without realizing that, in fact, these biases are arbitrary and a construction of the particular society—not based on a priori universal truth).

Any political unit’s SORVs will narrow the range of issues considered appropriate for agenda placement. For instance, in the states of the Deep South before the 1960s, white supremacy values would make it literally impossible for the governor of one of those states to suggest placing on the agenda the issue of integration of racially segregated schools. This issue would be “mobilized out” of consideration for agenda placement, as it went against basic values accepted by whites and their elected representatives. Biases, then, can reduce the scope of debate in the political process radically and prevent certain policy alternatives from ever being examined, no matter their objective merits.

CONCLUSION

Thus values have a role to play in the process of policy analysis. Indeed, it would be amazing if this were not so. The world of the policy analyst is one in which values structure the process and shape outcomes.^[8]

REFERENCES

1. Lasswell, H.D. The policy orientation. In *The Policy Sciences*; Lerner, D., Lasswell, H.D., Eds.; Stanford University Press: Stanford, 1951; 9, 15 pp.
2. Paris, D.C.; Reynolds, J.F. *The Logic of Policy Inquiry*; Longman: New York, 1983; 60 pp.
3. Weimer, D.L.; Vining, A.R. *Policy Analysis*, 3rd Ed.; Prentice-Hall: Upper Saddle River, NJ, 1999; 14, 48–53.
4. Hirschman, A.O. *Exit, Voice, and Loyalty*; Harvard University Press: Cambridge, MA, 1970.
5. Stone, D. *Policy Paradox*; Norton: New York, 2002; 37, 40–41. Revised edition.
6. Peters, B.G. *American Public Policy*, 3rd Ed.; Chatham House: Chatham, NJ, 1993; 373 pp.
7. Schattschneider, E.E. *The Semi-Sovereign People*; Dryden Press: Hinsdale, IL, 1975; 69 pp, reissue edition.
8. Heineman, R.A.; Bluhm, W.T.; Peterson, S.A.; Kearney, E.N. *The World of the Policy Analyst*, 2nd Ed.; Chatham House: Chatham, NJ, 1997.

Virtue Ethics

James S. Bowman

Askew School of Public Administration, Florida State University, Tallahassee, Florida, U.S.A.

INTRODUCTION

Virtue ethics, the dominant outlook in classical Greek philosophy, holds that excellences of character are of central importance in moral life. The analysis of subjective personal traits is ill-suited, however, to subsequent Enlightenment theories that grounded ethics not in individual character but in moral actions subject to universal, rational rules. These approaches concentrate on morally right action as determined by its consequences (in utilitarian, teleological ethics, actions should produce the greatest good for the greatest number) or absolute moral duty (in Kantian, deontological ethics, it is imperative that actions categorically follow principles irrespective of results). Contemporary dissatisfaction with these schools of thought has led to a revival of virtue ethics—those traits that characterize a morally good person. That is, one needs to look for moral norms, not in consequences or duties, but in personal virtues that permit human beings to flourish.

Virtue ethicists claim that consequential and duty ethics are incomplete, because they are derived from something more fundamental: right actions are defined by those done by moral people.^[1] The actual substance of morality cannot be understood by predicting outcomes or adhering to duties. Goodness is the result of internal imperatives to do right, not sanctions from moral rules or rewards from expected consequences. Formulation of a problem can never be a purely technical procedure. People do good not because of reasoning about moral dilemmas, typical of the other perspectives. Instead, they do good as a result of personal conviction, incremental development, behavior patterns, and the everyday choices stemming from them. Sound judgement is based on good character; beliefs, sensitivity, and experience are key in ethical life and should not to be taken for granted, as alternative schools of thought tend to do.

Reason may be essential in carrying out decisions, but behind every action is an individual; before we can do, we must be. The primary faculty is moral sense and intuition, not intellect; reason is easily led into error, and many do not have the capacity or training for analytical logic. Virtue ethicists contend that answers to the question of “What to do?” have little to do with depersonalized calculation of consequences or resolving conflicting duties, and everything to do

with what kind of person one is. It is more important to “know thyself” than to know anticipated results or inflexible rules. Thus, the philosophy is rooted in the reality of living and how decisions are made; indeed, our everyday moral judgements often are judgements about someone’s virtues and vices.

If a professional is defined as having technical competencies (what to do) and moral competencies (whom to be), then consequential and duty theories build on the former, and virtue ethics the latter.^[2] Virtue requires an individual not simply to know how to predict results or apply duties, but also why to do so; ethical questions are more than simply decision puzzles. This is because to be a human being is to possess the potential to be excellent, excellence that it is one’s responsibility to develop. Indeed, professionals may be required to perform supererogatory acts, or do more than is normally required, demanded, or expected (e.g., a mental health worker has a stronger duty to provide help to someone at risk than a member of the public). It is evident, then, that ethics is not an imposition or constraint in professional life, but the foundation of everything a professional is and does. As such, virtue theory may even help resolve the “dirty hands” dilemma^[3] encountered by public officials who may be conflicted between utilitarian concerns for the public interest and the duty to adhere to laws.

TERMINOLOGY

Critical concepts—character, virtue, integrity, practical wisdom—need examination to comprehend the range of virtue ethics. Character is a person’s pattern of behavior related to the moral qualities of individual self-discipline and collective social responsibility; one cannot be understood apart from the larger community in which one participates. Good character is not an accident; it is a lifelong process of self-discovery requiring commitment to ethical values. Character is one’s moral center or “naked soul.” One test of character is what one would do if one were invisible. Does the person, in modern vernacular, have the “right stuff” to do the right thing? Good people make moral choices. The reply to “What is a good life for a human being?” is to create a moral character, to achieve excellence in all aspects of life, to live virtuously with others,

in brief, to become a “virtuoso.” In reciprocal fashion, virtuous citizens build moral communities that, in turn, encourage further character formation.

Virtue is a trait or excellence of individual and societal character. Virtues are principles of goodness and rightness in character and conduct (e.g., justice, courage, temperance) that lead toward moral nobility and away from depravity. Woven into character, they are ingrained dispositions to act rightly, developed by practice, just as vices are tendencies to act wrongly. Not easily honed, virtues make people sensitive to ethical issues and shape how they think and act. One becomes courageous, for example, by repeatedly behaving in a courageous manner. Virtue is found between the vicious extremes of excess and deficiency—Aristotle’s “golden mean.” In every situation, a person of character will select actions based on reason and experience that are appropriate—neither excessive nor defective—for that circumstance. The virtue of mildness, for instance, rules that one may get angry, but only in certain proportions warranted by occasion. What is prudent in one situation may not be in another. Such individuals are guided by their principles, not popularity or expedience. Virtues are more than habitual skills enabling right action; they promote the attainment of what it means to be human, to function well, to be moral.

Integrity, a product or synthesis of virtues, is a predisposition to maintain one’s virtues and act in character. In today’s parlance, this is consistency between belief and action, to “walk the talk,” to “have it all together.” It is the aptitude to contemplate one’s roles and obligations and nurture a moral soundness as a person as well as a member of the community. It grows from internal values, experiences, and the ability to act thoughtfully; it is one’s “moral identity.”

To act thoughtfully requires practical wisdom (judgment, prudence), the capacity to distinguish right from wrong and take sensible action. The reflective person recognizes the distinction between knowing and doing. Moral virtue directs one to the right end; practical wisdom directs one to the right means. Practical wisdom enables one to discern the golden mean in particular situations, a process requiring experience that cannot be reduced to rules.

Essential for an ethical life, practical wisdom engages and unifies the intellectual and the moral: right action requires intellectual judgment, which requires moral virtue. It has no standard external to itself; to act rightly is to do what a good person would do in a particular circumstance. One deliberates as long as necessary and then acts decisively with compassion, precision, and speed. One must see a duty and do it; in fact virtuous action, founded upon arduous practice, may be nearly spontaneous. To summarize briefly, character (personal behavior patterns) is enriched by

virtues (disposition to act); a synthesis of them constitutes integrity, something that can be achieved through the exercise of practical wisdom, which is known as the preeminent virtue.

VIRTUE THEORY PRINCIPLES

Over thousands of years, diverse cultures have constructed various accounts of life to serve as moral guidelines (often included are hope, love, honesty, prudence, temperance, justice, courage, faithfulness, loyalty, wisdom). Compiling such lists misses an important point made by virtue ethicists: to avoid a formulaic approach to ethics by emphasizing its context and personal nature. This means that the content of a norm cannot be defined in general principles but by mediating upon the lives of admirable people.

Virtue ethics, then, is a way of life more than a method of analysis. That is, the mere possession of virtues is not enough, because character is more than the result of individual traits added together; instead, character is transformative, as one must also have practical wisdom to know how and when to exhibit individual virtues.^a Thus, Aristotle,^[5] who believed that virtues are learned concurrently, did not produce a definitive list but identified assorted virtues on the grounds that they were the mean between extremes; the essence of virtue is between the vices of excess and deficiency (e.g., honesty is between bluntness and deceit). Such counsel encourages decision makers from veering too sharply in one direction or another, thereby violating fundamental values; balance is a moral ideal when important values are pitted against each other.

For students of administration and policy, some standards have appeared in the professional literature.^[6] Bailey’s^[7] mental attitudes (recognition of moral ambiguity, contextual forces, and paradox of procedures) and moral qualities (optimism, courage, fairness) are useful to officials as ethical custodians of public organizations and democracy (Table 1, top panel). Denhardt^[8] recommended honor, benevolence, and justice (Table 1, bottom panel). Hart^[9] saw public administration as a moral endeavor demanding “superior prudence” (the power to transcend self-interest and seek the public good), which includes moral heroism, care for humankind, trust in the citizenry, and continuous improvement of honorable character.

Whether long or short, a list will have coherence when the habits and dispositions to achieve the good life are mastered. Virtue theory contends that the whole of

^aBen Franklin describes in his autobiography how he nurtured some dozen virtues he thought important to a meaningful life. A modern-day approach to such habituation is found in Ref. ^[4].

Table 1 Selected administrative virtues*Bailey*

Mental attitudes—recognition of:

- Moral ambiguity of people and policy (dirty hands).
- Contextual forces that condition priorities in public service (lie to a terrorist).
- Paradoxes of procedures (friend of deliberation, fair, law but may be enemy of progress and speech).

Moral qualities:

- Optimism (incentive to improve quality of life).
- Courage (needed to face ambiguities, duty to act).
- Fairness tempered with charity (compensates for inadequate information).

Denhardt

Honor: magnanimity (great mindedness), honesty, acting honorably (high standards).

Benevolence (promote welfare of others): sympathy, enthusiasm, commitment to serve.

Justice (respect for and commitment to the dignity and individual worth of each person): includes promoting informed participation.

Source: From Refs.^[7,8]

character is greater than the sum of its parts; a compilation of virtues is not as important as blending a set of desirable qualities to form an ethical character. No easy task, it is hardly surprising that some prominent leaders demonstrate not only moral strengths but also weaknesses. The leadership challenge is to deal thoughtfully with clashes between virtues; such conflicts will then yield integrity and personal growth. As a result, ethical managers are likely to be prepared to deal with unexpected problems, recognize dilemmas when they arise, and justify decisions on proper grounds.

ADVANTAGES AND DISADVANTAGES

For virtue ethicists, other approaches do not provide an adequate account of moral existence, because they focus on intellectualistic efforts to resolve ethical problems after individual characters are presumably set. While these theories maintain that they incorporate aspects of virtue ethics, virtue ethics presents a different conception of morality as well as what is important in moral evaluation by emphasizing personal development and reflective judgment. This philosophy contains advantages and drawbacks.

The chief advantage is its claimed realism by attending to the actual dynamics of life in several respects. First, by valuing being rather than doing, it emphasizes everyday experiences, not dramatic cases. People do not decide what to do by simply relying on abstract rules or estimating consequences. Instead, they learn

ethics and sound judgment by living ethically, more than they do by coping with extraordinary incidents. The sensitivity of a virtuous person is central. Knowing the distinction between right and wrong is necessary, but not sufficient; caring about the difference is also required.

Consequential and duty ethics, in contrast, are cold, calculating, and callous, because they leave out an essential part of reality: human feelings. A theory of ethics must deal with the whole person. Virtual ethics centers on the question, “What is a moral life?,” whereas other approaches ask, “What is the right thing to do in a situation?” In so doing, the latter represents a minimalist approach, because the instrumental focus makes minimal demands on people (“ethics without virtue”); the former is a maximal approach, because it demands deep reflection and high ideals.^[10]

Second, virtue ethics maintains that people become moral from what they demand of themselves in their characters (e.g., the reason to be considerate of others is not the impact on them but the effect on one’s self). Indeed, cultivating virtue helps them from being entrapped by moral dilemmas in the first place. The disposition to do the right thing is a prerequisite to, and more vital than, merely knowing what is right.^b Proper action results from thoughtful judgment informed by virtuous concern. Virtues are a kind of second nature that lead one to do the correct thing and provide pleasure in so doing. To consider character as static and unimportant, as the other schools tend to do, misses the essence of ethics.

Weaknesses in virtue ethics include a key limitation of, and several flaws in, the theory. The limitation is that it, by definition, lacks a theory of action; right conduct is not codified into general guidelines. While alternative theories compel decisions (although they may not provide clear guidance in specific cases), virtue philosophy says little about what should be done. Virtues may generate instructions for action (e.g., justice provides motivation to act justly), but what, exactly, does a just person do in problematic circumstances? Many situations cannot be reduced to character

^bThis distinction between knowing and doing can be illustrated by former Texas Governor Ann Richards, past Senator George Mitchell, and one-time Reagan White House Chief of Staff former Senator Howard Baker. In 1997, the tobacco industry hired them to lobby Congress. Each widely-known for integrity, Richards (a recovering alcoholic, knew well the problem of addiction), Mitchell (a peace negotiator, could work for life in Ireland and death in America), and Baker (whose wife died of lung cancer) earned huge sums of money for their efforts. In comparison, Reagan’s Surgeon General, C. Everett Koop, upon leaving office in 1989 refused to endorse breakfast cereals and condoms—actions that were at least consistent with good health. Not wanting to cheapen his office by selling his name, he turned down millions of dollars.

As one moves out from the triangle, the ethical basis for action is weakened and unethical actions may seem justifiable. In addition, reinforcement from other bases is decreased. As one moves in toward the triangle, the reinforcement among bases is increased.

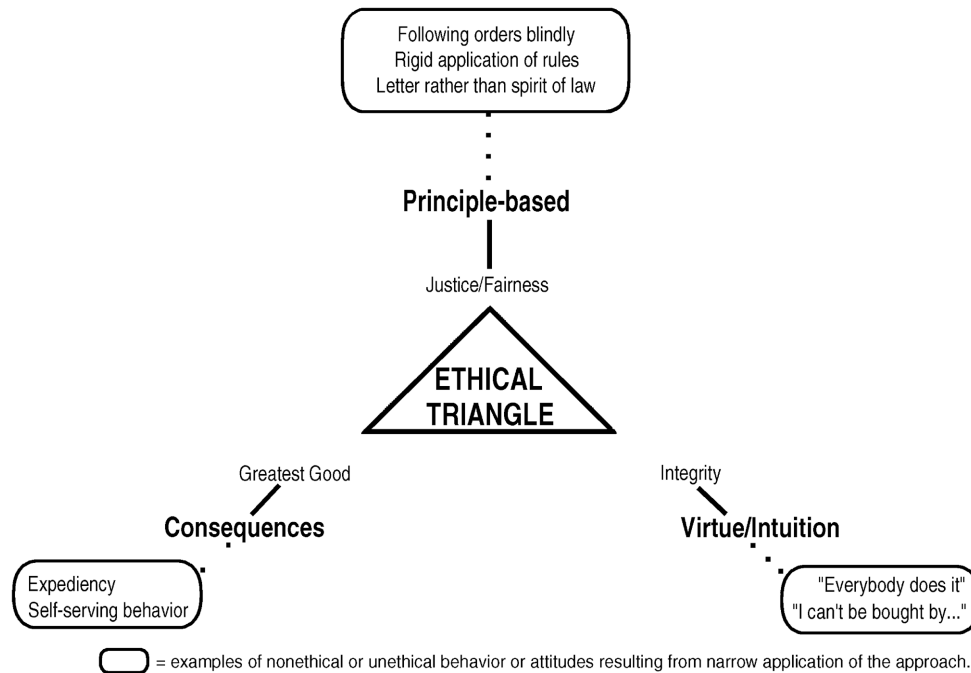


Fig. 1 Ethical triangle. From Ref.^[13].

virtues, traits which, paradoxically, are often defined as how a good person would act. A further complication is when someone “acts out of character” in a particular circumstance.

Relatedly, the initial theoretical flaw, ironically, is that virtue ethics lacks integrity—one may be good but may not know how to do good. Confusing being with doing can lead to self-righteousness; if one thinks he is good, then it is not hard to believe that what one does is good.^c Too, when confronted with conflicts among virtues, the only way to resolve them may be through the use of other moral philosophies.

Another flaw is that virtue ethics lacks context, as it is only in the *polis* that virtues can be exercised and the good life achieved. “To lead a good life and to be a good person,” according to the Greek saying, “one must live in a great city.” But without appropriate opportunities in the community, humans may be incapable of virtue. Thus, the infamous Milgram^[11] electric shock experiments demonstrated that significant numbers of people are unable to act morally in situations calling for obedience to authority [also see the notorious 1971 Zimbardo “Stanford prison experiments” (<http://www.prisonexp.org>)].

^cDuring the US Department of Housing and Urban Development (HUD) scandals in the 1980s, for example, agency official Marilyn Harrell received notoriety as “Robin HUD.” She testified in 1989 that she diverted public funds to the poor—and to herself.

Finally, the approach lacks universality, as there is no general theory of human virtue. Virtue seems to vary from time to time and place to place, and the same virtues may not apply to all people (e.g., men versus women, young versus old). Personal characteristics seem to invite an arbitrariness into moral considerations that the dominant moral philosophies of the last 200 years have attempted to avoid. If feelings are central, then ethical life is built on little more than the subjectivity, human emotions, and inconsistencies of daily living.

REINTEGRATION OF CHARACTER AND ACTION

Modern ethical systems distinguish between an individual’s character and an individual’s action, with action being more morally salient, because it affects others. By reviving the classical view that a person’s life—to be of general moral significance—must be one based on character, virtue ethicists have stimulated a reexamination of morality, the nature of ethical problems, and the extent to which they admit to solution. Accordingly, they argue for the reintegration of character and action, because personal virtue can be an important determinant of action.^[12]

Because ethical practices must deal with the entire person, management professionals are obligated to develop virtues, predict consequences, and adhere to rules—to be conversant with virtue, consequential, and duty ethics. The complementarity and interdependence

of these three theoretical perspectives is illustrated by Svvara's^[13] "ethical triangle" (Fig. 1). Each school, represented in the triangle, provides a way to reframe or clarify what is at stake in a problem. As one moves out from each point, the justification for action—and reinforcement from the other angles—weakens. Conversely, operating within the triangle mitigates against the key shortcoming of virtue (self-justification), consequential (expediency), and duty (rigid application) ethics.

The three angles often inform and supplement one another in a synergistic manner. When they do not, at least all perspectives will have been considered, and a reasoned choice—based on benevolence, the public interest, and moral imagination—can be forged and defended. As such, this integrative, robust geometric tool promotes an ethic of compromise and social integration, a key tenet of democracy. It is a parsimonious analytic device that helps professionals from straying too far from virtues, outcomes, and duty. The revival of virtue ethics, accordingly, has encouraged a scrutiny of moral theory and what it might reasonably accomplish, and has painted a more complete portrait of morality in public affairs.

REFERENCES

1. Cooper, T., Wright, N., Eds.; *Exemplary Public Administrators*; Jossey-Bass: San Francisco, 1992; 352 pp.
2. Bowman, J. The lost world of public administration education. *J. Public Adm. Educ.* **1998**, *4* (1), 27–31.
3. Waltzer, M. Political action and the problem of dirty hands. *Philos. Public Aff.* **1973**, *2* (1), 160–180.
4. Covey, S. *Principle-Centered Leadership*; Summit Books: New York, 1991; 332 pp.
5. Aristotle nicomachean ethics. In *Complete Works of Aristotle*; Ross, W., Ed.; Princeton University Press: Princeton, NJ, 1984. Trans; rev. by Urmson, J.; ed. and rev. by Barnes, J.; 2,487 pp.
6. Luke, J.; Hart, D. Character and conduct in the public service. In *Handbook of Administrative Ethics*; Cooper, T., Ed.; Marcel Dekker: New York, 2000; 529–553.
7. Bailey, S. Ethics and public service. *Public Adm. Rev.* **1964**, *24* (1), 234–243.
8. Denhardt, K. Unearthing the moral foundations of public administration. In *Ethical Frontiers of Public Management*; Bowman, J., Ed.; Jossey-Bass: San Francisco, 1991; 91–113.
9. Hart, D. The virtuous citizen, the honorable bureaucrat, and 'public' administration. *Public Adm. Rev.* **1984**, *44*, 111–119. (Special Issue).
10. Norton, D. 'Character ethic' and organizational life. In *Papers on the Ethics of Administration*; Wright, N., Ed.; Brigham Young University: Provo, UT, 1988; 47–66.
11. Milgram, S. *Obedience to Authority*; Harper's: New York, 1974; 224 pp.
12. MacIntyre, A. *After Virtue*; University of Notre Dame Press: Notre Dame, IN, 1984; 286 pp.
13. Svvara, J. The ethical triangle. *Public Integr. Annu.* **1997**, *2*, 33–41.

Watergate

Carol W. Lewis

Department of Political Science, University of Connecticut, Storrs, Connecticut, U.S.A.

INTRODUCTION

On June 17, 1972, a security guard at Watergate, a Washington hotel and apartment complex along the banks of the Potomac River, notified police when he saw that a stairwell door had been taped open. At 2:30 a.m. in the Democratic National Committee's headquarters, the three responding officers arrested five intruders, who were there to adjust bugging equipment installed in May and to photograph documents. This incident, which the president's press secretary (Ron Ziegler) labeled a "third-rate burglary,"^[1] was the public beginning of a series of events collectively termed *Watergate*. Its full story unfolded slowly,^[2] in bits and pieces, and ultimately developed into protracted political discord that "captured headlines... permeated the national consciousness and evolved into an ethical and constitutional landmark."^[3]

Within a few days, a relationship was established between the burglars and a former White House aide (E. Howard Hunt) and the general counsel for the president's re-election committee (G. Gordon Liddy). Seven indictments came down from a federal grand jury within several months, while the president repeated his denial of knowledge about and involvement in the incident.^a

BACKGROUND

In 1968, after serving as vice president under President Dwight D. Eisenhower and losing the presidency to John Fitzgerald Kennedy in 1960, Richard Milhous Nixon defeated Hubert Humphrey in one of the closest elections in U.S. history. When sworn in as the 37th president on January 20, 1969, he took the oath stipulated in the eighth clause of Article II, Section 1 of the U.S. Constitution. "I do solemnly swear (or affirm) that I will faithfully execute the Office of President of the United States, and will to the best of my Ability, preserve, protect and defend the Constitution of the United States." The fact that Nixon was re-elected

on November 2, 1972, less than five months after the Watergate break-in, in a landslide victory, adds a poignant and perplexing note to the affair that ultimately led to the first resignation of a president in U.S. history. (Nixon won re-election by more than 60% of the vote, carried 49 of the 50 states, and defeated the Democratic candidate, Senator George McGovern of South Dakota.)

An explanation (but by no means an excuse) for the wrongdoing by the president and advocates of his re-election, followed by deceit, denials, and illegalities, lays at least in part in the context surrounding this White House. Operating in a political world broader than electoral calculations, this was an administration besieged by the myths and realities of deep discord associated with the war in Vietnam, the Civil Rights Movement, political assassinations of the 1960s, and the decline in confidence and trust in government already evident in national surveys.^b Indeed, this was an era of domestic turmoil and social change.^c Also, and especially important in this context, a Republican president saw himself and his policies repeatedly thwarted by a Democratic majority in Congress and an unresponsive bureaucracy. Moreover, Nixon is said to have been "a man obsessed" by his 1960 defeat and by President Kennedy and his legacy.^[8]

The links between these factors and Watergate—and the mentality prevailing in the White House—are evident in the events surrounding the Pentagon Papers. In June, 1971, a year before Watergate, first *The New York Times* and later *The Washington Post* begin publishing the Pentagon Papers, the Defense Department's secret history of the Vietnam War leaked by Daniel Ellsberg, a former defense analyst. The White House plumbers, a staff group so named for their job to plug leaks in the administration, burglarized a psychiatrist's office to find materials on Ellsberg.

Another factor came into play as Watergate unfolded and would serve to further undermine the administration's credibility. Vice President Spiro T. Agnew took office in 1969 but resigned late in 1973,

^aA detailed chronology is available at <http://www.washingtonpost.com/wp-srv/national/longterm/watergate/chronology.htm>. A detailed 'cast of characters' is available in Ref.^[4].

^bOn public confidence and trust in government, see the National Election Studies, University of Michigan, formerly at http://www.umich.edu/-nes/nesguide/toptables/tab5a_5.htm and Ref. ^[5] and at <http://www.people-press.org>.

^cOn the political and social context, see Ref.^[6]. See also Ref.^[7].

after pleading no contest for evading income taxes on kickbacks he received when governor of Maryland.^[9] Under the authority of the 25th Amendment to the U.S. Constitution, President Nixon appointed U.S. Representative Gerald Rudolph Ford. Elected by voters in a single congressional district in Michigan, albeit 12 times, Ford would become the 38th president upon Nixon's resignation nine months after he was sworn in on December 6, 1973.^d

COVER-UP AND DISCLOSURE

The links to the Oval Office got stronger and stronger as information trickled out or was ferreted out. Initially at least, the protagonists were not talking, and links with the White House and the president's campaign were denied repeatedly. The cover-up could be dated as early as two days after the break-in because, on June 19, 1972, former Attorney General John Mitchell, the head of Nixon's re-election campaign, denied any association with the affair.

Watchdog Press

The cover-up was undone by a free media and by hearings held by a Congress dominated by Democrats.^e The press undeniably fulfilled its promise in a democratic polity, despite threats and intense pressure.^[1] The *Washington Post* and its two investigative reporters, Carl Bernstein and Bob Woodward, played a major role in following the trails of abuses and disclosing the cover-up. CBS News and the *Los Angeles Times* also chased the story (and the *New York Times* first published the Pentagon Papers). Woodward and Bernstein's work would win the *Washington Post* a Pulitzer Prize for public service and result in the book and film, *All the President's Men*.^[11]

The first paragraph of the first big story by Woodward and Bernstein was published on August 1, 1972. It declared, "A \$25,000 cashier's check, apparently earmarked for President Nixon's re-election campaign, was deposited in April in a bank account of one of the five men arrested in the break-in at Democratic National Headquarters here June 17."^[12] The burglars were connected to the president's re-election committee. Their article, published under the headline "FBI Finds Nixon Aides Sabotaged

Democrats" on October 10, began, "FBI agents have established that the Watergate bugging incident stemmed from a massive campaign of political spying and sabotage conducted on behalf of President Nixon's re-election and directed by officials of the White House and the Committee for the Re-election of the President."^[13] These and other stories about the break-in apparently had little or no effect on the election, and Nixon won his landslide victory.

Whitewash at the White House

Approximately five months later, President Nixon offered the strongest case he would make in his own defense. On April 30, 1973, he reaffirmed his innocence in a nationally televised speech. In what would prove to be a useless attempt to offer up scapegoats and deflect attention from his own role, he announced the resignation of his two senior White House aides (H. R. Haldeman and John Ehrlichman). He described these sacrificial lambs as "two of the finest public servants it has been my privilege to know." The resignation of the attorney general (Richard Kleindienst) was announced, along with the dismissal of the president's special counsel (John W. Dean III), whose revelations subsequently implicated the president in the cover-up.

The president characterized his newly appointed Attorney General, Elliot Richardson, as "a man of unimpeachable integrity" and gave him "absolute authority to make all decisions bearing upon the prosecution of the Watergate case and related matters" and to name a special prosecutor. Of Richardson, Nixon remarked, "I am confident that with him in charge, justice will be done." (His words would be remembered and fuel the fires ignited by the so-called Saturday Night Massacre, discussed below.)

In his speech, President Nixon also affirmed that "in matters as sensitive as guarding the integrity of our democratic process, it is essential not only that rigorous legal and ethical standards be observed but also that the public, you, have total confidence that they are both being observed and enforced by those in authority and particularly by the president of the United States."

In this speech, the president "made promises that would prove to be both important and empty."^[3]

Whatever may appear to have been the case before, whatever improper activities may yet be discovered in connection this whole affair, I want the American people, I want you to know beyond the shadow of a doubt that during my term as president, justice will be pursued fairly, fully, and impartially, no matter who is involved. This office is a sacred trust and I am determined to be worthy of that trust . . . We must

^d"Gerald Ford served in the House of Representatives from January 3, 1949 to December 6, 1973, being reelected twelve times, each time with more than 60% of the vote." Gerald R. Ford Biography at <http://www.ford.utexas.edu>, accessed September 2001.

^eA detailed chronology is available at <http://www.washingtonpost.com/wp-srv/national/longterm/watergate/front.htm> and, for a point-by-point rendition, see Ref. ^[10].

maintain the integrity of the White House, and that integrity must be real, not transparent. There can be no whitewash at the White House.

Over time, it would become clear that this speech was an instance of outright deceit. It is to this deceit that Theodore H. White, a best-selling chronicler of presidential politics, attributed Nixon's downfall. In his book entitled *Breach of Faith*, White wrote that a government official "must know when to choose trust and faith over political need. If exposure of his acts threatens to contradict his words, he must renounce his acts and keep his word, because the people must trust his words at whatever cost—or he cannot govern."^[7]

Saturday Night Massacre

The president's "assurances explain why the whirlwind touched down on October 20, 1973, the date of the [so-called] Saturday Night Massacre."^[7] The controversy over releasing taped presidential conversations—the authority of Congress to issue subpoenas versus claims of executive privilege—provoked President Nixon to order Attorney General Richardson to fire Archibald Cox, the special prosecutor Richardson had appointed subsequent to Nixon's speech the previous April. (Cox was a former solicitor general.) Rather than comply, Richardson and his deputy (William D. Ruckelshaus) resigned. Solicitor General Robert Bork became acting attorney general and fired Cox. "Seeing discharging the special prosecutor as a legal matter, not a moral one, Bork explained that he was trying to maintain the viability and integrity of the Department of Justice and, indeed, the executive branch of government."^[14] A Federal District Court ruled in 1973 that the firing was illegal. President Ronald Reagan nominated Bork, by that time a federal judge, to the Supreme Court, but Bork was not confirmed.

Political analyses often draw upon historical precedents, which are useful in understanding the role of the special prosecutor. President Ulysses S. Grant, whose administration was disreputable at best, was the first president to appoint and the first to fire a special prosecutor. President Harry Truman's attorney general did likewise in 1952. Nonetheless, the political controversy escalated in response to the Saturday Night Massacre, as did public pressures and congressional resolve to investigate. Assurance had degraded into deceit and, for some, a sense of having been betrayed.

Congressional Investigation and the Tapes

The Senate Select Committee began its Watergate hearings—televised nationally—in May, 1973, under

the chairmanship of Senator Sam Ervin (Democrat, North Carolina). In early June, the public learned that a former counsel to the president (John Dean) had informed Watergate investigators that he had discussed the cover-up with the president almost three dozen times. In his congressional testimony less than two months later, the president's former appointments secretary (Alexander Butterfield) disclosed the existence of the White House's tape-recording system. Quickly, attention shifted to the tapes. On July 23, 1973, the president refused to release tapes to the Senate committee or special prosecutor. A battle would rage over the tapes in the press and in court for another year. In December, 1973, the disclosure of a gap of 18 1/2 minutes in one of the subpoenaed tapes would further undermine White House credibility and cement suspicion in some quarters.

IMPEACHMENT

Impeachment proceedings were initiated in the House Judiciary Committee, chaired by New Jersey Democrat Peter Rodino (with a quarter-century of House experience). Article I, Section 2 of the U.S. Constitution gives the House of Representatives "sole power of impeachment" (which amounts to an indictment). According to House procedure, the Judiciary Committee takes up the matter first and then makes its recommendation to the House. Congress conducted investigations of presidents as early as President James Monroe in the early nineteenth century.^[14]

At the end of April, 1974, the White House released more than 1200 pages of transcripts to the Judiciary Committee. The committee objected to the edited transcripts and demanded the tapes. On July 24, 1974, the Supreme Court, presided over by Chief Justice Warren Burger, rejected the claim of executive privilege and unanimously ruled that the president must deliver recordings of 64 conversations. Among them was the so-called "smoking gun," direct evidence of a specific illegal act that meets the stipulation in the Constitution (Article II, Section 4) of "treason, bribery or other high crimes and misdemeanors" for impeachment and conviction. While a majority of the members of the Judiciary Committee were persuaded by a pattern of illegal and abusive behavior, others objected that there was no single piece of evidence linking the president directly to impeachable action; still others argued that the proceedings were partisan in motive and less than objective and fair.

The link was established and an instance of obstruction of justice made certain by the tape of a conversation on June 23, 1972, between the president and his advisor (Haldeman) about using the CIA to put a stop to the FBI's investigation of the break-in. This

tape (and others released in later years) confirmed that Nixon knew about the cover-up from the start. The question repeatedly posed by Senator Howard H. Baker, Jr. (Republican, Tennessee), vice chairman of the Senate investigating committee—"What did the president know and when did he know it?"—was answered.^f

On July 27, 1974, after seven months of inquiry, in a roll call taking 3 minutes, the House Judiciary Committee approved the first of three articles of impeachment. Rooted firmly in the president's constitutional oath, it read in part:

In his conduct of the office of President of the United States, Richard M. Nixon, in violation of his constitutional oath faithfully to execute the office of President of the United States and, to the best of his ability, preserve, protect, and defend the Constitution of the United States, and in violation of his constitutional duty to take care that the laws be faithfully executed, has prevented, obstructed, and impeded the administration of justice.

The first article, charging obstruction of justice, was passed 27 to 11; six Republicans joined 21 Democrats in this historic decision.^[15] Note, however, that the vote preceded disclosure of the critical tape, the smoking gun. Chairman Rodino had predicted, "This is a turning point, whatever we decide," and he was correct. Impeachment precedents number about 50, but only about a dozen reached trial and only a single case had involved a president.^g The committee subsequently passed another article alleging abuse of power and a third alleging contempt of Congress. This committee would report for impeachment to the House.

RESIGNATION AND PARDON

The president now came under intense pressure from many quarters, including longtime political associates, Republican congressional leaders, and close advisors, to spare the country the prolonged, preoccupying, and now apparently fruitless drama of impeachment in the House and then a trial in the Senate. On August 8, 1974, the 37th president announced his resignation. (See Fig. 1.) He was the first U.S. president to do so. The White House biography of Nixon explains, "Faced with what seemed almost certain impeachment,

Nixon announced on August 8, 1974, that he would resign the next day to begin 'that process of healing which is so desperately needed in America.'"^[16]

According to one description of the events surrounding the resignation:

The 16 minute speech was delivered with remarkable restraint, given the circumstances, and without a trace of demagoguery or self-pity. There were no attacks on his old enemies, no visible bitterness. There was also no concession of anything more serious than "mistakes" in his handling of Watergate, and no hint of remorse except one line regretting "any injuries that may have been done in the course of the events that led to this decision."^[17]

Nixon stated that he was leaving office the following day, because his "political base in the Congress" was gone. His letter of resignation, dated August 9, 1974, was addressed to the Secretary of State, Henry A. Kissinger, and read simply, "I hereby resign the Office of President of the United States."^[18]

Gerald Ford became the 38th president and the first to assume the presidency by virtue of the 25th amendment. By signing Proclamation 4311 on September 8, he exercised his constitutional authority (Article II, Section 2) to grant pardons. The authority to grant pardons does not extend to cases of impeachment. President Ford granted Nixon a "full, free, and absolute pardon."^h

Among the reasons Ford gave as the basis for his decision to grant a pardon were his understanding of his constitutional duties; his conscience; his religious beliefs; and values such as mercy and justice. He also stressed the public good that was otherwise threatened by political polarization, perhaps even paralysis, along with the reduced credibility of governmental institutions, were Nixon subjected to a long criminal trial. (See Fig. 2.) Speaking to the House Judiciary Committee on October 17, Ford further explained that "a government which deceives its supporters and treats its opponents as enemies must never, never be tolerated."

Shortly after pardoning Nixon, in an act of reconciliation that supported his explanations, President Ford proclaimed an amnesty for those who had evaded military service during the war in Vietnam. Ford's standing in opinion polls suffered from the pardon and the amnesty, and he lost his bid for re-election.

^fIn a move heavy with symbolism, President Reagan appointed Senator Baker as his chief of staff in response to the Tower Commission's investigation of the Iran-Contra affair.

^gThe Senate convicted neither President Andrew Johnson in 1868 nor President William Jefferson Clinton. To date, no U.S. president has been convicted.

^hThe complete text of President Ford's statement is available at the Gerald Ford Library at <http://www.ford.utexas.edu>, accessed August 2001. President Ford received the John F. Kennedy Profile in Courage Award in 2001 for pardoning Nixon. See Ref. ^[19].

In all of the decisions I have made in my public life, I have always tried to do what was best for the Nation. Throughout the long and difficult period of Watergate, I have felt it was my duty to preserve, to make every possible effort to complete the term of office to which you elected me

In the past few days, however, it has become evident to me that I no longer have a strong enough political base in the Congress to justify continuing that effort....

I would have preferred to carry through to the finish whatever the personal agony it would have involved, and my family unanimously urged me to do so. But the interests of the Nation must always come before any personal considerations....

I have never been a quitter. To leave office before my term is completed is abhorrent to every instinct in my body. But as President, I must put the interest of America first. America needs a full-time President and a full-time Congress.... To continue to fight through the months ahead for my personal vindication would almost totally absorb the time and attention of both the President and Congress in a period when our entire focus should be on the great issues of peace abroad and prosperity without inflation at home

Therefore, I shall resign the presidency effective at noon tomorrow. Vice President Ford will be sworn in as President at that hour in this office

As we look to the future, the first essential is to begin healing the wounds of this Nation; to put the bitterness and divisions of the recent past behind us and to rediscover those shared ideals that lie at the heart of our strength and unity as a great and as a free people

By taking this action, I hope that I will have hastened the start of that process of healing which is so desperately needed in America.

I regret deeply any injuries that may have been done in the course of the events that led to this decision. I would say only that if some of my judgments were wrong, and some were wrong, they were made in what I believed at the time to be the best interest of the nation....

Fig. 1 Excerpts from President Richard M. Nixon's resignation speech, August 8, 1974.

IMPACT

Watergate's most notable impact was the president's resignation. About 40 government officials eventually were indicted in proceedings related to Watergate; some served prison terms and others were dismissed, fined, or disbarred. The affair's successive revelations, duration, and aftermath profoundly affected the federal government beyond the White House:

While White House staff, senior political appointees, and members of a political campaign committee were heavily implicated and ultimately prosecuted, the federal bureaucracy was deeply affected as well. The Internal Revenue Service, Federal Bureau of Investigation, Central Intelligence Agency, and the Department of Justice suffered the impact of manipulated operations, suspicion, and illegal obstruction of justice and cover-up. Allegations and investigations riveted attention and absorbed administrative resources. Executive decisions were hard to come by and much of the non-routine business slowed to a standstill.^[3]

While elected and appointed officials were involved, not civil servants, neither the U.S. Civil Service

Commission nor human resources managers protected the career civil service.^[20]

Public opinion registered a considerable decline in confidence and trust in government. The affair also expanded the "popular conception of what constituted governmental corruption" to include waste and abuse, as well as more traditional focus on fraud.^[20] Measures designed to ameliorate these impacts were taken. "Largely driven by the revelations of Watergate, lawmakers in the 1970s passed a series of 'good government' bills designed to repair the political process and rejuvenate public confidence in their elected officials."^[21] In the aftermath of Watergate, 19 states adopted codes of conduct. The American Bar Association adopted an accreditation standard for instruction in legal ethics in 1974.^[9] The 1974 elections swept 92 new members into the House of Representatives and 11 new members to the Senate. Both chambers adopted codes of conduct in 1977 that ended office "slush funds," restricted outside income, limited lobbying by former members and staffers, and more. In 1978, the U.S. Office of Government Ethics was established; the 1978 Ethics in Government Law established what later became the controversial office of the independent counsel.^[22]

I have come to a decision which I felt I should tell you and all of my fellow American citizens, as soon as I was certain in my own mind and in my own conscience that is the right thing to do....

My customary policy is to try and get all the facts and to consider the opinions of my countrymen and to take counsel with my most valued friends. But these seldom agree, and in the end, the decision is mine....

I have promised to uphold the Constitution, to do what is right as God gives me to see the right, and to do the very best that I can for America....

The Constitution is the supreme law of our land and it governs our actions as citizens. Only the laws of God, which govern our consciences, are superior to it....

Theirs [Nixon and his family] is an American tragedy in which we all have played a part. It could go on and on and on, or someone must write the end to it. I have concluded that only I can do that, and if I can, I must....

I deeply believe in equal for all Americans.... The facts, as I see them, are that a former President of the United States, instead of enjoying equal treatment with any other citizen accused of violating the law, would be cruelly and excessively penalized....

During this long period of delay and potential litigation, ugly passions would again be aroused. And our people would again be polarized in their opinions. And the credibility of our free institutions of government would again be challenged at home and abroad....

As President, my primary concern must always be the greatest good of all the people of the United States whose servant I am. As a man, my first consideration is to be true to my own convictions and my own conscience....

My conscience tells me it is my duty, not merely to proclaim domestic tranquility but to use every means that I have to insure it. I do believe that the buck stops here, that I cannot rely upon public opinion polls to tell me what is right. I do believe that right makes might and that if I am wrong, 10 angels swearing I was right would make no difference....

Now, therefore, I, Gerald R. Ford, President of the United States, pursuant to the pardon power conferred upon me by Article II, Section 2, of the Constitution, have granted and by these presents do grant a full, free, and absolute pardon unto Richard Nixon for all offenses against the United States which he, Richard Nixon, has committed or taken part in during the period from July (January) 20, 1969, through August 9, 1974....

Fig. 2 Excerpts from President Ford's explanation to the public of his pardon of Richard Nixon, September 8, 1974.

LEGACY

President Nixon asked in his televised speech on April 30, 1973, "How could it have happened? Who is to blame?" An answer is needed to begin to understand the cause(s) or at least the seeds of Watergate and to take measure of its legacy. A good part of the answer lays in the organizational culture induced or condoned by the president, his appointees, and his staff, to which they fell prey. An historian likened the culture to a "cult of personality" in which "loyalty to the leader seemed to override everything else."^[23] They were inadequately prepared for their stature and task, too easily tempted by power, and then overwhelmed by it.^[9]

At the request of a Senate committee, a panel of the National Academy of Public Administration produced a report in which was described a climate of zealotry

characterized by ideological commitment to a particular president and his policy rather than by judicious commitment to the office.^[24] In his April, 1973, speech, Nixon expressed a similar view by describing subordinates as "people whose zeal exceeded their judgment and who may have done wrong in a cause they deeply believed to be right." "It appears that conviction preempted judgment and loyalty trumped competing ethical claims in this White House."^[3]

The *Washington Post*, commemorating in June, 1997, the 25th anniversary of the burglary, described Watergate as the "most famous political scandal in American history." While that may prove a temporary designation, it surely left a lasting mark on the American polity and society. "Watergate continues to reverberate through politics and journalism as the scandal and investigative story against which all others

are measured.’’^[25] Watergate was the backdrop against which were viewed the Iran-Contra scandal and scandals surrounding President Clinton’s Administration. The boundaries between public and private behavior shrank as media standards for public disclosure shifted, along with public expectations. A culture of scandal and scandal-mongering changed the nature of the presidency and its relationship with the media, Congress, oversight and investigative offices, and the public.^[26]

“The events and repercussions collectively called Watergate symbolize breaking faith, breaking trust, and breaking laws” and evolved into a symbol of abuse of public office and deception.^[3]

It was a test of the American political system, and perhaps its greatest test. But it fell short of a constitutional crisis and instead demonstrated the strengths of the system. Nixon offered this assessment in his speech in April 1973: “Some people, quite properly appalled at the abuses that occurred, will say that Watergate demonstrates the bankruptcy of the American political system. I believe precisely the opposite is true. Watergate represented a series of illegal acts and bad judgments by a number of individuals. It was the system that has brought the facts to light and that will bring those guilty to justice.” Twenty-five years later, Bob Woodward agreed with the central proposition (if not with the identity of the guilty) and said, “I believe Watergate shows that the system did work.’’^[27] Nevertheless, the public’s trust in government and idealism about politics and political leaders were severely shaken by the Watergate scandal that left an indelible mark on American public life.

REFERENCES

- Graham, K. *The Watergate Watershed: A Turning Point For A Nation And A Newspaper*; Washington Post, January 28, 1997; D01 at <http://www.washingtonpost.com/wp-srv/national/longterm/watergate/stories/graham.htm>, accessed August 2001. Graham was the publisher of the Washington Post during Watergate.
- Parts of the story such as the identity of “Deep Throat,” the background informant to Washington Post reporter Bob Woodward, are still unknown. See <http://www.washingtonpost.com/wp-srv/national/longterm/watergate/front.htm>.
- Lewis, C.W. Watergate: a public service parable. *PA Times* **1992**, (newspaper of American Society for Public Administration) (June), unpaginated insert.
- Kutler, S. *Abuse of Power: The New Nixon Tapes*; Free Press, Simon & Schuster: New York, 1997; ix–xii.
- Pew research center for the people and the press. *Deconstructing Distrust, How Americans View Government*; 1998; 15 pp.
- Schell, J. *The Time of Illusion*; Random House: New York, 1975.
- White, T.H. *Breach of Faith, The Fall of Richard Nixon*; Dell: New York, 1975; pp. 412, 413.
- Weiner, T. Presidential tapes: nixon wanted to show up JFK. *N. Y. Times* **Feb. 28, 1999**, at <http://www.nytimes.com>, accessed August 2001. Sixty-three hours of the Nixon White House tapes are available for public listening at the National Archives and Records Administration’s Office of Presidential Libraries. See also the Richard Nixon Library at <http://www.nixonfoundation.org>.
- Clark, K. The legacy of watergate for legal ethics instruction. *Hastings Law J.* **2000**, 51 (4), p. 673. See also the roster of participants on pp. 678, 679.
- Congressional quarterly. *Watergate: Chronology of a Crisis*; Congressional Quarterly: Washington, DC, 1974; Vol. 1–2.
- Bernstein, C.; Woodward, B. *All the President’s Men*; Simon & Schuster: New York, 1974.
- Bernstein, C.; Woodward, B. *Bug Suspect Got Campaign Funds*; Washington Post, August 1, 1972; A01 at <http://www.washingtonpost.com/wp-srv/national/longterm/watergate/articles/080172-1.htm>, accessed August 2001.
- Bernstein, C.; Woodward, B. *FBI Finds Nixon Aides Sabotaged Democrats*; Washington Post, October 10, 1972; A01 at <http://www.washingtonpost.com/wp-srv/national/longterm/watergate/articles/101072-1.htm>, accessed August 2001.
- Eastland, T. *Ethics, Politics and the Independent Counsel*; National Legal Center for the Public Interest: Washington, DC, 1989; pp. 9, 14–16.
- Lyons, R.; Chapman, W. *Judiciary Committee Approves Article to Impeach President Nixon, 27 to 11, 6 Republicans Join Democrats to Pass Obstruction Charge*; Washington Post, July 28, 1974; A01 at <http://www.washingtonpost.com/wp-srv/national/longterm/watergate/articles/072874-1.htm>, accessed August 2001.
- White House biography of President Richard M. Nixon at; <http://www.whitehouse.gov/history/presidents/rn37.html> accessed September 2001.
- No precedents. *Time Mag.* **August 19, 1974**, at <http://www.time.com/time/magazine/special/nixon/nixon2.html>, accessed June 2001.
- A facsimile of Richard M. Nixon’s resignation letter (NWCTC-59-PI1E767-Nixon) is available at the National Archives at <http://www.nara.gov:80/cgi-bin/starfinder/6354/standard.txt>, accessed September 2001.
- Pagnozzi, A. Real heroes stand test of time. *Hartford Courant* **May 8, 2001**, at <http://www.ctnow.com>, accessed June 2001.
- Henry, N. *Public Administration and Public Affairs*, 7th Ed.; Prentice Hall: Upper Saddle River, NJ, 1999; pp. 197, 311. One effort to sidestep the civil service and its emphasis on apolitical merit was the “Malek Manual,” named after a White House official responsible for seeing to it that partisan politics dominated appointments at executive levels in the civil service. According to this source, “Abuse may be defined as

the inappropriate, unethical, or illegal misuse of the power vested in one's public office''.

21. Stencel, M. *The Reforms*; Washington Post, June 13, 1997; at <http://www.washingtonpost.com/wp-srv/national/longterm/watergate/front.htm>, accessed August 2001. The author goes on, "But holes in the new laws, and the lengthy investigations the reforms sometimes spawned, also deepened voter cynicism''.
22. A 1999 count (based on information from the General Accounting Office) put the number of independent counsel investigations of senior officials, including cabinet secretaries, at 19 since the law's enactment in 1978. Independent Counsels at a Glance, Washington Post, at <http://www.washingtonpost.com>, as of October 14, 1999, accessed November 1999. After reauthorizations, the independent counsel law was allowed to expire in June 1999.
23. Scandal in perspective, interview with Daniel J. Boorstin. *Congr. Q.* **July 7, 1973**, 31 (27), 170.
24. Mosher, F.C.; et al. Watergate, implications for responsible government. In *A Special Report at the Request of the Senate Select Committee on Presidential Campaign Activities of the National Academy of Public Administration*. Foreword by Senator Sam J. Ervin, Jr.; Basic Books: New York, 1974.
25. Washington Post Watergate. **June 1997**, at <http://www.washingtonpost.com/wp-srv/national/longterm/watergate/front.htm>, accessed July 2001.
26. Woodward, B. *Shadow: Five Presidents and the Legacy of Watergate, 1974–1999*; Simon & Schuster: New York, 1999.
27. Interview with Bob Woodward in 1997 at <http://discuss.washingtonpost.com/zforum/97/woodward.htm>, accessed July 2001.

BIBLIOGRAPHY

1. CNN/Time. <http://www.cnn.com/ALLPOLITICS/1997/gen/resources/watergate/index.html>.
2. Gerald Ford Library. <http://www.ford.utexas.edu>.
3. Government Printing Office for federal documents and reports. http://www.access.gpo.gov/su_docs/locators/coredocs/index.html.
4. Library of Congress for federal statutes and current and, historical documents. <http://thomas.loc.gov>.
5. National Archives and Records Administration. <http://www.nara.gov> and <http://www.nara.gov/fedreg/codific/index.html>. Office of Presidential Libraries at <http://www.nara.gov/fedreg/presdoc.html>.
6. PBS's Online NewsHour, June 27, 1997 transcripts on 25th anniversary of Watergate. <http://www.pbs.org>.
7. Pew Research Center for the People and the Press. <http://www.people-press.org>.
8. Richard Nixon Library. <http://www.nixonfoundation.org>.
9. Washington Post. <http://www.washingtonpost.com>.

Welfare and Social Security Policy

Mark Carl Rom

*Georgetown Public Policy Institute (GPPI), Georgetown University,
Washington, District of Columbia, U.S.A.*

INTRODUCTION

The U.S. government provides income support to citizens in two main ways: social insurance and public assistance. (The best single source of program descriptions and data for both types is the *Green Book*.)^[1] Social insurance programs provide benefits to those who qualify by virtue of having contributed through taxes to the program. The largest social insurance programs are social security, which primarily affords pensions to the elderly; Medicare, which supplies health insurance to the elderly; and unemployment insurance (UI), which provides cash benefits to those who have temporarily lost jobs. Public assistance—colloquially called “welfare”—comprises various programs that provide support to the poor. The federal welfare system includes Temporary Assistance to Needy Families (TANF), Food Stamps, Medicaid, Supplemental Security Income, and child nutrition programs, among numerous others. Together, these programs provide cash benefits, health insurance, housing, job training, and other social services to those in poverty.

Social insurance and welfare have different purposes, serve dissimilar groups, and witness divergent politics. After a brief introduction to the distinction between social insurance and public assistance, this article focuses on the policies and politics of social security, the largest social insurance program, and TANF, the most controversial of the public assistance programs.

BODY

Social insurance and public assistance (welfare) programs are distinct in several ways. First, welfare programs are “means tested”—that is, persons can qualify for public assistance only if their income and assets are below programmatic thresholds. These programs explicitly exist to assist the poor. Social insurance programs, in contrast, make benefits available to anyone who has contributed through taxes to the program and has become eligible to receive payments from it; eligibility is determined primarily by

age, for social security and Medicare, and loss of employment, for UI. Social insurance programs are not means tested, so benefits are available to recipients regardless of their income. Second, welfare programs are designed primarily to help those who are incapable of supporting themselves and their families through work. (Determining exactly who can and should be expected to work is usually controversial, however.) Social insurance, in contrast, is to support workers from the loss of income due to age, illness, or job loss. Third, social insurance programs are broadly popular, in part, because they are seen to offer benefits to those who “deserve” them: those who have contributed to the program, the temporarily unemployed, the elderly—and the vast majority of Americans will fall into at least two of those three categories during their lifetimes. Public assistance is far more controversial, for a couple reasons. Although the public shows broad support for helping the “truly needy,” large proportions of the public believe that welfare recipients are lazy, shady, and dependent, and hence undeserving of governmental support.^[2] Moreover, a substantial majority of the public never receives welfare benefits and so apparently believes it has little self-interest in supporting such programs. Finally, as a disproportionate segment of the poor are racial and ethnic minorities, especially African Americans and Hispanics, welfare programs are often seen through racial stereotypes.^[3]

Origins of Social Security and Welfare

Prior to the 1930s, individuals and families were expected to fulfill their own economic requirements, supplemented in times of need by charities or local governments. The onset of the Great Depression pushed millions of Americans into destitution, swamped charitable and local efforts at remedy, and raised questions about the future of capitalism. To address all three concerns, in 1934, President Franklin Roosevelt appointed the Committee on Economic Security (CES) to develop comprehensive legislation to enhance “the security of men, women and children of the Nation

against certain hazards and vicissitudes of life.’’^[4] The legislation the CES proposed contained social security, UI, and Aid to Dependent Children [ADC, which later became Aid to Families with Dependent Children (AFDC) and, in 1996, transformed into TANF], among other programs to assist the blind and disabled. Congress swiftly enacted this legislation as the Social Security Act in 1935. This act continues today as the basis for social security and welfare.

Social Security

As the Great Depression wiped out the life savings of many elderly, leaving their children unable to provide for them, social security was conceived of as an insurance program that would nonetheless begin immediately providing cash benefits to the elderly. Social security has been periodically revised, with most of the initial revisions expanding eligibility and increasing benefits. Originally covering just over one-half of the workforce (those employed in commerce and industry), by the late 1990s, social security covered approximately 96% of all workers, with the main exceptions being those employed by state governments and some federal employees. By 2000, social security paid out about \$387 billion in benefits to the retired, the disabled, and those eligible by the death of an insured worker. It was the most expensive federal program.^[5]

Social security establishes eligibility and benefits in the following way. All “covered” workers pay taxes on a fixed proportion (6.2%) of their wages or salary up to a fixed limit (\$76,200 in 2000). The employer contributes the same proportion to the same limit. Benefits are paid to workers when they retire or become disabled if they have worked (i.e., contributed) long enough. Although the formula varies, persons become “fully covered” (i.e., insured for life) when they have earned one quarter of coverage each year from when they turn 21 until they reach 62 (40 quarters); workers are insured against disability if they have worked at least 20 of the last 40 quarters.^[1]

Workers are eligible for benefits if they retire at age 62, but can receive larger monthly benefits if they defer retirement until age 65 (or, in the future, later). Benefits are calculated based on a formula accounting for the worker’s lifetime “covered” earnings. The formula is fairly complex (based on such factors as when a worker started employment, as well as the duration and continuity of wages), but the essential ideas are these: the longer one has worked and the higher one’s wages, the greater the social security benefit. Benefits are indexed to inflation. There is a special minimum benefit awarded to those who worked many years at very low

wages. For those with relatively high postretirement income, one-half of the social security benefits are subject to income taxes.^[1]

Social security has as one of its primary goals the reduction of poverty among the elderly. To achieve this goal, those with lower wages receive relatively higher social security benefits. For instance, in calculating a person’s benefits, in 2000, social security “replaced” roughly 90% of the first \$531 in monthly earnings, 32% of the next \$2671, and 15% of the rest. Individuals thus received a higher proportion of their preretirement income the lower their wages; they received a higher dollar amount, the higher their income while working.^[1]

Workers’ contributions are not contributed to individual accounts, which means that workers are not literally saving for their retirement; they are paying into a program that will support their retirement. Most revenue collected goes immediately out to pay the benefits of current recipients; the remainder forms one large pool of money, called the Trust Fund. The Trust Fund is invested in government securities (i.e., the government purchases its own bonds), where it earns the prevailing rate of interest on such bonds. The Trust Fund is not invested in a tangible asset, but in an IOU the government has written to itself.^[1]

Social security has long been considered the “third rail” of American politics—touch it and you die. In general, the elderly depend on social security, they believe they have earned its benefits, and they believe it is nearly a sacred contract with the government. Moreover, the elderly vote at higher rates than the public as a whole. It has always been much easier to expand the program than contract it. Politicians who are perceived as willing to tampering with social security often face retribution on election day. The Democratic party, which has traditionally been seen as the main defenders of social security (as it was created by a Democratic president and a Democratic Congress), routinely pillories the Republican party as a threat to the program.

The political untouchability of social security may be changing. The reasons for the change are demographic, economic, and ideological. The demographic and economic reasons are linked. In the coming decades, the American public will grow substantially older on average; the ratio of those retired to those working will shift from about 1 in 5 to 1 in 2.5.^[6] As a result, there will be far fewer workers to pay social security taxes and far more recipients obtaining benefits (Fig. 1). This means that the costs of the program are likely to rise substantially, and the revenue raised to fund it will not keep pace. The Social Security Administration estimates suggest that the Trust Fund will be running an annual deficit (i.e., more is paid out in a given year than is raised) beginning about

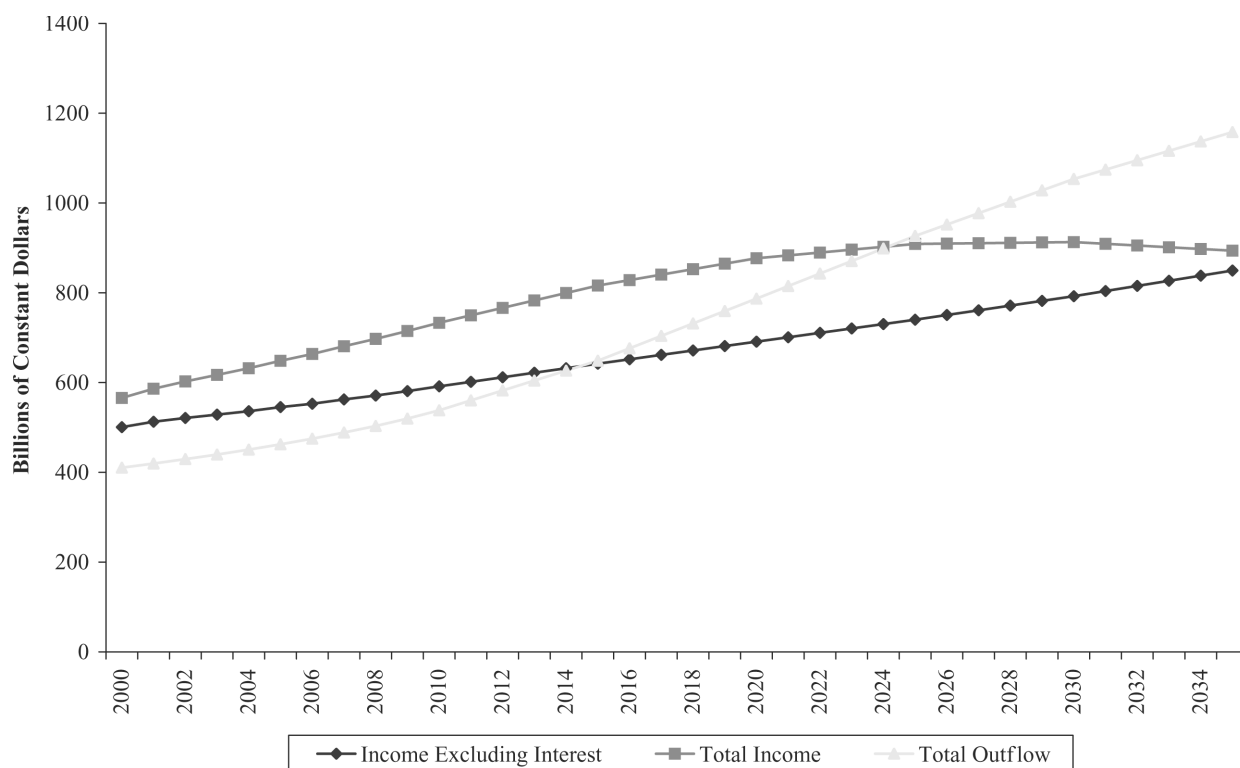


Fig. 1 Social security income, outflow.

2030 or so (depending on many assumptions), and will begin incurring debts (i.e., the Trust Fund will be exhausted) by around 2040.^[6] Although some might argue that the social security system is going bankrupt, it is virtually unthinkable that the federal government would default on its obligations.

Nonetheless, given current projections, most analysts agree that, at some point, either taxes must be raised, benefits reduced, or the retirement age must be raised, unless some way can be found to increase the rate of return on revenue brought into the Trust Fund.^a Raising the rate of return seems to be the most attractive choice politically, as it is the only one that appears not to hurt anyone. The rate of return could conceivably be increased if some or all the Trust Fund was invested in non-governmental assets (i.e., stocks, commercial bonds, etc.), or if individuals could directly invest some of their contributions in such assets. Those holding a promarket ideology certainly favor such options, and plans to privatize social security have become more reputable in recent years. Critics raise three main objections to privatization: 1) the transition

from the current pay-as-you-go system to a privatized system would be costly; 2) higher administrative costs would probably erode the higher returns; and 3) higher returns would come only with (unacceptably) higher risks to the beneficiaries.

One political dilemma is that the program's long-term financial difficulties will be easier to fix technically (i.e., smaller changes will be required) the sooner they are addressed. Yet, as long as the financial problems seem far in the future, the political incentives to face them now are fairly weak. Unfortunately, the longer policy makers wait to address the problems, the larger they are likely to become, and the harder the political choices will be.

TANF

TANF's predecessor, AFDC, was created as a minor and uncontroversial element of the original Social Security Act. Aid to Families with Dependent Children was designed to provide cash benefits to families where the father had died or was otherwise absent from the home. The federal government set broad program standards and provided a majority of the funding, but left program administration to the states. The states established a eligibility criteria, set benefit levels,

^aProjecting the long-term future (e.g., 30 or more years from now) is fraught with uncertainties regarding demographics and economics. For various assumptions and the wide variation in the state of the economy and social security, see Ref.^[7].

and contributed a portion of benefit payments. Eligibility, benefits, and participation varied widely from state to state.^[8]

Aid to Families with Dependent Children was intended to be a temporary program providing income support to widows and their families. It did not turn out that way, however. Fifty years after its creation, AFDC provided assistance to more than 14 million Americans. The vast majority of the parents it assisted were not widows, but parents who were divorced, separated, or never married. Only a small percentage of the mothers receiving AFDC had paid employment, even though by the 1990s, most other mothers did earn income by working. (Aid to Families with Dependent Children did have a work element, the so-called “JOBS” program, but only about 10% of AFDC adults actually participated in it.) At any given time, almost two-thirds of adults receiving AFDC had been “on the dole” for 8 or more years. Finally, almost two-thirds of the recipients were non-white minorities.^[1] In short, a large proportion of adult AFDC recipients were minority single mothers with little earned income who remained on the rolls for years. For these reasons, AFDC came to be seen as a program that discouraged marriage and work, encouraged out-of-wedlock childbearing and dependency, and primarily served minorities. The critics of AFDC held that welfare recipients became dependent on the government because, in part, the government did little to encourage independence.

Presidents since Nixon had called for major restructuring or elimination of AFDC, and President Clinton’s most popular campaign pledge was to “end welfare as we know it.” In 1996, a Republican-controlled Congress took the President at his word and enacted legislation that eliminated AFDC and created TANF in its place; President Clinton, against the advice of his more liberal advisors, signed the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) into law.^[2]

Temporary Assistance to Needy Families differed from AFDC in several important ways. (Program details are summarized in the *Green Book*).^[1] First, it eliminated the individual entitlement to benefits. Poor families would no longer be guaranteed benefits because the states were given authority to decide which families were eligible, to divert families from entering TANF or to sanction those already receiving benefits. The law also set a 5-year lifetime limit on individuals receiving federal benefits (although 20% of a state’s caseload may be exempt from this provision due to “hardship”). Second, it made work, rather than income support, the primary focus of the program. To encourage work, PRWORA set a specified and gradually rising proportion of each state’s welfare

caseload that must be working at least 30 hours per week. The law required that beneficiaries be working within 2 years of obtaining benefits, and allowed states to set shorter time limits. (Wisconsin, for example, requires potential beneficiaries to be engaged in work efforts before they receive their first benefit check.) Third, PRWORA gave the states substantially more flexibility to design and implement TANF; the law allowed the states to delegate responsibility to the counties (e.g., as California has done). The law also eliminated open-ended funding to the states by providing them a fixed-amount block grant each year from 1997 to 2002 (funding levels must be reauthorized in 2002). States, for their part, had to spend at least 75% as much on welfare as they were when PRWORA was enacted. Personal Responsibility and Work Opportunity Reconciliation Act did authorize certain bonus payments to the states if they, for example, reduced non-marital birth rates while also reducing abortion rates.

Temporary Assistance to Needy Families impact on welfare caseloads is nothing short of astonishing. Although caseloads had already begun shrinking by 1994, by the end of 2000, the number of individuals enrolled in TANF had fallen by almost 60% from its all-time high (Fig. 2). In 2002, there are fewer families receiving welfare than at any time since the 1960s. There is substantial debate regarding the relative weight of economic conditions and TANF implementation on caseload declines, but by all accounts both have played substantial roles. The greatest change in welfare policy in 60 years occurred in a uniquely favorable economic climate.

Temporary Assistance to Needy Families has also increased the work effort of those on welfare and potentially eligible for it. By 1998, approximately 35% of those enrolled in the program were meeting their 30 hour per week work requirements, up from about 5% in the early 1980s. The true impact of the program on employment is not known, in large part because a substantial portion of the AFDC caseload apparently was working off the books, and also because the program does not track the work effort of those diverted from the program or those who are eligible but have not applied.

Perhaps the most controversial question regarding TANF has been its impact on poverty rates and the income of the poor, especially given the positive economic conditions of the late 1990s. Some substantial fraction of families that had been eligible for AFDC had improved incomes after TANF was implemented, primarily because they had found unsubsidized employment in the private sector. It appears that a smaller portion were worse off because they received smaller benefits, or had been driven off the rolls and

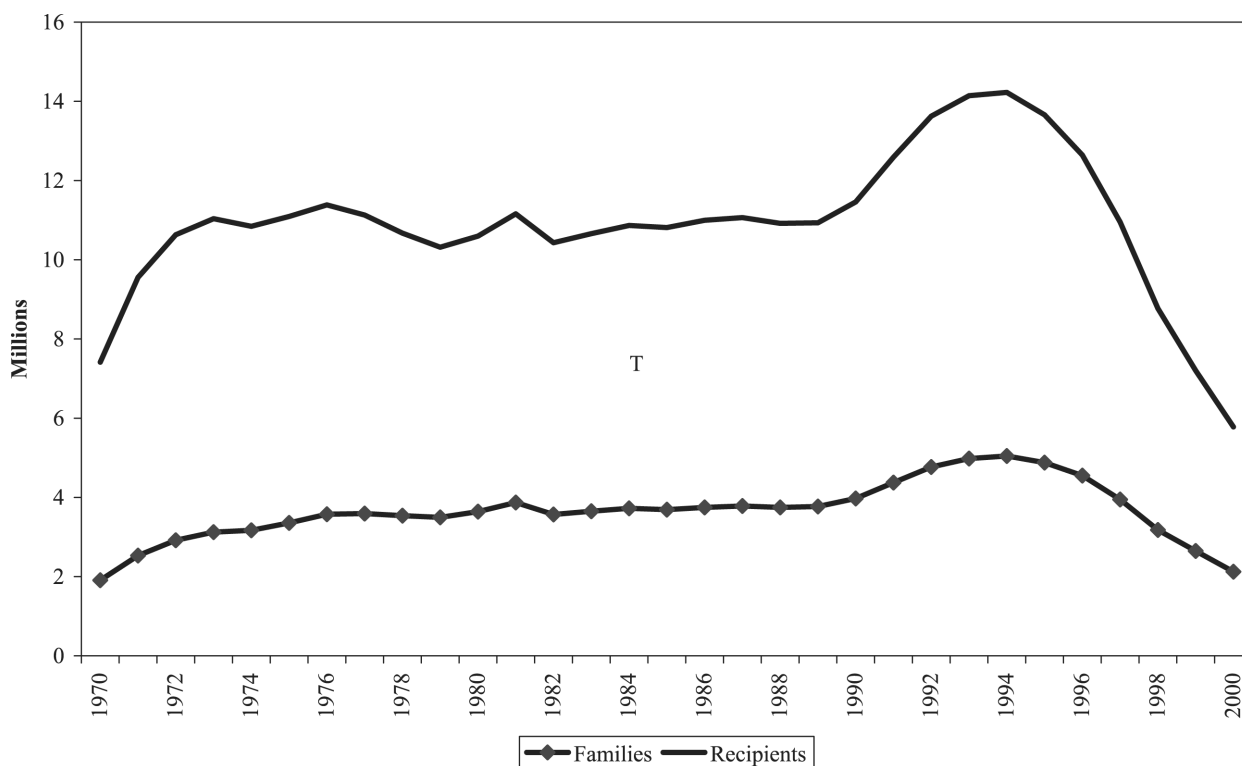


Fig. 2. AFDC/TANF caseloads.

had not found employment. A large segment of the AFDC population's income did not change much under TANF.^[9]

Both Democrats and Republicans accept TANF's basic structure, and no major voices call for a return to AFDC. The major controversies concern work requirements, funding, and moral provisions. The Bush administration and Republicans in Congress have called for increasing the proportion of the caseload that must be working, and the number of hours they must work weekly. Democrats (and most governors) find these provisions unduly onerous. Regarding funding, those who prefer cuts focus on the major declines in the caseloads; those who favor increased (or at least level) funding contend that the remaining caseload is more expensive to serve. Finally, the Bush administration has called for TANF funding to promote marriage. The public strongly values marriage as an institution, but is strongly skeptical of the government's need or ability to promote it.

CONCLUSION

Social security is the largest and most important social insurance program. It has been a tremendous success economically, socially, and politically because it has

helped millions of elderly (and disabled) Americans to have a decent income after their working years have ended. Social security has substantially lowered the poverty rate among the elderly and has provided generations of Americans the confidence of having a financially secure retirement. However, it does face challenges for the future.

Welfare programs have been much smaller and more controversial. Aid to Families with Dependent Children was widely seen as a failure by liberals because it provided scant support for the poor and by conservatives because it encouraged dependency. Although many analysts viewed systematic reform as nearly impossible, the creation and implementation of TANF as AFDC's replacement is commonly interpreted as a major political and programmatic success. Whether TANF can be modified and sustained in ways that keep caseloads low while also reducing poverty among families remains an open question.

REFERENCES

1. U.S. House of Representatives, Committee on Ways and Means. *Green Book*; Government Printing Office: Washington, DC, 2000; 4–10, 16–27, 420–422.
2. Weaver, R.K. *Ending Welfare as We Know It*; Brookings Institution: Washington, DC, 2000; 172–175.

3. Gilens, M. *Why Americans Hate Welfare: Race, Media, and the Politics of Antipoverty Policy*; University of Chicago Press: Chicago, 1999; 61–79.
4. U.S. House of Representatives Hearings before the House Committee on Ways and Means. *Economic Security Act*; Government Printing Office: Washington, DC, 1935; 13 pp.
5. Office of Management and Budget. *United States Budget, FY 2001*; Government Printing Office: Washington, DC, 2001.
6. Social Security Administration. *Annual Report of the Trustees of the Old Age, Survivors, and Disability Trust Funds*; Government Printing Office: Washington, DC, 2001. Table V.A2, Table III.B2.
7. Gist, J.; Verma, S. *The Federal Budget, the Economy, and the Future of Entitlements*; AARP Public Policy Institute: Washington, DC, 2002.
8. U.S. House of Representatives, Committee on Ways and Means. *Green Book*; Government Printing Office: Washington, DC, 1996; 473, 474, 505.
9. Weil, A. *Ten Things Everyone Should Know About Welfare Reform*; Urban Institute: Washington, DC, 2002.

Welfare Economics

Glenn A. Daley

RAND Graduate School, Santa Monica, California, U.S.A.

INTRODUCTION

Welfare economics is the study of how economic well-being is (or is not) advanced by markets, and thus by government intervention in markets. To some, this implies a value judgment inconsistent with the ideal of scientific objectivity. One way to approach this is by analogy with the field of medicine: just as understanding the body helps in prescribing actions that contribute to health, understanding how the economy works helps in prescribing policy. Nevertheless, it is useful to remember the differences between understanding why something happens, analyzing the consequences of proposals, and making values-based recommendations. In economics, an example of the first would be explaining why the price of gasoline rises and falls; the second would be predicting the consequences of a price ceiling on gasoline; and the third would be recommending for or against a price ceiling based on criteria such as efficiency or fairness.

Welfare economics includes all three of these pursuits, but ultimately aims toward the third. Its method is to make assumptions about well-being and then to explore the implications of these assumptions: 1) by raising questions about the assumptions in an attempt to keep them limited, flexible, consistent, compatible with experience, and yet capable of generating useful results; and 2) seeking to identify areas where the real world does not match the assumptions of the model, areas where other values might be accommodated without abandoning the model, and areas where the methods gained from the model are valid enough to use in the real world. Welfare economics provides the foundation for the applied discipline of benefit-cost analysis.

WELFARE AND UTILITY

The first task for welfare economics is to define economic welfare for an individual. Although economists may disagree on exactly where to draw the line, there is widespread agreement on the basic approach described here (which we refer to as the “standard model”). This model only applies to goods and services that are, or potentially could be, obtained from

other people or provided to other people, whether as a gift, for money, through barter, or in some implied quid pro quo arrangement. (Hereafter, “goods” refer to such goods and services, understood broadly to include leisure time and other things that might be given up for something else.) To simplify the mathematics, we assume that people prefer more of these goods to less; in the case of a preference for less of something (e.g., air pollution), the absence of the thing may be treated as a good. Furthermore, each person may have different preferences between goods, so individual welfare is based on a person’s unique set of preferences. This is consistent with the tradition of democratic individualism in allowing everyone to decide for themselves what constitutes their own well-being (with exceptions, such as parents deciding for children).

Thus, in one sense, an individual’s economic well-being is whatever value, satisfaction, or happiness that the individual receives (or expects to receive) from economic goods as defined above. The word “utility” is used to identify this kind of well-being. This definition makes contemporary economists uneasy, though, because it is based on subjective effects such as “satisfaction” rather than on observable behavior. Thus, the definition used today is that utility is the unobserved content of an individual’s preferences. Although preferences are subjective, in principle they may be revealed in a person’s choices—without any further guessing about the person’s emotional state or degree of happiness. So utility is whatever it is that people prefer more of when they show preferences between goods; this provides a highly flexible definition of welfare that defers moral and subjective judgments to the individual rather than to the economist.

Given this concept of individual utility, what is the economic well-being of society as a whole? In the utilitarian tradition, society’s welfare is defined as some aggregate of individual utilities. This means that any values that apply to the community as a whole, completely independent of the utilities of the members of the community, are excluded. This is not to deny their validity or importance; it is only to say by definition that they are not the same as economic welfare. However, some of the community-level values that people care about, such as equity in the distribution of wealth,

may still be accommodated in an economic model through a social welfare function, such as the weighting of individual utilities.

The concept of aggregate utility raises the question of how to combine and compare individual utilities. Early welfare economists attempted to define utility in a way that could potentially be measured and summed together on a cardinal scale (i.e., a measurement system with countable units that are equal and comparable to each other), just as the aggregate weight of passengers on an airplane is the sum of the individual weights. This proved to be a difficult concept; it requires a standard unit of measurement and a zero point on the scale, and it assumes that utility can be compared between people. If my utility and yours are to be added together, then some third person must, in principle, be able to tell whether you have more utility than I do. This leaves the door open for some person or institution to take on the paternalistic role of defining utility for you and me, something we want to avoid. Because of these difficulties, the assumption of cardinal utility—the basis of “old welfare economics”—has been abandoned.

The “new welfare economics” is based on the ordinal concept of utility, the idea that utility levels can be ordered or ranked by an individual, even though they might not be measurable in additive units. This is consistent with the definition of utility as a matter of preferences. The question is not how much utility a person gets from a bundle of goods, but whether the person prefers one specific bundle in comparison to another specific bundle.

Ordinal utility is identified by pairwise comparisons between alternative bundles of goods. For every possible pair of bundles, which we can label as X and Y, a person either prefers X, prefers Y, or is indifferent between X and Y. The map of all these pairwise preferences defines utility for an individual. If people prefer X to Y, we say that their welfare, or utility, is greater from getting X than from getting Y, and, if they have the opportunity to exchange X for Y, they will voluntarily do so, thereby obtaining an increase in utility.

THE PARETO CRITERION

The ordinal concept of utility still does not tell us how to think about society’s welfare because it cannot be added between individuals. The answer is to concentrate on changes in utility at the margin (i.e., preferences between a change and its absence). A transaction or a policy decision might make one individual better off (or worse off) in terms of that person’s preferences, but the same event might affect other people in different ways. For instance, a policy that takes goods from me and gives them to you would clearly reduce

my utility, and increase your utility. So far, we cannot tell whether your gain is greater than my loss; we might have different preferences for these goods in the context of the other goods we have. So, this example is ambiguous in its effect.

However, we can imagine an event that unambiguously represents an improvement in society’s welfare. Suppose I have bundle X but prefer Y, and you have bundle Y but prefer X. I can increase my own utility, and you can increase yours, by swapping. Because both of us are better off, and assuming no one else is made worse off, we have increased society’s welfare without having to add or compare utility between individuals. An event improves the welfare of society if it makes at least one person better off in terms of that person’s preferences, and does not make anyone else worse off in terms of their own preferences. Such a change is called a Pareto improvement, after Vilfredo Pareto, the nineteenth-century Italian social scientist who first stated this rule. In terms of economic welfare defined as preference-based utility, Pareto improvements are unquestionably desirable.

The Pareto rule might seem too restrictive because nearly any policy change will make some members of society worse off. However, the losers might be compensated. If person A gains from a change, and person B loses, perhaps A’s gain is large enough that A is willing to pay B for the loss. As long as A prefers or is indifferent to the change even after making this payment, and B prefers or is indifferent to the change conditional on receiving the payment, no one has suffered a net loss of utility, and the Pareto rule is satisfied. This sort of compensation occurs all the time in private market transactions; in public policy, though, such payments are sometimes outlawed as “bribes.” Still, it is often possible to design policy programs to include such compensation, and thus satisfy the Pareto criterion. Yet, the Pareto criterion is never satisfied when the winners do not gain enough to compensate the losers.

Suppose a point is reached where no more Pareto improvements are possible; society has exhausted every opportunity to increase anyone’s welfare without reducing the welfare of someone else. Such a point is called a Pareto-optimal point, or a Pareto-efficient point. However, there are many possible Pareto-efficient points, not all of which are equally achievable or desirable. The set of all Pareto-efficient points is an important concept in theory, but practical policy-making focuses on making continuous Pareto improvements rather than formally defining the end state. The Pareto criterion is neither sufficient nor necessary for ethical judgments about society’s well-being, but it does establish a benchmark: it places the burden of justification on any policy proposal that violates it, and lends support to any policy proposal that satisfies it.

COMPETITIVE MARKET OUTCOMES

To apply the Pareto criterion to economic outcomes, we need to identify what we mean by a competitive market equilibrium. This occurs when supply and demand are in balance, and markets clear. In other words, there is no unmet or excess demand at the market price, and no unused or excess supply at that price. If the market is perfectly competitive, then every participant is a price taker; everyone faces the same prices for the same goods, and no one buys or sells enough to influence prices. Market prices (and, thus, quantities bought and sold) are determined by supply and demand.

Both supply and demand depend on the utility-seeking efforts of many individuals. Consumers seek to maximize utility by purchasing goods; they are constrained, however, by a budget—their available income or wealth. As long as goods have prices, reflecting the fact that others also want them, one cannot obtain them without giving up something. Given a certain set of prices and a certain income or wealth constraint, a person will choose the quantities that maximize total utility. Across all individuals, these are the quantities demanded in the market.

Likewise, goods are provided by firms or individuals who want to maximize their net gain by selling. We commonly assume that firms seek to maximize money profits, which are simpler to calculate than utility. However, firms' costs depend on the hired labor of utility-seeking workers, and some firms operate as though pursuing other goals or utilities, rather than simple profits. So, utility maximization is essential to supply as well as demand. At a given set of prices, firms and individuals will supply the quantities that maximize their own utilities (or profits).

One of the mathematical results of welfare economics^[1–3] is that, given some basic assumptions about the competitiveness of markets and the rationality of preferences (“rationality” referring to consistency in how preferences relate to each other, *not* whether the preferences themselves are based on “reason”), a person will buy quantities of goods such that the marginal utility of each good—the contribution to utility of the last unit purchased—is proportional to its market price. (If utility can be measured in monetary terms, we might even say that the marginal utility is equal to the price. This is a simplification; in terms of preferences, the person is indifferent between having one more unit of a good and using the same money toward buying one more unit of a different good.) Another result of these assumptions is that a firm will sell quantities of goods such that the marginal cost of each good—the contribution to cost of the last unit sold—is equal to its price. As a result, goods are produced and purchased in just the right quantities so that the

marginal utility of *each* good to *each* consumer is exactly proportional (or equal) to its marginal cost to society of production, across all goods and consumers.

In the short run, when goods are bought and sold from existing inventories (or existing workers' available time and skills, in the case of services), a market simply consists of buying and selling decisions. Over time, though, more or less of each good may be produced, so a market equilibrium also comes to represent the cumulative production decisions of firms; thus, it determines the allocation of society's resources between the various goods possible. As a result of utility and profit seeking, this equilibrium allocation represents the production level of each good where the marginal cost equals the price and, thus, the marginal utility to consumers. A competitive market equilibrium provides the best possible match between the allocation of resources to the production of goods, and the preferences and budget constraints of all consumers.

THE FIRST FUNDAMENTAL THEOREM

The First Fundamental Theorem of welfare economics says that any such perfectly competitive market equilibrium is in fact a Pareto-efficient allocation (given certain assumptions). This can be proven mathematically,^[1] but it can also be seen by intuitive reasoning. It is a later version of Adam Smith's idea that society's welfare is promoted by the market's “Invisible Hand.”

Welfare economics assumes that, if given the opportunity, people will always exchange goods they possess but do not prefer for goods that they prefer but do not possess; however, a voluntary exchange will *not* occur if one of the parties is made worse off by the transaction. (Here, we ignore asymmetric information, risk, and derived utilities, such as the satisfaction of giving. Such complexities can be added later as extensions of the model or recognized as violations of the assumptions.) Thus, a voluntary exchange transaction is a Pareto improvement. Assuming that all people who have or want exchangeable goods at the same price can find each other, and firms are able to sell all they make at this price, the resulting allocation must be Pareto efficient.

The analysis of markets in terms of Pareto efficiency can be illustrated with a graphic device called the Edgeworth box, shown in Fig. 1. This represents a simplified economy in which there are only two persons and two goods, and all goods must be held by the two persons. The axes represent quantities of goods X and Y held by persons A and B, with A's holdings increasing up and to the right, and B's holdings increasing down and to the left. The curves in the graph are indifference curves; each one represents a set of combinations of X and Y that an individual is

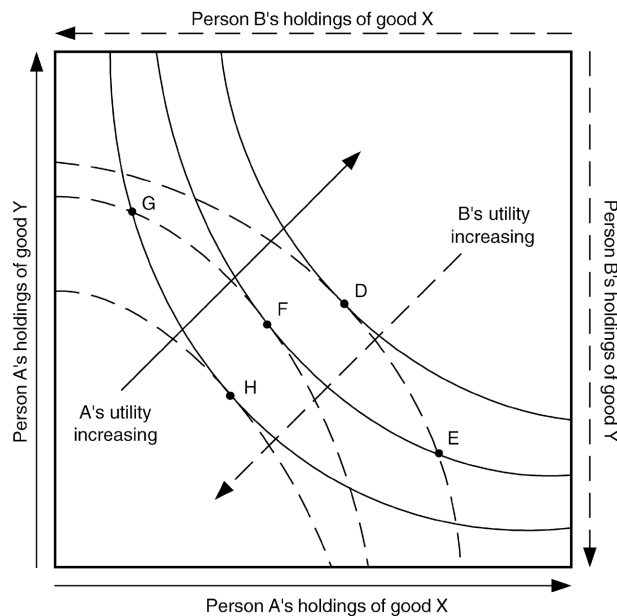


Fig. 1 Edgeworth box with indifference curves. A prefers D to either E or F; is indifferent between E and F; prefers E or F to G or H; and is indifferent between G and H. B prefers H to either F or G; is indifferent between F and G; prefers F or G to D or E; and is indifferent between D and E.

indifferent among. The convexity of these curves is based on certain assumptions about the rationality of preferences, and the observation that people generally prefer mixed combinations of goods rather than having all of one thing and none of another (again, a simplification; for more, see Refs.^[1,2,4]). There is an infinite number of indifference curves, or at least enough so that every possible point (X,Y) lies on *one and only one* indifference curve for each person.

Assuming goods cannot be held by both persons, the only way to increase one's utility is to exchange goods, but an exchange must be agreeable to both. So each person faces a constraint. What is the highest utility possible for one person with the other person's utility fixed? In Fig. 2, with person B's utility fixed at curve P, person A achieves the highest possible utility on curve R, which touches curve P at only one place: point K, where the two curves are tangent. Mathematically, this is where the local slopes (or derivatives) of the two curves are equal. This is another way of saying that the rate at which A is indifferent to giving up the last unit of X for another unit of Y is exactly equal to the rate at which B is indifferent to giving up the last unit of Y for another unit of X. The set of all such points where indifference curves are tangent is called the contract curve; whenever the two parties are not on that curve, they will exchange goods until, and only until, they reach a point on it. This curve represents the combinations of X and Y at which both people will be unable to achieve any more gains through exchange;

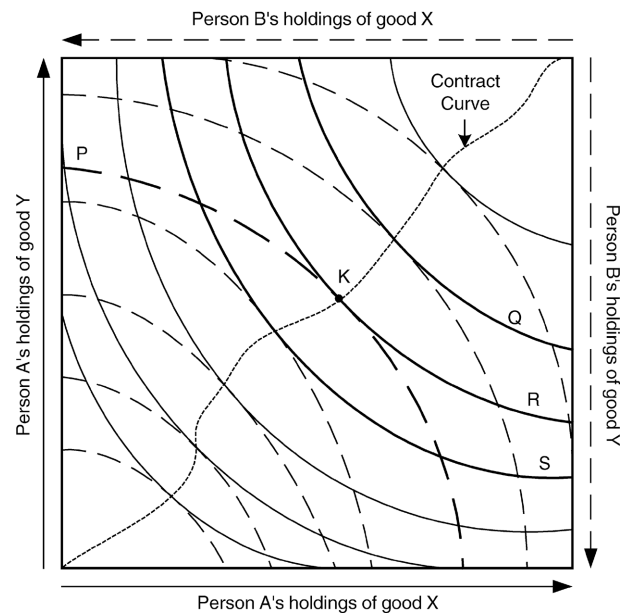


Fig. 2 Edgeworth box with contract curve. With B's utility fixed on indifference curve P: A cannot achieve utility on curve Q because B will not agree. A will not agree to utility curve S because R is better. The highest utility for A (that B will agree to) is on curve R. And the only place on curve R that A can achieve is at point K, where curve R is tangent to curve P. The set of all such tangent points is the contract curve.

once on that curve, they will not mutually agree to move off of it.

The First Fundamental Theorem says that all the points on this contract curve are Pareto efficient points. Any final outcome produced by voluntary exchange between individuals will be a Pareto-efficient point. This simple result can be extended mathematically^[1] to a competitive market of many individuals and firms that produce, sell, and buy goods for money at market prices rather than merely swapping; any such allocation is Pareto efficient.

The First Fundamental Theorem requires some strong assumptions about the market: 1) No third party can be involuntarily affected. If a third party is harmed without compensation, the Pareto criterion is not satisfied, and the price does not equal the marginal cost to society. 2) If a third party stands to gain from a potential transaction that does not take place in sufficient quantity because the primary parties are satisfied with their outcomes, then all possibilities for Pareto improvements have not been achieved, and the marginal benefit for society does not equal the price. These third-party effects are called externalities or public goods, and they violate the assumptions of the theorem. The mathematical proof of the theorem across many buyers and sellers also requires assumptions about the market: 3) All parties are price-takers (i.e., everyone pays the same price because of free

competition), and no one has the power to affect the market price as a monopoly does. 4) Buyers and sellers have all the information and power they need to locate each other, to assess the value of goods, and to make enforceable contracts. Where any of these four assumptions are violated, the outcome may not be a Pareto efficient allocation.

THE SECOND FUNDAMENTAL THEOREM

While the First Fundamental Theorem states that every competitive market equilibrium is Pareto efficient, the Second Fundamental Theorem of Welfare Economics asserts that every possible Pareto-efficient point can be reached by a competitive market from some initial distribution of wealth. Any point not on the contract line provides an opportunity for the parties to make an exchange that will move one or both of them to a higher utility (a more preferred indifference curve), without causing either of them to move to a lower utility. Clearly, which points can be reached depend on the initial distribution. In Fig. 3, point M^* can be reached by Pareto improvements from any point in the region M_0 , whereas point N^* can be reached by Pareto improvements from any point in the region N_0 .

One criticism of the Pareto rule is that it is too conservative because it limits the range of acceptable

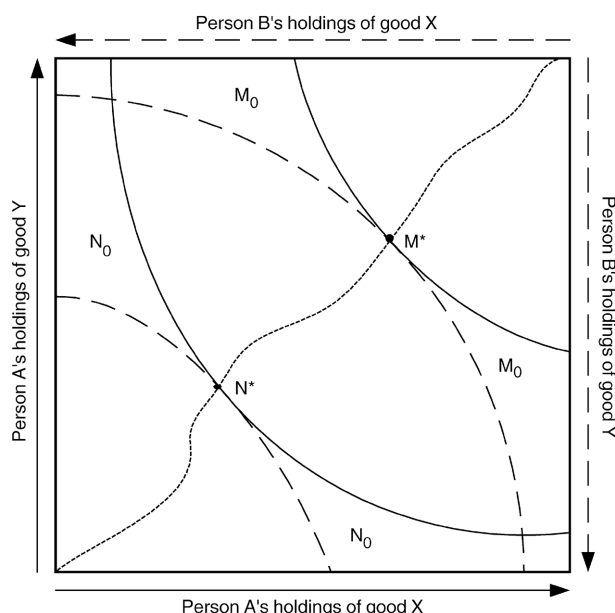


Fig. 3 Pareto improvements from initial endowments. Point M^* is a Pareto-efficient point that can be reached by Pareto improvements from any initial endowment in region M_0 . Point N^* is a Pareto-efficient point that can be reached by Pareto improvements from any initial endowment in region N_0 .

changes to those that can be achieved from the status quo distribution of wealth. Even though an endowment might have resulted from previous events that harmed people and failed the Pareto criterion, any present action to redress the historical harm is likely to reduce someone's current welfare and fail to satisfy the criterion. Even aside from past injuries, a society might believe that less inequality is good for society as a whole and, thus, will object to Pareto efficient points that represent gross inequalities.

The Second Fundamental Theorem shows that such distributive concerns can be addressed by lump-sum redistributions of wealth. In other words, it is possible to separate the issues of economic efficiency and distributive equity. Society can, if it wants, achieve distributive goals through lump-sum redistributions, and let the natural workings of the market (Adam Smith's "Invisible Hand") result in a Pareto-efficient allocation where every individual's marginal utility equals society's marginal cost for all goods. This theorem does not claim that lump-sum redistributions strictly satisfy the Pareto criterion, because they do reduce someone's welfare, but it implies that, if remedying a past injury or reducing inequality is considered important to society's welfare, this can be achieved by a lump-sum redistribution. The Second Fundamental Theorem provides a theoretical basis for separating distributive goals from economic efficiency goals as much as possible in public policy.

The Second Fundamental Theorem requires the same assumptions about a competitive market as the First Fundamental Theorem, but it also requires stricter mathematical assumptions about preferences, particularly the convexity of indifference curves.^[1] Furthermore, the project of assigning lump-sum redistributions requires government to know everyone's utility function and to be able to administer personalized taxes and transfers across the whole population. However, government clearly lacks this knowledge and capacity. In theory, lump-sum taxes and transfers should not distort people's decisions at the margin. In reality, though, most redistribution programs are done through proportional and conditional taxes and transfers (i.e., taxes and transfers that depend on the ongoing economic choices of the individual, rather than representing lump-sum changes in the individual's starting position); these have the effect of changing the slopes of budget constraints and, thus, distorting economic behavior. So, the Second Fundamental Theorem is severely limited in applicability.

THE KALDOR-HICKS CRITERION

There are many situations where the competitive market assumptions do not hold; therefore, a

Pareto-efficient contract line cannot be reached through voluntary exchanges from any initial point. In such cases, the idea of compensation provides a useful decision rule. We have already seen that actual compensation can be included in the strict Pareto criterion, but what if the compensation is not made? If winners gain enough in welfare that they would be willing to pay the compensation and losers would be willing to accept the compensation, then the net welfare of society has been increased, even without actual compensation. It is often impossible or excessively costly to find all the winners and losers and arrange the correct compensations. A modification of the Pareto criterion, called the Kaldor-Hicks criterion after two twentieth-century British economists, says that a policy change that increases net welfare should still be made.

Is it justifiable to make someone worse off because someone else is made better off by a greater amount? The Kaldor-Hicks criterion seems to violate the idea of fairness. However, it must be understood in light of the Second Fundamental Theorem. In theory, all the wealth of society is subject to redistribution. Thus, it makes sense to increase the total wealth of society as much as possible and to deal with questions of fairness in the same way that we remedy past injustices: by lump-sum redistributions, if possible. A policy that satisfies the Kaldor-Hicks criterion is one that increases the net welfare of society. Concerns about who is advantaged or disadvantaged can be addressed, if necessary, through redistribution of that net welfare. The Kaldor-Hicks criterion leads to this fundamental rule of benefit-cost analysis and policy making: that policies should be chosen to produce the greatest possible net welfare of society, as measured by subtracting the losses of losers from the gains of winners.^[5,6]

We started with the claim that welfare cannot be compared between individuals and cannot be measured in units, but can only be ranked on an individual, ordinal scale. Yet, we ended with a decision rule that makes just such a comparison. The idea of compensation provides the needed tool for this purpose: gains and losses in individual utility can be measured in terms of the monetary compensation that winners are willing to pay and losers are willing to accept. Thus, the monetary values of such hypothetical compensations serve as a useful approximation of welfare changes at the margin, allowing comparisons between gains and losses of different individuals.

This concept of compensation is closely associated with consumer surplus, the extra utility obtained by a consumer who can buy the demanded quantity of a good all at one price; such a consumer might be *willing* to pay a different price for every unit, equivalent to that unit's marginal contribution to utility (see Refs.^[1,2] for the mathematics). If a good represents a small fraction of one's total expenditures (a condition

approximated by an assumption called “quasilinear demand”), the increase in consumer surplus resulting from a change turns out to be the same as the willingness to pay compensation for that change. In Fig. 4, the demand curve (shown as linear) represents the consumer's willingness to pay for each additional unit at the margin. The area under this curve is the total willingness to pay, or the monetary equivalent of utility. The rectangular area under price line P_0 is what must actually be paid to get the quantity Q_0 , which is demanded at that price. So the triangular area above P_0 is the difference between willingness to pay and actual expenditures at that price; this is consumer surplus. If the price goes down to P_1 , the quantity purchased increases to Q_1 , total utility increases, and consumer surplus increases by the trapezoidal area between the two price lines. This increase in surplus is the consumer's willingness to pay compensation for a policy that effectively reduces the price. So, the Kaldor-Hicks criterion, to select the policy that yields the greatest net benefit, can often be measured in terms of net gains in consumer surplus—a standard tool in benefit-cost analysis.

DIFFICULTIES AND EXTENSIONS

In the real world, of course, the competitive market assumptions of the First Fundamental Theorem often

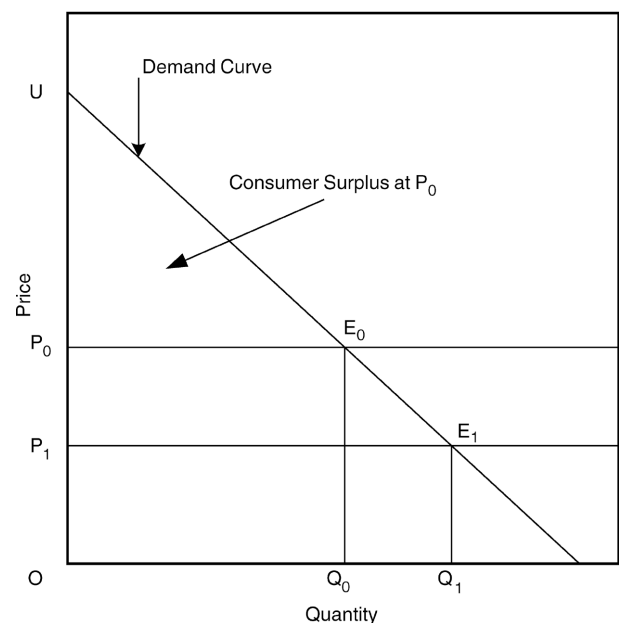


Fig. 4 Consumer surplus. At price P_0 , the consumer chooses quantity Q_0 . Total utility is the area of trapezoid OQ_0E_0U . Total expenditure is the area of rectangle $OQ_0E_0P_0$. Consumer surplus is the area of triangle P_0E_0U . With a price decrease to P_1 , the increase in consumer surplus is the area of trapezoid $P_1E_1E_0P_0$.

do not hold, and government lacks the knowledge and administrative ability necessary to make the individualized lump-sum redistributions required by the Second Fundamental Theorem. One of the basic tasks of welfare economics is to identify when these assumptions are not met, to evaluate the resulting loss of welfare, and to analyze whether a specific government intervention would still be a Pareto improvement (or Kaldor-Hicks improvement) over the market outcome. Many economists view the two theorems as an argument against too much government intervention, and treat violations of the assumptions as exceptional cases to be addressed on a case-by-case basis. Other economists assert that the assumptions are so rarely satisfied that there exist many opportunities for government to create genuine Pareto improvements by intervention.^[7]

Some new perspectives on welfare economics are being provided by transaction cost and institutional economics. One is the insight that transaction costs result in market imperfections even more widespread than the traditionally recognized market failures. This argues against the assumptions of the two theorems and suggests a stronger role for government. However, there is the insight that economic, social, and cultural institutions may evolve so as to achieve efficiency by reducing transaction costs through means other than strictly competitive market exchanges. This suggests that even an imperfect market system has strengths unrecognized in the price equilibria of the standard model and may already be achieving efficiencies greater than those achievable through policy interventions.

There is also a new awareness of political economy. When social welfare values other than economic efficiency are to be pursued, the market process is insufficient and must be complemented with other public choice processes to reveal society's overall

preferences. However, when there are parties whose gains and losses might be altered by public policy, they have incentives to influence public decision making for their own advantage, so government intervention may result in neither Pareto improvements nor overall gains in social welfare.

In its standard model, and in these extensions, welfare economics provides arguments both for and against government interventions. It is still necessary for policy makers and analysts to consider how well each of these arguments applies to any particular decision situation.^[8]

REFERENCES

1. Mas-Colell, A.; Whinston, M.D.; Green, J.R. *Microeconomic Theory*; Oxford University Press: New York, 1995; pp. 5–159, 311–343, 515–683.
2. Varian, H.R. *Microeconomic Analysis*, 3rd Ed.; W.W. Norton & Company: New York, 1992; pp. 1–171, 215–230, 314–357, 387–438.
3. Graaf, J.de V. *Theoretical Welfare Economics*; Cambridge University Press: Cambridge, 1967; 1–171.
4. Varian, H.R. *Intermediate Microeconomics: A Modern Approach*, 5th Ed.; W.W. Norton & Company: New York, 1999; pp. 1–177, 245–305, 508–569.
5. Stokey, E.; Zeckhauser, R. *A Primer for Policy Analysis*; W.W. Norton & Company: New York, 1978; pp. 134–158, 257–319.
6. Gramlich, E.M. *A Guide to Benefit-Cost Analysis*; Waveland Press: Prospect Heights, IL, 1990; 1–45.
7. Stiglitz, J.E. *The Invisible Hand and Modern Welfare Economics*; National Bureau of Economic Research: Cambridge, MA, 1991; 1–41.
8. Little, I.M.D. *A Critique of Welfare Economics*; Oxford University Press: Oxford, 1957; 274–279.

Welfare Reform, Management of

Edward T. Jennings Jr.

Department of Political Science, University of Kentucky, Lexington, Kentucky, U.S.A.

INTRODUCTION

In the contemporary version of welfare reform, instead of processing claims for benefits, welfare organizations become agents for change, responsible for moving recipients and potential recipients into jobs and self-sufficiency. This requires a shift of organizational culture from one that views clients as victims needing support and help to one that views clients as responsible for their needs, requiring encouragement, motivation, opportunity, and services to accept that responsibility. All of this grows out of the dramatic change in welfare policy initiated in welfare-to-work experiments of the early 1980s, facilitated by the Family Support Act of 1988, and conclusively championed in the Personal Responsibility and Work Opportunity Reconciliation Act (PRWOR), or Personal Responsibility Act, of 1996.^[1]

To accomplish this transformation of welfare from an emphasis on entitlements to an emphasis on work and self-sufficiency, states had to accomplish several things. One was to develop a new set of policies. Those policies had to reflect choices between a variety of competing approaches.^[2] Beyond this, the states and their local implementing agencies had to undertake important management challenges: designing a client process; determining necessary services; identifying service providers and creating a service delivery network; locating, allocating, and using financial resources; mediating the intergovernmental system; managing contracts; designing and using information systems; changing the organization; and measuring results.^[3] In the end, from the perspective of many observers, the most critical task was to align public welfare organizations with policy goals.^[2,4-8]

CONTEXT AND SHAPE OF REFORM

The contemporary approach to welfare policy, one that emphasizes work and self-sufficiency, grew out of dissatisfaction with the system of public support for the poor that emerged from the Social Security Act of 1935, revisions to that legislation, and new initiatives to expand access, support, and services for the poor in the War on Poverty of the 1960s.^[9] Central to the debate over welfare reform was Aid to Families

with Dependent Children (AFDC), a federal-state program to provide cash assistance to families unable to support themselves. Focused originally on women with children, the program evolved over a 30-year period beginning in the 1960s to offer aid to some two-parent families.

There is insufficient space here to trace the economic, social, and political changes that created pressure for transformation of the American welfare system. We can simply note several important developments.^[1,3,10,11] One was the movement of women into the workforce in recent decades. Another was the growth of two-earner families. A third was a substantial increase in births to unwed mothers. A fourth was the shift of AFDC from a program serving widows and their children to one serving divorced, deserted, and never-married women and their children. Finally, there was the substantial growth of AFDC rolls in the late 1960s and the 1970s, followed by another spurt in the late 1980s and early 1990s. These developments changed perceptions of the welfare system and created widespread support for change.

It is important to recognize that public welfare, or support for the poor, encompasses a broad range of public programs and mechanism for meeting the needs of low-income families and individuals. Indeed, even though AFDC provided the lightning rod for political discontent with welfare, other forms of assistance—such as Medicaid, the Earned Income Tax Credit, and food stamps—provided a far greater share of support and grew rapidly in the 1980s, while spending for AFDC declined in constant dollars.^[1] Despite that range of programs and spending, the reform of welfare focused substantially on AFDC, and that reform is the subject of this entry on welfare reform.

A quick review of three phases of the American welfare system demonstrates the dramatic nature of the changes while permitting illustration of the central policy and management challenges of each phase. The first phase is the AFDC program, as it existed in the 1970s. The second phase is the period of experimentation with welfare-to-work programs by the states in the 1980s, first under waivers to AFDC policies and then under provisions of the Family Support Act of 1988 with its Job Opportunities and Basic Skills program. The third phase is one of time limits on benefits and mandatory work under the Personal Responsibility

Act of 1996 and its Temporary Aid for Needy Families (TANF) program.

Under AFDC in the 1970s, the primary concern was to provide cash assistance, food stamps, and Medicaid to those in need while avoiding eligibility errors. These were entitlements, and the primary thrust was to make sure that those with needs were served. The second phase in the 1980s represented a response to concerns about welfare dependency and the need to help the poor become self-sufficient. Both before and after the Family Support Act, a mix of experiments sought to move recipients from welfare to work through some combination of education and training, work requirements, and supportive services. The principal pathway to this was not better jobs and wages, but mandates to participate in work programs or lose benefits. Sanctions were limited during this phase, services were a central focus, and most states in practice stressed voluntary participants. States tended to follow a human capital approach, which emphasized developing the knowledge, skills, and abilities of welfare clients so they could better compete in the labor market.

The third phase, under the Personal Responsibility Act and TANF, represented a dramatic departure from the welfare of the 1970s. Federal law placed a 5-year limit on lifetime benefits, required engagement in work activities within 2 years after beginning receipt of benefits, allowed states to disallow benefits to a mother who has a child while receiving welfare benefits, and gave states much greater leeway on how they manage their programs. Under TANF, the emphasis was on a work-first approach that involved getting welfare clients as quickly as possible into work activities. Rather than developing skills through education and training, the immediate goal was a job. Training and education can come later, after the client accepts personal responsibility for the well-being of family and self and experiences the success of work.

Thus, the three phases represent very different approaches to the needs of the welfare population based on very different theories of the sources of their problems and the ways to solve those problems.

POLICY CHOICES IN REFORM

In the quest to reform welfare and put in place a new system, policy makers and managers designing a reformed system had to make choices between competing conceptions of public priorities and the best way to achieve the ends of reform.

As states experimented with different versions of welfare-to-work programs, there were two basic differences in design choices.^[3,10,12] One involved the question of whether client participation would be voluntary or mandatory. Proponents of voluntary participation

believed that welfare participants wanted jobs and would take advantage of opportunities if they were made available. Proponents of mandatory participation believed that many welfare recipients did not want to work and would not do so unless it was required. The other basic design difference involved the question of whether the path to success would be found through work experience or education. Proponents of a human capital development approach point to the low educational attainments of many welfare recipients as the principal barrier to self-sufficiency. They believe that welfare recipients need basic job skills, education, and training if they are to leave the rolls and achieve self-sufficiency. Proponents of a work-first approach believe that many welfare recipients have not succeeded in educational activities in the past and are better served by programs that get them into a job so they can enjoy success.

The Personal Responsibility Act requires mandatory participation of a substantial portion of the welfare population. Many states have been even more insistent on mandatory participation. While many states emphasized education and development of literacy and other basic skills in their Job Opportunities and Basic Skills programs, most shifted quickly to a work-first orientation as they implemented TANF. Thus, the move was clearly to tougher requirements, more work, and the development of skills on the job instead of in the classroom.

States made a series of policy choices as they implemented the Personal Responsibility Act. This included, first, a set of choices that shaped participation in TANF, decisions about time limits, work exemptions, asset and income limits, sanctions, participation requirements, and family caps. States varied in their choices on these matters. Those choices made it more or less difficult for individuals to receive cash incentives and provided greater or lesser degrees of incentives for individuals to leave or fail to enter the welfare rolls.^[8,13] Second, states made a set of choices about how to pursue the work-first agenda. Gais, Nathan, Lurie, and Kaplan identified six variations of the work-first model.^[2] These included two motivational strategies, an employment service strategy, two adaptive service strategies, and a compound system.

One motivational strategy has work participation as its primary goal. States following this strategy try to decrease the cost of employment and increase its gains. The purpose is to increase the financial rewards of work through high earnings disregards and high benefit levels that allow participants to continue to receive benefits while working. A second motivational strategy has reduction of dependency as its primary goal. This is achieved by minimizing dependence on cash benefits. Wisconsin, for example, pays families only for the hours they participate in work and offers extensive career and employment services. An employment

services strategy uses employment, training, and education services to build job-related skills and place participants in jobs. One adaptive services strategy has work as its primary goal. States following this strategy offer a diverse array of services and give workers and administrators considerable discretion to tailor the services to the needs of the individual client. They place little emphasis on caseload reduction. Other adaptive service strategies have reduction in dependence as their primary goal. They offer a wide array of services and programs but place more emphasis on caseload reduction than states that make work the primary goal. Finally, Gais and his colleagues suggested that some states are not so easily classified.^[2] These compound systems exist because the states devolved a range of policy and administrative decisions to counties and cities.

MANAGING REFORM

These strategic approaches and the demands of implementation more generally required states to design a client process; determine the necessary services; identify service providers and create a service delivery network; locate, allocate, and use financial resources; mediate the intergovernmental system; manage contracts; change organizations; design and use information systems; and measure results. This was a complex set of management activities that reflected a strategic set of choices made by the states.

Designing a Client Process

The client process under welfare reform is a set of steps by which clients are directed to work activities, assess their skills and work readiness, have their eligibility determined, and are provided services. The sequencing of these steps is guided by program philosophy, strategic considerations, and individual client characteristics. In many states following a work-first approach, potential welfare recipients are sent to look for jobs before any effort is made to determine their eligibility for welfare benefits. If that search is not successful or the state is one that does not so heavily emphasize work first, two alternative processes might be initiated. One is assessment to determine the job readiness, work experience, education, and skills of the client. That information is useful in designing a set of activities to move the individual into a work role. The other is determination of eligibility for welfare benefits. In addition to this, steps are taken to determine what supplementary service the client needs to prepare for work or support engagement in the workforce.

Related to the client process design is the question of who will guide the client through the system. The

typical response to that has been the case manager. Case managers have replaced the traditional social worker whose primary responsibility was to see that the client received benefits, if eligible. Case managers encourage and assist the clients' move to work, arrange supportive services, and link the different parts of the system.

Determining the Necessary Services

Welfare-to-work programs generally provide a diverse set of services for clients. These services help clients seek and obtain jobs, acquire knowledge and skills employers require, meet family obligations, and reduce barriers to work. Clients may need and be eligible for a variety of services to address barriers to work or support employment. Services to prepare clients to work can include job readiness training, basic education, general equivalency degree preparation, and vocational training. Barriers that might be addressed include health and mental issues, substance abuse, disabilities, family problems, lack of transportation, and family situations. Employment services help place clients in jobs. Common services to support work include child-care and transportation. Case managers have to assess the needs of individual clients to determine the need for these services. Assuming the need is present, they help arrange the service.

Identifying Service Providers and Creating a Service Delivery Network

In the contemporary social service arena, the full range of services is seldom integrated in a hierarchical organization. Instead, services are provided by a complex array of public, not-for-profit, and private organizations. Networks of these organizations create diverse opportunities for service delivery but also introduce significant coordination problems. While it is possible for the welfare agency to deliver all of the needed services, it generally is not practical, effective, or politically feasible to do so.

Over the years, states have evolved in different directions in selecting service providers. They differ in how extensively services are integrated in the welfare agency or dispersed to other organizations. In some states, responsibility for welfare-to-work was transferred to a new organization or placed in an agency with a stronger work focus. States may assign certain services to particular public agencies. They may contract out services to not-for-profit or private organizations. With such diversity in service delivery, contract management and monitoring can become critical dimensions of the management of welfare reform.

A critical management challenge here is to decide which services the agency can best handle itself and which it should outsource. This includes deciding which mix of private entities, not-for-profit organizations, and public agencies will serve program objectives in a cost-effective manner. Another critical challenge in this network setting is to insure that the network of service providers works effectively. Numerous studies document that this is a different and more challenging task than managing an integrated, hierarchical organization.^[14–16]

Locating, Allocating, and Using Financial Resources

Financial resources were not a source of concern for the managers of welfare reform during the first five years of TANF. Because federal support for welfare was put on a block grant basis not tied to the number of recipients and the rolls were declining throughout that period, states did not face significant financial constraints. In fact, some states did not spend all of the funds available. Thus, there was less pressure to search for ways to leverage resources of other agencies, like employment and training programs or school systems. As the economy weakened in late 2001 and the rolls started to increase again, there were hints that states would face financial pressures. These pressures were brought on in part by significant expansions in Medicaid spending and questions about whether federal funding would be reduced when the program was reauthorized.

Mediating the Intergovernmental System

In many ways, the Personal Responsibility Act reduced federal control and gave the states greater leeway to design and operate their welfare programs. This was not completely the case, because some mandates were put in place, such as those that would limit the amount of time that most recipients could receive benefits supported by federal funds. In addition, states that sought to integrate their welfare program with other federally supported services had to maneuver through the shoals of intergovernmental complications. More important for some states was the need to make decisions about decentralization within their own bounds. To what extent would local communities or counties be allowed to assume control of the program? Some states, like North Carolina, gave significant policy responsibility to local communities. Others, like Kentucky, let sub-state regions and local communities play a significant role in program management.

Managing Contracts

With welfare agencies relying on contracts to acquire the services needed to make reform work, there is increased need to monitor the performance of contracting organizations. Contracting is a well-established tool of contemporary public management, but its effectiveness depends on several factors. The agency has to clearly specify what the contractor will provide. There has to be competition for contracting to be effective. Standards are needed to gauge performance. Finally, the agency has to be able to monitor the contract. While a clearly written contract with specific performance standards aids monitoring, the agency must create effective processes to do the monitoring.

Changing the Organization

As states moved to reform their welfare systems in the 1980s and 1990s and implement the Personal Responsibility Act after 1996, the single biggest challenge they typically faced was to change the culture of the organizations responsible for welfare reform. This meant changing the attitudes, orientations, and behaviors of case managers and other service providers in the system. Studies by Meyers, Glazer, and MacDonald,^[4] Myers, Riccucci, and Lurie,^[5] Corbett,^[17] Hagen and Lurie,^[18] Ewalt,^[19] and others pointed to the critical problems. As Corbett^[15] put it, “Ultimately, the culture of welfare is defined by what clients experience when they interact with welfare agencies and with people who run those agencies.”

As many managers and policy makers have discovered, changing the culture is difficult. The various tools of change have to be deployed effectively. In welfare reform, managers have sought to produce change by clearly communicating the new priorities of welfare and the message of change, altering client processing and office arrangements to reflect the new priorities, redefining the roles of service providers, training staff, and modeling the culture change themselves.^[3] Symbols are an important part of the culture change. Labeling is one way to convey symbolic importance—Work First is a prime example.

Information System Design and Use

Information is the lifeblood of the design, management, and assessment of welfare reform.^[20,21] Good information systems provide case managers the information they need to work effectively with clients. Mandates like time limits mean states have to be able to track individuals to insure that they do not exceed the limits. Performance management means that states have to be able to generate data on performance and

use it. Effective information management allows states to do such things as making accurate payments to service providers on the basis of performance, make programmatic adjustments in response to fluctuations and trends in performance levels, and report performance information to the U.S. Administration for Children and Families.

Designers of information systems have to determine who needs information, when they need it, and in what form they need it. They have to determine who can best provide the information and how it can be entered in the system. This knowledge facilitates the selection of hardware and design of software.

Measuring Results

Policy makers, managers, clients, citizens, and interest groups believe that it is critical to measure the success of welfare reform. To judge the effectiveness of reform, we have to know the goals and objectives and be able to measure their attainment. Additionally, performance measures are central to much contracting. Despite this, and in spite of modern public management practices reflected in the Government Performance and Results Act,^[22] states and the national government have been slow to gather systematic data on performance.

Most studies of the impact of welfare reform and the Personal Responsibilities Act have focused on changes in the number of individuals and families receiving welfare benefits.^[8,23–25] The rolls, of course, declined dramatically in the years following TANF, although they had begun to decline before its adoption.^[3] States have also generated a set of one-shot studies of welfare-leavers (individuals and families who have left the welfare rolls), but they have not created systematic measures over time of what has happened to those who left welfare or those who did not enter the rolls. Those one-shot studies of leavers have examined the employment status, income, and well-being of former TANF recipients.^[26] Typically, 50 to 60% of welfare leavers are working at the time of state surveys. For the first quarter after exit, 50 to 68% of leavers report working. Thirty-nine to 62% were working in the fourth quarter after exit. Forty percent or more reported unemployment in one or more quarters during the first year after leaving the rolls. In the states that conducted studies, earnings were typically well below the poverty line.

The critical point to note is that welfare reform can have a variety of consequences.^[27] It can reduce welfare rolls and welfare costs. It can increase self-sufficiency. It can increase employment. It can increase or decrease poverty. It can leave families better or worse off, take children out of risk or put them at risk. It can affect teen pregnancy rates and out-of-wedlock

birth rates. While it can do all of these things, little effort is being made to gather systematic data with respect to these consequences. Performance is largely unmeasured.

CONCLUSION

Reform has dramatically changed the character, symbolism, and management of welfare programs in America. Policy makers and managers have struggled to craft policies and programs that alter the outcomes of the system. In pursuit of the new goals, welfare managers have had to change organizations, manage effectively in networks, develop useful information systems, create new approaches to client processing, craft a new role of case management, select appropriate services, navigate the intergovernmental system, and allocate resources effectively. While the changes have been dramatic, the bottom line remains unclear. We know that the rolls have declined, but we have little systematic information about the impact of reform on the well-being of the poor.

REFERENCES

1. Weaver, R.K. *Ending Welfare as We Know It*; The Brookings Institution: Washington, DC, 2000.
2. Gais, T.; Nathan, R.P.; Lurie, I.; Kaplan, T. Implementation of the personal responsibility act of 1996. In *The New World of Welfare*; Blank, R.M., Haskins, R., Eds.; The Brookings Institution Press: Washington, DC, 2001; 35–69.
3. Jennings, E.T., Jr. Administration of welfare reform in the states. In *Handbook of State Government Administration*; Gargan, J.J., Ed.; Marcel Dekker, Inc.: New York, 1999; 483–518.
4. Meyers, M.K.; Glaser, B.; MacDonald, K. On the front lines of welfare delivery: are workers implementing policy reforms? *J. Policy Anal. Manage.* **1999**, *17*, 1–22(Winter).
5. Meyers, M.K.; Riccucci, N.M.; Lurie, I. Achieving goal congruence in complex environments: the case of welfare reform. *J. Public Adm. Res. Theory* **2001**, *11*, 165–201(April, 2001).
6. Mead, L.M. Welfare policy: The administrative frontier. *J. Policy Anal. Manage.* **1996**, *15*, 587–600(Fall).
7. Mead, L.M. Optimizing JOBS: evaluation versus administration. *Public Adm. Rev.* **1997**, *57*, 113–123(March/April).
8. Jennings, E.T.; Ewalt, J.A.G. Driving caseloads down: welfare policy choices and administrative action in the states. In *Governance and Performance: New Perspectives*; Heinrich, C., Lynn, L.E., Jr., Eds.; Georgetown University Press: Washington, DC, 2000; 109–139.

9. Cammisa, A.M. *From Rhetoric to Reform? Welfare Policy in American Politics*; Westview Press: Boulder, CO, 1998.
10. Mead, L.M. *The New Politics of Poverty: The Non-working Poor in America*; Basic Books: New York, 1992.
11. Bane, M.J.; Ellwood, D. *Welfare Realities*; Harvard University Press: Cambridge, 1994.
12. Nathan, R.P. *Turning Promises into Performance: The Management Challenge of Implementing Workfare*; Columbia University Press: New York, 1993.
13. Gallagher, L.J.; Gallagher, M.; Perese, K.; Schreiber, S.; Watson, K. *One Year After Federal Welfare Reform: A Description of State Temporary Assistance for Needy Families (TANF) Decisions as of October 1997 Preliminary Edition*; The Urban Institute: Washington, DC, 1998, Accessed at <http://newfederalism.urban.org/temp/1yeara.html>, March, 1998.
14. O'Toole, L., Jr. Treating networks seriously: Practical and reserch-based agendas in public administration. *Public Adm. Rev.* **1997**, 57, 45–52(January/February).
15. Provan, K.G.; Milward, H.B. How networks are governed. In *Governance and Performance: New Perspectives*; Heinrich, C., Lynn, L.E., Jr., Eds.; Georgetown University Press: Washington, DC, 2000; 238–262.
16. Jennings, E.T., Jr.; Krane, D. Coordination and welfare reform: the quest for the philosopher's stone. *Public Adm. Rev.* **1994**, 54, 341–348(July/August).
17. Corbett, T.J. Changing the culture of welfare. *Focus* **1994**, 16(Winter), Accessed at gopher://gopher.ssc.wisc.edu/00/irpgopher/publications/focus/16.2a/changing_gfr, September, 1998.
18. Hagen, J.L.; Lurie, I. Implementing JOBS: from the rose garden to reality. *Social Work* **1994**, 40 (4), 230–238.
19. Ewalt, J.A.G. *An Analysis of the Job Opportunities and Basic Skills Program in Kentucky: Determinants of Component Choice*; Dissertation; University of Kentucky: Lexington, KY, 1998.
20. Nathan, R.P.; Ragan, M. *Federalism and the Challenges of Improving Information Systems for Human Services*; The Nelson A. Rockefeller Institute of Government: Albany, NY, 2001, Accessed at <http://www.rockinst.org/>, December, 2001.
21. U.S. General Accounting Office!01). In *Welfare Reform—Improving State Automated Systems Requires Federal Effort*; 2000Washington, DC; GAO/HEHS-00-48.
22. Newcomer, K. *Using Performance Measurement to Strengthen Public and Nonprofit Programs*; Jossey-Bass Publishers: San Francisco, 1997, Editor.
23. Council of economic advisers. In *Technical Report: Explaining the Decline in Welfare Receipt, 1993–1996*; 1997, Washington, DC.
24. Council of economic advisers. *The Effects of Welfare Policy and the Economic Expansion on Welfare Caseloads: An Update*; 2000, Washington, DC.
25. Martini, A.; Wiseman, M. *Explaining the Recent Decline in Welfare Caseloads: Is the Council of Economic Advisers Right?* The Urban Institute: Washington, DC, 2000, July.
26. National campaign for jobs and income support. In *Leaving Welfare, Left Behind: Employment Status, Income, and Well-Being of Former TANF Recipients*; 2001, Washington, DC; Accessed at <http://www.nationalcampaign.org/incomesupport/LeaversRelease9.25.htm>, January 2002.
27. WELPLAN. *Welfare Reform: How Will We Know If it Works? Midwest Peer Assistance Network, Family Impact Seminar*; 1998, Washington, DC; Accessed at <http://www.welfareinfo.org>, July, 1998.

Welfare State

Paul B. Kirby

*University of Massachusetts Medical School, Center for Health Policy and Research,
University of Massachusetts, Worcester, Massachusetts, U.S.A.*

INTRODUCTION

At its most basic level, the term “welfare state” refers to a state’s use of its policy-making authority to guarantee some minimal level of material well-being to its citizens. Welfare states seek to attain this goal through various means, such as direct redistribution through progressive taxation, state-administered social insurance programs, regulation (of wage levels, working conditions, etc.) and state provision of services such as health care. Although the welfare state is generally understood to be a defining characteristic of Western democracies in the period after World War II, its origins actually date to the late nineteenth century. Further, aspects of the welfare state model have been adopted in non-Western countries.

This article discusses the welfare state phenomenon from three perspectives that are of particular interest to students of public policy and public administration. The first section discusses the normative rationales for the creation and expansion of the welfare state put forth by prominent political and scholarly supporters. Second, scholarly work seeking to provide causal explanations for the emergence of welfare states generally, and to account for the wide variety of specific modalities of the welfare state, is reviewed. Finally, the more recent turn to evaluation of the welfare state’s impact is discussed. That is, scholars have extensively analyzed the effects of welfare policies on a range of social and economic indicators, and their findings have had considerable influence over ongoing policy debates.

RATIONALES FOR THE WELFARE STATE

The Industrial Revolution entailed a dramatic dislocation of traditional social relations in Western countries. Social critics argued that industrial wage labor, fundamentally detached from any direct connection to the reproduction of daily existence, actually worsened the living conditions of the lower classes. Industrialization, in this view, removed the largely self-sufficient peasantry from its traditional agricultural pursuits, offering only repetitive, dehumanizing factory labor and utter dependence on cash wages in return. As a historical matter, this view is not entirely satisfactory. There is

ample evidence that rural workers began to participate in the cash economy well before widespread industrialization, through paid agricultural labor and small-scale production of intermediate goods, such as textiles, for sale (literal “cottage industries”).

Nonetheless, this critical view was broadly accurate and compelling. Generations of nineteenth- and twentieth-century social, political, and economic theorists struggled with the question of how to improve the lives of the workers. As the pace of industrialization accelerated in the later nineteenth century, concern over increasing social inequality and economic insecurity generated widespread agitation for reform. Prominent among the various rationales for reform were fears that loss of social cohesion would inevitably result from the impoverishment of workers; social critics wanted to create a harmony of interests among the various classes in society.

The nature of industrial production created new risks for workers, such as the increased potential for disabling accidents, and hardship due to unemployment and old age. Total reliance on wage earnings made such events catastrophic. The earliest welfare state policies, accordingly, were social insurance programs designed to insulate workers from these risks. Advocates insisted that the state would have to administer, and subsidize through general tax revenues, these programs, which private insurance markets would not provide. The state would also have to mandate universal participation in social insurance to guard against the problem of adverse selection.^[1]

In the 1930s, as the great depression demonstrated the inadequacy of the early social insurance programs in dealing with mass unemployment and destitution, supporters of the welfare state began to argue for a much more assertive program of expanded social insurance, state intervention in labor markets to bring about full employment, and direct redistribution through cash transfers where those measures prove insufficient. In the United States, President Roosevelt created a Committee on Economic Security, charged with creating policy proposals to send to Congress. In their 1935 report, the committee argued that, with wages falling in depression conditions, economic insecurity had begun to affect even full-time workers. In response, the committee called for “(a) program of economic security” having “as its primary aim the

assurance of an adequate income to each human being in childhood, youth, middle age, or old age—in sickness or in health.”^[2] Although the goals expressed in the committee’s report anticipated a much more extensive welfare system than that which the United States would later develop, the rhetoric of the report mirrored the widespread belief that economic stability would ultimately have to rest on vigorous government action.

As the welfare state continued to expand and become entrenched in the postwar period, arguments in support of it became more abstract. Marshall, for example, famously argued that the welfare state should be understood as the logical end of a three-stage historical process by which the concept of “citizenship” developed in the Western democracies.^[3] In his view, *civil rights*, or the idea of equality before the law, were generally established in the eighteenth century, laying the foundation for *political rights*, that is, the right of all citizens to exercise an equal share of political power through the vote, which prevailed in the West by the early twentieth century. With the civil and political spheres of citizenship established, the natural next step for citizenship was to extend to *social rights*.^[3] Marshall defined this broadly as “the whole range from the right to a modicum of economic welfare and security to the right to . . . live the life of a civilized being according to the standards prevailing in the society.”^[3]

ORIGINS AND DEVELOPMENT OF THE WELFARE STATE

In recent decades, scholars have turned their attention to the question of how the welfare state came to exist, both as a generic phenomenon and in its wide variety of manifestations. Early attempts at causal explanation of the welfare state phenomenon were marked by a tendency toward teleological “laws of development.” In this perspective, the welfare state is seen as the natural and necessary outcome of capitalist, industrial development.

The process of welfare state development, in this view, appears as a response of the early critiques of industrial capitalism made by welfare state proponents. As industrialization upset traditional societies, the resulting social problems led to the welfare state as a functional necessity (see, e.g., Refs.^[4,5]). At the same time, industrial development facilitated the emergence of the welfare state in two critical ways: 1) by giving rise to bureaucracy as the dominant organizational form, in politics as well as economics; and 2) by generating enough surplus wealth that redistribution and capital accumulation would no longer be seen as mutually exclusive (see Ref.^[6] for a discussion and critique of this theory). Variation between existing

welfare states is attributed primarily to the “economic level and its demographic and bureaucratic correlates” present in different countries, where these “correlates” are an aging population and an increasingly entrenched, self-perpetuating welfare bureaucracy.^[7]

Although the industrialization theory is an intuitively satisfying explanation for the existence of the welfare state in general, it fails to account for the significant variation in actual welfare states. A second broad strain of thought attempts to overcome this weakness by focusing on the political and electoral strength of the working class, and labor-based political parties, in explaining differences in the size, scope, and character of welfare states. Korpi, for example, analyzed the electoral power of the working class (as exercised through labor-based parties) along two dimensions: 1) mobilization, as measured by the percentage of the labor force that is unionized and the percentage of the electorate voting for labor-based, left-wing parties; and 2) control over government, as measured by representation of these parties in cabinets and parliaments.^[8] He found that countries in which the working class is highly mobilized and labor-based parties have governed for long periods of time, welfare states are at their most extensive and inequality is most dramatically reduced, as in Sweden and Norway.^[8]

Other scholars in this tradition have emphasized the ability of labor-based parties to form alliances with other classes, in particular, with farmers. Esping-Andersen, for example, argued that Social Democratic forces in the Scandinavian countries overcame opposition to their welfare state policies by creating broad coalitions with farmers, gaining their support in exchange for agricultural price subsidies, and other policies addressing their needs.^[9]

The discussion of the working-class strength thesis is at its strongest when explaining the emergence of large, comprehensive, and solidly institutionalized welfare states in the Scandinavian countries. Similarly, the industrialization theory presents a reasonably compelling explanation for the greater prevalence of welfare states among the rich, highly developed countries. But neither perspective can adequately account for the most significant outlier in crossnational comparisons of the welfare state—the United States. As many analysts have noted, the U.S. welfare state has lower benefit levels, is less extensive in its coverage, and intervenes less in the market economy than its European counterparts. Despite its great wealth, rapid industrialization, and abundance of eloquent social reform advocates, the United States did not develop anything comparable to the broad package of social insurance programs developed in Britain during the early twentieth century, much less the combination of universal, nonmeans-tested social benefits and Keynesian commitment to full employment that emerged in Sweden

during the 1930s and 1940s. This “American exceptionalism” has inspired a great deal of analysis in recent decades, leading to a school of thought that emphasizes the role of state institutional structures in shaping welfare policy regimes.

Most closely associated with Skocpol and colleagues (see, e.g., Refs.^[10–13]), this institutionalist theory charges both the industrialization and the working-class strength theories with being overly deterministic. Instead of seeing the welfare state as the natural outgrowth of social forces, this perspective sees state structures determining the range of policy responses to social and economic problems that a particular state might select. For example, both Britain and the United States confronted the (supposedly) objective “need” for social insurance programs to protect industrial workers from the insecurities of old age, sickness and disability, and unemployment. Yet, Britain developed a comprehensive array of such programs during the Liberal reform period of the early twentieth century, while the United States did not develop truly national social insurance of any kind until well into the Great Depression, and then only after considerable hesitation.

The institutionalists find the explanation for this discrepancy in the unique character of American political institutions, as compared with those found in Western Europe. The fundamental cause, in this view, is the weak administrative capacity of the national government in the United States, as compared with many European countries. One obvious cause of this weakness is the federal structure of the U.S. government, which for most of the nation’s history limited the power of national government institutions.^[14] Until the 1930s, state and local governments combined were larger than the federal government, both in terms of spending and in terms of exercise of regulatory power. The executive branch of the national government remained relatively weak until it was strengthened by the twin challenges of the Great Depression and World War II. Meanwhile, the U.S. Congress, influenced heavily by local concerns, had neither the ability nor the incentive to formulate any sort of national social welfare policy.^[10]

The deeper causes for the missing American welfare state posited by the institutionalist view are more complex, however. Somewhat ironically, the early extension of the right to vote to the white male working class set in motion a process of what these theorists call *state formation* that would later hinder the development of a welfare state. Although the question of male working-class suffrage caused bitter political struggles in most of Western Europe, white male American workers obtained the vote by the middle of the nineteenth century, with relatively little controversy. Because the expansion of suffrage preceded the

bureaucratization of state institutions, American political parties competed for votes on the basis of patronage, or the exchange of government offices and resources for votes. In the European countries, by contrast, universal male suffrage came much later, with economic restrictions on the franchise in place until the early twentieth century. This late democratization came after European state institutions had become firmly bureaucratic, with access to government jobs and funds strictly controlled by a professionalized, non-political civil administration. Thus, political parties had to compete for the votes of newly enfranchised workingmen on the basis of policy programs, rather than patronage.^[11]

Whether a country’s political parties were programmatic or patronage based had a critical impact on the formation of welfare states. The legacy of patronage is most clearly seen in the proliferation of government pensions for veterans of the Union Army (and their survivors) in the American Civil War. Eligibility for these pensions, which initially were limited to demonstrably disabled veterans, was gradually liberalized to the point that any veteran of age 62 years or older could receive benefits.^[12] But instead of signaling the emergence of an American protowelfare state, the Civil War pensions became a means for the Republican Party to secure its electoral dominance in the postwar period. By assisting constituents with their pension applications, and by fervently advocating liberalization of eligibility rules, Republican members of Congress bolstered their re-election prospects. Because Confederate veterans were not eligible, southern Democratic politicians could not take similar advantage of the patronage opportunities the pensions offered.^[12]

As a result, early twentieth-century social reformers in the United States, who were closely associated with the largely middle- and upper-class Progressive movement, were more concerned with combating the corruption endemic in patronage electoral democracy than with expanding the power of the national government to address the social dysfunctions of an emerging industrial capitalist economy. Progressives, consequently, opposed the extension of the Civil War pensions into a true national system of old-age pensions, preferring instead to press for various social insurance programs at the state level, to be financed through payroll contributions from workers rather than through general government revenues. The lack of support by Progressive reformers for non-contributory old-age pensions soured the relationship between Progressives and the labor movement, headed by the increasingly powerful American Federation of Labor (AFL) umbrella group. The AFL refused to support the Progressives’ call for mandatory, contribution-based social insurance programs; in the absence of broad, crossclass support, few such programs were

established.^[12] In contrast, British reformers, centered in the governing Liberal Party, enacted a non-contributory pension system in 1908, which helped them to maintain their cooperative relationship with the emergent Labour Party and its union supporters. This cooperation was essential to the enactment of contributory unemployment and health insurance programs under the National Insurance Act of 1911.^[13]

Only when faced with the severe crisis of the Great Depression was the United States able to create national social insurance programs covering old age, unemployment, and (later) disability. Compared with their European counterparts, America's social insurance programs have remained limited in terms of coverage and benefit levels, particularly in the areas of health care and disability. In contrast to most of Western Europe, the United States has only developed meager programs of direct social assistance to working-age adults. Such programs are generally available only to single mothers and offer very low benefit levels. Other hallmarks of European (in particular, Scandinavian) welfare states, such as active intervention in labor markets to ensure full employment, have received little, if any, serious consideration in the American political system. The institutionalist perspective is useful in understanding this lack of welfare state development in the United States.

The theories discussed above apply, of course, mainly to the economically advanced Western democracies. In recent years, a number of scholars have turned their attention to welfare states in non-Western, particularly East Asian, countries. Although the modes of operation of non-Western welfare states do not fit the Western model, many East Asian countries have, in functional terms, what may be considered welfare states. For example, Japan and South Korea have extensive social insurance programs, covering health care, disability, and pensions. However, these programs involve more regulation than direct expenditure by the state, which tends to act as a coordinator of insurance plans run by private companies, mutual-aid societies, and local authorities. Consequently, by standard measures such as government expenditure levels, these programs may not be recognized by Western commentators as constituting a welfare state.^[15] Other analysts have viewed welfare policies as an adjunct to Asian states' overall policies of state-guided economic development (see Ref.^[16] for the application of this view to Japan). Finally, some authors argue that East Asian countries have in fact followed a pattern of welfare state development that is broadly similar to the European model.^[17] As East Asian countries continue to develop economically, and as capitalism emerges in formerly Communist Central and Eastern Europe, the study of newly developing welfare states will undoubtedly expand.

EVALUATING THE WELFARE STATE

Although the question of the historical origins of the welfare state and its variations has primarily been of interest to academic social scientists, the issue of the welfare state's effects on economies and individuals has attracted a broader audience. Studies of welfare state effects have looked at two sets of variables. Some have concentrated on strictly economic variables, both microeconomic (e.g., individual labor market behavior) and macroeconomic (e.g., fiscal limitations on the welfare state and affects on the competitiveness of national economies). Others have assessed the influence of the welfare state on personal behaviors, such as decisions to marry and give birth. This division is not a strict one, however, as some studies have drawn connections between family formation and labor market behavior.

Concerned with the dramatic slowing of economic growth in the Western industrialized countries that began in the 1970s, economists have become highly critical of the welfare state's impact on growth, international competitiveness, and employment. These concerns are especially sharp in the Western European context, in which more extensive welfare states and more pronounced economic slowdowns have been combined. Many commentators, both academic economists and business journalists, have argued that economic growth in Western Europe has lagged behind that of the United States since the 1970s, largely due to the impact of the welfare state. The critical point in this critique is that the welfare state, by protecting workers from the negative consequences of economic change, hinders the process of economic adjustment to new conditions. Generous unemployment benefit programs, labor market protections in the form of legal restrictions on the ability of companies to fire workers, and other features of European welfare states have, in this view, slowed the European transition from economies heavily reliant on manufacturing to a more service-based model.

Indeed, the term "sclerosis" has often been used to describe the difficulty of Western European economies to adjust to new international conditions, resulting in persistently high rates of unemployment. This "sclerosis" has several components. Labor costs, particularly in the manufacturing sector, are among the highest in the world, due to the cost of mandated employer contributions to social insurance programs and the role of collective bargaining in setting wages. High labor costs, in combination with employment protection, began to cause dramatically slower growth in manufacturing sector employment, leading to high levels of joblessness, especially among the young. Meanwhile, the welfare state reduced incentives for economic actors—firms as well as workers—to adjust

to changing economic conditions. High wages, resulting from social insurance costs and legal regulation of minimum wages, prevented economies from shifting toward lower-productivity, lower-pay service industries, while essentially unlimited benefits reduced the incentive for the unemployed to seek jobs outside manufacturing.^[18,19] A deeper concern is that the persistence of this situation will alter long-term social norms about the importance of work and avoidance of relying on welfare benefits.^[20] Such changes in social norms threaten both the economic growth and the long-term viability of the welfare state itself.

Although the above discussion applies most obviously to Western Europe, the debate over public old-age pensions has assumed great importance on both sides of the Atlantic. There is little mainstream controversy surrounding the basic desirability of pensions, yet there is great concern over funding them. The funding problem stems from the fact that most public old-age pensions operate on a “pay-as-you-go” basis, in which the payroll taxes of active workers pay for retirees’ pensions. At present, the value of payroll taxes collected from workers is sufficient to fund current pension obligations in most industrialized countries; in the United States, much more tax is collected than presently needed, leading to a large surplus in the social security system. However, demographic trends common to all industrialized countries threaten this stable situation. Because birthrates in the wealthy countries have fallen so sharply, the ratio of workers to pensioners is projected to decrease dramatically during the first-half of the twenty-first century. Pension obligations could lead to a fiscal crisis across the industrialized world.

The economic projections involved are highly sensitive to changes in baseline assumptions (particularly in economic growth rates); however, the pension issue has attracted much public attention. Proposals in the United States to reform the social security system (which is actually in a much better fiscal position than public pension programs in most of Europe and in Japan) generally advocate the diversion of some portion of current payroll taxes to higher-return investments in equities, rather than, as at present, government bonds.^[21] Advocates of this change argue that it will increase the pool of available funds for future retirees; most proposed plans would also raise the minimum age at which retirees can begin receiving benefits. Proposed reforms of European public pension plans are similar to those in the United States. Although preliminary reforms have begun to narrow the projected gaps in future pension funding in a number of European countries, further changes will be needed to meet the needs of rapidly aging societies.^[22]

Finally, a great many critics, particularly in the United States, have argued that the welfare state has

severe negative effects on the behavior of recipients, in the critical areas of work and family formation. The focus of these critics is on public assistance programs for working-age adults, primarily single mothers (these programs, and not others such as social security, comprise what is generally thought of as “welfare” by the American public). Murray maintained that the expansion of welfare benefits during the 1960s actually worsened the lives of those it was intended to help—the poor and, especially, racial minorities.^[23] He asserted that programs such as Aid to Families with Dependent Children (AFDC) created perverse incentives for recipients, encouraging women to have children outside of marriage. More broadly, Murray argued that the expansion of public assistance reduced the stigma attached to receiving benefits instead of working, causing the poor to view welfare as a legitimate alternative to low-wage work.

Although Murray’s conclusions have been hotly contested (see, for example, Ref.^[24]), they have had great impact, both among other critics and among policy makers. The idea that public assistance programs have created a “culture of dependency” among recipients has become commonplace, leading many to call for welfare policies to reverse this trend. Mead, for example, argued that welfare policy in the United States should put less emphasis on the right to benefits as an entitlement, and instead focus more on responsibilities of recipients.^[25] In particular, he called for public assistance programs to force recipients to participate in training, job search, and “workfare” activities. More recently, Mead advocated a “new paternalism” in welfare policy, in which public assistance recipients’ progress toward employment goals is closely monitored by caseworkers.^[26] The criticisms of Murray, Mead, and other authors strongly influenced the 1996 Welfare Reform legislation, which ended the federal entitlement to AFDC benefits, replacing this with Temporary Assistance for Needy Families, a time-limited program explicitly designed to move recipients into jobs as quickly as possible.

The critiques discussed above have had obvious effects in the political arena, in both the United States and in the Western Europe. Although there have been differences in emphasis—for example, American critics of the welfare state have focused on the behavior of recipients, whereas Europeans have been more concerned about overall economic outcomes—the end result on both sides of the Atlantic has been a reconsideration of the scope, and even the desirability, of many aspects of the welfare state. The welfare state, which in the early postwar years commanded a broad favorable consensus across the industrialized world, is now the object of serious political contention.

REFERENCES

1. Churchill, W. Speech on National Insurance Bill. In *Hansard's Parliamentary Debates*; series: House of Commons, Fifth Series; 25 May, 1911; cols 495–510.
2. Committee on Economic Security. Report of the Committee on Economic Security. Available on Social Security Administration website, at <http://www.ssa.gov/history/reports/ces/ces5.html>.
3. Marshall, T.H. Citizenship and social class. In *Class, Citizenship, and Social Development: Essays by; Marshall, T.H., Ed.*; Doubleday: Garden City, NY, 1964.
4. Wilensky, H.L.; Lebeaux, C.N. *Industrial Society and Social Welfare*; The Free Press: New York, 1965.
5. Flora, P.; Alber, J. Modernization, democratization and the development of welfare states in Western Europe. In *The Development of Welfare States in Europe and America*; Flora, P., Heidenheimer, A., Eds.; Transaction Books: New Brunswick, NJ, 1981.
6. Esping-Andersen, G. *The Three Worlds of Welfare Capitalism*; Princeton University Press: Princeton, NJ, 1990.
7. Wilensky, H.L. *The Welfare State and Equality: Structural and Ideological Roots of Public Expenditure*; University of California Press: Berkeley, 1975.
8. Korpi, W. Social policy and distributional conflict in the capitalist democracies. A preliminary comparative framework. *West Eur. Polit.* **1980**, 3 (3), 296–316.
9. Esping-Andersen, G. *Politics Against Markets*; Princeton University Press: Princeton, NJ, 1985.
10. Weir, M.; Orloff, A.S.; Skocpol, T. Understanding American social politics. In *The Politics of Social Policy*; Weir, M., Orloff, A.S., Skocpol, T., Eds.; Princeton University Press: Princeton, NJ, 1988.
11. Orloff, A.S. The political origins of America's belated welfare state. In *The Politics of Social Policy in the United States*; Weir, M., Orloff, A.S., Skocpol, T., Eds.; Princeton University Press: Princeton, NJ, 1988.
12. Skocpol, T. *Protecting Soldiers and Mothers: The Political Origins of Social Policy in the United States*; Harvard University Press: Cambridge, MA, 1992.
13. Orloff, A.S.; Skocpol, T. Why not equal protection? Explaining the politics of public social spending in Britain, 1900–1911, and the United States, 1880s–1920. *Am. Sociol. Rev.* **1984**, 49 (6), 726–750.
14. Lowi, T. Why is there no socialism in the United States? A federal analysis. *Int. Polit. Sci. Rev.* **1984**, 5 (4), 369–380.
15. Kwon, H.-J. Beyond European welfare regimes: comparative perspectives on East Asian welfare systems. *J. Soc. Policy* **1997**, 26 (4), 467–484.
16. Peng, I. A fresh look at the Japanese welfare state. *Soc. Policy Admin* **2000**, 34 (1), 87–115.
17. Hort, S.E.O.; Kuhnle, S. The coming of East and South-East Asian welfare states. *J. Eur. Soc. Policy* **2000**, 10 (2), 162–184.
18. Organization for Economic Cooperation and Development. *The OECD Jobs Study: Evidence and Explanations*; OECD: Paris, 1994.
19. Ljungqvist, L.; Sargent, T.J. The European unemployment dilemma. *J. Polit. Economy* **1998**, 106 (3), 514–550.
20. Lindbeck, A. *Hazardous Welfare-State Dynamics*; Industriens Utredningsinstitut: Stockholm, 1995.
21. Advisory Council On Social Security Report of the 1994–1996. Advisory Council on Social Security. Available on Social Security Administration website, at <http://www.ssa.gov/history/reports/adccouncil/report/toc.htm>.
22. Organization for Economic Cooperation and Development. *OECD Economic Outlook*; series: *Number*; OECD: Paris, 2001; Vol. 69.
23. Murray, C. *Losing Ground: American Social Policy, 1950–1980*; Basic Books: New York, 1984.
24. Wilson, W.J. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*; University of Chicago Press: Chicago, 1987.
25. Mead, L. *Beyond Entitlement: The Social Obligations of Citizenship*; Free Press: New York, 1986.
26. Mead, L. Welfare employment. In *The New Paternalism: Supervisory Approaches to Poverty*; Mead, L., Ed.; Brookings Institution Press: Washington, DC, 1997.

Whistleblowing, Corporate, and Public Policy

Michael W. Austin

*Department of Philosophy and Religion, Eastern Kentucky University,
Richmond, Kentucky, U.S.A.*

Michael A. Harper

University of Arkansas, Fayetteville, Arkansas, U.S.A.

INTRODUCTION

The phenomenon of whistle-blowing may stir up images of the uncovering of unethical practices in various businesses, but in a loose sense, it has occurred since the mid-1800s within corporate organizational structures. Generally, it is someone telling the truth about wrong actions taken by an entity to some agency, which can bring to bear some type of ramifications on the entity at hand. To better understand what whistle-blowing is, we consider the salient features of a few contemporary definitions, the context in which it occurs, and the ethical tensions involved. We then turn to important public policy questions related to whistle-blowing.

SALIENT DEFINITIONS

Jubb^[1] offers an analysis of contemporary definitions and one's definition of whistle-blowing. Jubb's definition is: "Whistle-blowing is a deliberate non-obligatory act of disclosure, which gets onto public record and is made by a person who has or had privileged access to data or information of an organization, about non-trivial illegality or other wrongdoing whether actual, suspected, or anticipated, which implicates and is under the control of that organization, to an external entity having potential to rectify the wrongdoing."

Jubb's definition flows from the understanding that whistle-blowing at the core presents ethical dilemmas for the agent involved in the action. First, a dilemma emerges for the agent with respect to the organization's wrongdoing in light of an organization's established values and/or the agent's personal values. If the organization decides to assent to the wrongdoing over its established values, the dilemma will pit the organization as a whole against the personal values of the agent. This is viewed as an explicit form of dissent within the organizational structure. For Jubb, the central dilemma that an agent faces hinges on one's loyalties to the organization versus loyalties to oneself and/or other entities. This dilemma is set in a context

of loyalty that is often implicit within the structure of an organization that hires a person to become an employee with access to the confidentiality surrounding ideas, information, and products. Thus the agent must pit one's loyalties to an organization that has provided employment against loyalties to oneself, colleagues, the profession, family, or the general public.

The last crucial element that must be underscored in Jubb's definition is that the wrongdoing must be disclosed to an external entity; otherwise, no ethical dilemma will arise for the agent. If the agent is not put in a position to choose to override the loyalty created by being able to access confidential materials as an employee, then the dilemma fails to materialize; thus the conditions seen as sufficient for the definition also are lacking. The formulation of a loyalty dilemma is a vital component for Jubb's definition and is the salient feature that requires emphasis.

Our second definition is expressed by Davis^[2] and shifts the central feature away from loyalty dilemmas toward understanding an agent's complicity in organizational activities that may involve wrongdoing. Davis' definition, which he calls a theory, is as follows: "You are morally required to reveal what you know to the public (or to a suitable agent or representative of it) when:

- (C1) What you will reveal derives from your work for an organization
- (C2) You are a voluntary member of that organization
- (C3) You believe that the organization, although legitimate, is engaged in a serious moral wrong
- (C4) You believe that your work for that organization will contribute (more or less directly) to the wrong if (but not only if) you do not publicly reveal what you know
- (C5) You are justified in beliefs (C3) and (C4)
- (C6) Beliefs C3 and C4 are true."

Davis sees his definition of whistle-blowing as differing primarily and most importantly in (C4). Here,

Davis expresses what he means by complicity. Complicity is important because it is the feature that ties the agent's work with the internal actions of the organization. An agent in an organization will remain an accomplice if he/she overlooks and remains with the organization and even if he/she quits. Whistle-blowing occurs when the agent is put in a position in which taking this revelatory action is the means to break these ties. The action of going public with the wrongdoing prevents complicity on the agent's part. Davis recognizes the issue of loyalty, but he sees complicity as a key feature encountered in many and various whistle-blowing cases, whereas on the other hand, Jubb also recognizes complicity but places emphasis on the need for an ethical dilemma to be formed for whistle-blowing to occur. These are two novel definitions of whistle-blowing that attempt to point out salient features of whistle-blowing. This is not an exhaustive list, but does hone in on important aspects of whistle-blowing that help us understand what it is.

TENSIONS FACING THE WHISTLE-BLOWER

Without reflection, it may seem like the whistle-blower sees the wrongdoings within one's organization and does what is right, but the tensions faced by the whistle-blower are often more substantive than just seeing the wrong and then doing the right thing. Any time issues that involve loyalty and means to a livelihood are at stake, moral and practical tensions will likely follow.

The seemingly most obvious aspects of whistle-blowing involve the moral criteria being brought to bear on the wrongdoing and the practical ramifications that might occur from the "disloyal" action. Nonetheless, Jensen^[3] highlights five substantive issues that reveal the deeper moral tensions. Generally, these moral tensions are as follows:

- Moral obligations to organization and colleagues
- Ethical standards of, and obligations to, a profession
- Adverse effects of the action on family or primary groups
- Moral obligations toward oneself
- Ethical obligations toward the general public
- Effects on bedrock values.

When considering moral tensions, it is obvious that some very practical matters seem to fall out, such as the possibility of financial hardships that could occur for the employer and employee. Further thought on these moral tensions can reveal more practical tensions. First, if loyalty is established between the employer and the employee at hiring, then that loyalty is placed under stress when whistle-blowing occurs. This can result in the suspension of the benefits the

company provides the employee and vice versa. So not only are the whistle-blower's personal interests affected, but many times the agent also considers the various ramifications that can result for an organization. Those damages can be significant—ranging from disruption of normal business functions, to loss of profits, to damaged internal and external credibility, to moral tensions that emerge for the agent's colleagues in that they can be adversely affected by the act. Of course, this may be superseded by the idea that long-term benefits will result, but it is not guaranteed. Second, a tension will exist between the need for an organization to make a profit and the credibility that certain professions may need when looked on by the public, such as medicine or law. Third, there will be potential for tensions from within and without the family of the agent, with the potential for increased publicity, financial hardship, loss of friends, and cruelty toward family members, which need to be weighed. Fourth, the agent should consider the personal risk taken in such actions, which may result in loss of a job, respect, confidence from others, and in oneself, as well as difficulties in future employment. Fifth, public reaction could be mixed regarding the revelation of the wrongdoing. The agent should consider the short-term and long-term ramifications of the act on the public. Lastly, Jensen considers that some deeply held values might encounter friction. For example, the difficult choice and potential consequences of telling the truth in a whistle-blowing scenario may be chosen over the friendship of a peer within the organization. Again, the issue of loyalty is a centerpiece for the culmination of moral tensions, and the practical ramifications of being implicated in wrongdoings within the company drive many of the practical tensions that accompany whistle-blowing.

PUBLIC POLICY AND WHISTLE-BLOWING

The ultimate goal with respect to public policy and whistle-blowing should be to eliminate as much as possible the need for engaging in it, by reducing its legitimate causes.^[4,5] Given the moral and legal failures of business scandals (e.g., Enron and MCI Worldcom), we must do all that we can, both institutionally and legally, to encourage the practice when it is justifiable or obligatory. Having good institutional procedures and policies in place, which reduce the need to go public or to the court, is highly desirable. Until then, laws (or regulations) could be formulated so that the potential whistle-blower is required by law (or regulations) to do so, which could remove some of the burden from the individual, as could anonymous hotlines, ombudsmen, and government inspectors general. For example, the Toxic Substances Control Act of 1977 requires

companies that produce chemicals to direct their employees to report chemicals that constitute a substantial risk to people's health or to the environment.

When is whistle-blowing morally justifiable and, hence, legitimate? DeGeorge^[6] has addressed this issue, and offers three conditions that must be met for whistle-blowing to be morally justifiable. First, the issue must involve a serious harm to the public or to particular segments of the public. For example, if an employee becomes aware of one's company manufacturing and selling unsafe tires as those of high quality, there is an obvious risk to the public. Second, an employee must first report the concern to one's immediate superior. An advantage of this condition is that, if the issue can be handled internally, this preserves loyalty to the company as well as the protection of the public interest. Third, if an employee sees that nothing is done to address the problem, he/she can take her concern up the ladder as far as necessary. If no satisfactory response ensues, then the employee can go public with the concern, whether it be to the media, the government, or some other group that can apply pressure on the company.

Of course, whistle-blowers have been subject to penalties, demotions, isolation in the company, blackballing in the industry, and discharge from their jobs. With this in mind, consider the other points regarding whistle-blowing raised by DeGeorge. It is arguably easier to change laws than it is to change the practices of all corporations, and so issues related to whistle-blowing must be addressed by the law. For example, it should be illegal for a company to fire or take other punitive actions against a whistle-blower, at any time, if the employee has satisfied the set of conditions discussed above. Additionally, the law could include penalties for those individuals within a corporation who are responsible for the defect or harm, instead of just fining the company and leaving the people who are actually morally responsible untouched. Another suggestion is that all companies of a certain size or larger could have (mandated by law) an inspector general whose job is to look for immoral and illegal practices, and be a resource for employees to go to to voice their concerns. Unfortunately, prohibiting the punitive treatment of whistle-blowers by corporations cannot be entirely accomplished by the law, given the subtle ways a company can punish a whistle-blower. However, some of the immoral treatments of whistle-blowers can be stopped by the practices discussed above. Another strategy for encouraging and protecting whistle-blowers is for unions and professional organizations to provide channels for whistle-blowing, as well as protection of those who call attention to an immoral or illegal corporate practice. Generally, professional organizations and societies have failed to protect their members. The problem is that such

organizations may not have sufficient resources, or sufficient power, to do anything, which again stresses the need for legal sanctions and motivations for ethical corporate behavior and protection of whistle-blowers.

The need for laws and regulations to promote ethical behavior in business is exacerbated by the complexity and new forms of business conduct.^[7] These complexity and newness may make an appeal to traditional ethical precepts less effective than in the past. In 1991, the United States Corporate Sentencing Guidelines went into effect, enjoining both uniform and severe penalties for corporations found guilty of criminal misconduct. In the guidelines, corporations are given incentives for the adoption of ethics programs. If they have such programs, the penalties are significantly reduced if and when a corporation is convicted of an offense but have a program in place prior to that offense. Additionally, it is often the case that both federal and state prosecutors consider refraining from prosecuting a company for misconduct if an ethics program was in place prior to the offense.

The Corporate Sentencing Guidelines offer seven steps to help guide the formation of an ethics program that will both meet government approval and be an effective deterrent of wrongdoing. First, a company's ethics standards and procedures must be specifically constructed so as fit to the company's particular business and needs. Second, particular individuals who are high-ranking officials in the company must be personally involved and responsible for compliance, legitimizing the company's stated ethical goals. Third, companies are to avoid delegating significant discretionary authority to individuals who have shown themselves to be inclined to illegal conduct. Fourth, company leaders should seek to communicate in an effective manner ethics policies and procedures to employees and other agents of the company. Fifth, companies should work for compliance with the standards through monitoring, audits, and a working reporting system. It is important that companies target likely offenses in their monitoring, and they also should have a functioning disciplinary process in place when misconduct is discovered. Establishing a culture of morality and compliance with the law can do much to discourage immoral and illegal behavior. Sixth, companies must have disciplinary procedures in place for enforcing the ethics program. This might include reprimands, fines, suspension, or firing. Lastly, companies must respond to offenses and seek to deter further ones. When misconduct does occur, it should be reported to the government. In addition, companies should modify their ethics programs and other policies, procedures, and structures, as needed, to avoid future problems.

Governmental agencies such as the Environmental Protection Agency, the Nuclear Regulatory Commission

(NRC), and the Department of Health and Human Services can be instructive when seeking to develop effective ethics programs. For instance, with respect to whistle-blowing, the NRC gives more specific and detailed guidelines than the Corporate Sentencing Guidelines to the nuclear industry, partly because of the significant harm that could be avoided by whistle-blowers coming forward. NRC policy includes the training of supervisors to understand the value of whistle-blowers to the corporation, and it pushes the importance of having incentive programs in place that will encourage whistle-blowers to report dangerous, immoral, or illegal practices.

How Effective Are Whistle-Blowing Laws?

The level of corporate change regarding whistle-blowing policies in response to legal changes has been analyzed.^[8] According to the research of Near and Dworkin, protective legislation at the state level has not been effective. When whistle-blowers suffer retaliation, the majority continues to employ wrongful firing tort suits instead of seeking legislative protection via state statutes. Near and Dworkin explore whether the whistle-blowing statutes enacted by many states caused firms within those states to establish internal mechanisms for reporting wrongdoing as a possible explanation of the fact that no increase in the number of whistle-blowers seeking legislative protection has occurred.

Near and Dworkin surveyed a representative sample of *Fortune 1000* firms with headquarters in states where legislative protection of whistle-blowers exists. From the data they compiled, they conclude that the majority of firms continued to do business as usual, such that the passage of the whistle-blower statutes had little effect on corporate whistle-blowing policies and mechanisms. Most companies that responded heavily relied on an open-door policy, which Near and Dworkin argue is equivalent to no policy at all because employees feel that such policies are neither effective nor protective. Moreover, the policies of the majority of companies responding to the survey do not meet the Corporate Sentencing Guidelines because an open-door policy is not sufficiently specific, lacks the involvement of high-level officials, and does not include protection from retaliation against whistle-blowers. Companies surveyed failed to meet the guidelines with respect to the dissemination of information regarding their ethics statements and open-door policies. Because of this, then, they would be susceptible to harsher

sentences and fines, unless they instituted new policies in line with the guidelines. Lastly, the research of Near and Dworkin shows that whistle-blowing statutes enacted by many states did not cause firms within those states to establish internal mechanisms for reporting wrongdoings, given that only a fifth of the firms responding to their survey indicated that they had established such mechanisms in response to the changes in state statutes. To remedy this situation, Near and Dworkin suggest that: 1) legislation, which effectively encourages employees who become aware of wrong doing to blow the whistle, must be enacted and 2) governmental incentives, which encourage firms to construct internal channels for whistle-blowers to employ, must be put into place.

CONCLUSION

Whistle-blowers who are justified in bringing wrongdoings to light arguably deserve protection from the backlash that often occurs. Although legislative changes are important, it appears that more must be done to encourage change within corporate structures such that whistle-blowing, when justified, is not only allowed but encouraged. Public confidence in corporations as well as the public interest require nothing less.

REFERENCES

1. Jubb, P.B. Whistleblowing: a restrictive definition and interpretation. *J. Bus. Ethics* **1999**, *21*, 78.
2. Davis, M. Whistleblowing. In *The Oxford Handbook of Practical Ethics*; LaFollette, H., Ed.; Oxford University Press: Oxford, 2003; 549 pp.
3. Jensen, J. Vernon ethical tension points in whistleblowing. *J. Bus. Ethics* **1987**, *6*, 324–326.
4. Bok, S. Whistleblowing and professional responsibility. *N. Y. Educ. Q.* **1980**, *11*, 2–10.
5. Rae, S., Wong, K., Eds.; *Beyond Integrity*; Zondervan Publishing House: Grand Rapids, MI, 1996; 295–304.
6. DeGeorge, R.T. *Business Ethics*; Macmillan Publishing: New York, 1982.
7. Kaplan, J.M.; Rebecca, S. Walker ethics and the regulatory environment. In *A Companion to Business Ethics. Blackwell Companions to Philosophy*; Frederick, R., Ed.; Blackwell Publishers: Malden, MA, 1999; 366–373.
8. Near, J.P.; Dworkin, T.M. Responses to legislative changes: corporate whistleblowing policies. *J. Bus. Ethics* **1998**, *17*, 1551–1661.

Win-Win Approaches to Domestic Peace, Prosperity, and Democracy

Stuart S. Nagel

Political Science Department, University of Illinois, Urbana, Illinois, U.S.A.

INTRODUCTION

Win-win domestic policy analysis involves applying win-win analysis to domestic problems that relate to peace, prosperity, and democracy. Win-win analysis refers to analyzing a policy problem in such a way as to determine what alternative might be proposed that would enable conservatives, liberals, and other major viewpoints to come out ahead of their best initial expectations simultaneously.

In this article, win-win analysis is applied to domestic peace issues that relate to internal violence stemming from drug addicts seeking money to pay for drugs or seeking to expand territory for selling drugs. The analysis is applied to prosperity issues that relate to simultaneously dealing with inflation and unemployment. The analysis is also applied to democracy issues that relate to increasing voter turnout.

All of these applications view peace, prosperity, and democracy in a domestic or national context rather than in an international or multinational context, as is done in a related article in the Summer 2001 issue of *Peace, Prosperity, and Democracy*. There, the peace problem is the problem of U.S. foreign policy since the Cold War and the relation between the United States and United Nations. The analysis is applied. The prosperity problem relates to tariffs and international exchange rates. The democracy problem is promoting international democracy through trade and other means.

PEACE ANALYSIS: DRUG VIOLENCE

Conservative and Liberal Positions

The last policy problem regards drug-related crime, which has become an increasingly serious issue. It currently consumes a great deal of money in the United States, and is causing enormous losses in productivity. It is causing governments to engage in corruption and police officers to use abusive behavior due to the tremendous frustration brought on by the drug war (Table 1).

The key conservative goal in this context is to eliminate or greatly reduce drug-dealing. This means reducing the buying and selling of illegal drugs, most

commonly heroin, cocaine, and marijuana. Cocaine and crack represent the largest problem, with involvement in a large percentage of crime that occurs. Sixty to seventy percent of all people who are arrested for crimes against property or crimes against person are arrested under circumstances that indicate that they are stealing in order to get money for drugs. Often, they are under the influence of drugs while committing crimes. The crimes against the persons usually involve power plays for staking out drug territory. They frequently lead to murder, because under the influence of drugs, people behave in a more aggressive way than they otherwise would.

Liberals are especially concerned about reducing the side effects of the war on drugs. These side effects refer to the police engaging in more illegal arrests, illegal searches, and illegal interrogations than before crack cocaine became widespread in the mid-1980s. They also include the tremendous amount of money necessary for the imprisonment of over 1,000,000 people in our prisons and 500,000 people in our jails. That money could be much better spent on health care, education, or other more useful purposes.

The conservative solution has been to try to repress the sale and possession of drugs. This law enforcement type of approach is an approach of repression, of prohibition. One of the more extreme liberal approaches, although becoming somewhat more popular, is legalization. The policy of drug prohibition and repression is analogous to the policy of liquor prohibition. Back then, the United States policy makers realized that liquor prohibition did not work, and that the policy's abolishment was better, even if it meant more alcohol-related crime. Legalization is the liberal alternative to conservative police enforcement and repression.

A Win-Win Alternative

The object of the win-win solution is to come up with some kind of idea that wipes out the drug dealers but will not increase the use of drugs. Another way to phrase it is to de-profitize drug dealing. If you de-profitize drug dealing, then you also eliminate the side effects. The organized crime cartels would fall without drug money to support them, and crime would

Table 1 Evaluating Policies for Dealing with the Drug Problem

ALTERNATIVES	GOALS	
	CONS	LIB
	Decrease Crime Committing	Societal Productivity
CONS		
Law enforcement crackdown	+ ^a	- ^b
LIB		
Legalization	- ^b	+ ^a
NEUTRAL		
1. Treatment and education		
2. Prohibit cocaine and heroin, not pot	0 ^c	0 ^c
Superoptimum solution or win-win		
Medicalization de-profitize drug sales	++ ^d	++ ^d

^a+ = The best of the alternatives on the goal before the superoptimum alternative is considered.

^b- = The alternative is relatively adverse to the goal.

^c0 = Neither conducive nor adverse.

^d++ = The best of the alternatives on the goal after the superoptimum solution alternative is considered.

decrease accordingly. Another side effect that would be eliminated is the corruption of government officials due to the availability of “mutually profitable arrangements” with large dealers. An ounce of pure cocaine has a street value higher than that of an ounce of gold. This causes a lot of frustration for the police, because as soon as one drug dealer gets arrested and imprisoned or killed, there is another set up to immediately take their place. The money is just too good for some people to turn down.

The way to de-profitize drug dealing is to treat drug addicts like sick people and to consider them criminals only if they engage in crimes other than possession and use of drugs. But, a key part of the treatment is to treat drug addicts as being in need of medical treatment under our current health care programs. As health care plans move in the direction of becoming universal, the need to cover drug addiction is important. If drug addicts are considered to be sick people and eligible for prescription drugs, that could include a cocaine substitute or even actual cocaine under a cocaine maintenance program. The profits of drug dealing would be wiped out, because addicts would be able to get what they need through their subsidized health maintenance organization (HMO). These maintenance prescriptions would usually be given on a long-term, phase-out basis until the person is clean. If they never kick their drug

habit, their doctors can prescribe additional doses. No more would drug dealers be interested in hooking 10-year-old kids with free samples; all they would be doing is creating new patients for the HMOs. The dealers would eventually be forced to look for some alternative occupation, as did the bootleg liquor dealers of the past. At the present time, that would be better than the drug dealing they were previously into.

The win-win solution is plagued by psychological feasibility problems. Many do not believe that such availability will have any effect other than to increase the use of drugs. The idea that people who are not drug addicts would go to their HMO and say “I’m a drug addict, I’m on cocaine,” when they want to experiment is farfetched. Such a declaration would likely cause them to sacrifice many opportunities that they might otherwise have had—employers are not too enthusiastic about hiring drug addicts. The real drug addicts, though, would have an incentive to go to the HMOs. There they could get a maintenance prescription without having to rob or kill for it and without having to run the risk of getting arrested or killed.

This kind of approach appeals to some libertarian conservatives who think that a repression program involves more problematic government interference than interference in the economy. It should also appeal to liberals. It is better than legalization, because legalization means that controls would be much less strict, and that anybody could have access to drugs, including children. Those children would likely grow up to be much less productive members of society. Legalization, in fact, would have a very bad effect on national productivity, whereas this kind of program, which is only for confirmed drug addicts, would have few new effects. There are not going to be a whole lot of new drug addicts created, because there will not be a set of drug dealers encouraging the younger generation to become drug addicts.

As said before, the feasibility problem here is psychological. People are resistant to the idea of giving drugs to anyone, let alone those who are already addicts. That sounds somehow evil, but it is less evil than the present system might be. Because this plan is psychologically hard to accept, it is difficult politically to adopt. Economically, this solution would cost far less than what is currently being done. Cocaine, for instance, is very cheap to produce. Its high price largely has to do with police regression, which causes the costs of drug dealing to rise, and therefore, the street price to rise also. There might also be some feasibility problems with regard to administering such a program. The drugs possessed by the HMOs and hospitals must be carefully monitored so that they do not fall into the wrong hands. This can be taken care of by authorizing only one doctor in the whole HMO to prescribe drug maintenance programs or drugs. This would greatly decrease the amount

of people with access and would thereby decrease the risk. Also, in order to be eligible to receive a drug maintenance prescription, a person must be a confirmed addict. This would require testing, over time, of the levels of drugs in a person's system to see if they qualify. The administrative part of the program should not require people to accept a phase-out arrangement. Such restrictions would cause people to avoid the programs and would cause the formation of a black market, which defeats the purpose. That is why a similar program did not succeed in England. As long as these people are treated as sick people, not criminals, this kind of administration is likely to succeed.

PROSPERITY ANALYSIS: JOBS AND PRICES

Conservative Alternatives

Doing nothing is not likely to worsen unemployment or inflation, but it is also not likely to help. Decreasing the money supply and increasing interest rates may decrease inflation but increase unemployment. The same is true of decreasing government spending and increasing taxes. The Reagonomics approach involved decreasing taxes to stimulate employment and decreasing domestic spending to reduce inflation.

The Democratic counterpart as of 1980 was to increase employment through government jobs and decrease inflation through price control (Table 2).

Raising interest rates to decrease inflation may have the effect of decreasing prices by reducing spending from borrowed money. Those benefits may be more than offset by the undesirable effects on reducing the ability of business to borrow for expansion, inventory, and other purposes. The reduction in spending may also have an adverse effect on employment. Raising taxes and decreasing spending to fight inflation may not be politically feasible. It would also reduce the ability of the government to give tax breaks and well-placed subsidies to increase productivity.

Liberal Alternatives

Increasing the money supply and decreasing interests may stimulate employment but increase inflation. The same is also true of increasing government spending and reducing taxes. Lowering interest rates to decrease unemployment may have little impact, because businesses are reluctant to borrow when they are reducing their operations and sales are down. Likewise, consumers are reluctant to borrow when they are already heavily in debt and fearful of a reduction in employment or

Table 2 Unemployment

ALTERNATIVES	GOALS	
	CONS	LIB
	1. Lower inflation to 3% 2. Free enterprise	1. Lower unemployment to 3% 2. Distribute inflation and unemployment
CONS		
1. Do nothing		
2. Decrease money supply and increase interest rates		
3. Decrease spending and increase taxes		
4. Decrease taxes and decrease domestic spending	+	-
LIB		
1. Increase money supply and decrease interest rates		
2. Increase spending and decrease taxes		
3. Create jobs and control prices		
4. Create tax breaks and subsidies	-	+
5. Decrease defense spending		
NEUTRAL		
1. Combine	0	0
Superoptimum solution or WIN-WIN		
Provide subsidies with strings attached for new technologies and training	++	++

^a+ = The best of the alternatives on the goal before the superoptimum alternative is considered.

^b- = The alternative is relatively adverse to the goal.

^c0 = Neither conducive nor adverse.

^d++ = The best of the alternatives on the goal after the superoptimum solution alternative is considered.

hours. Lowering taxes and increasing spending to fight unemployment may not be politically feasible when the national debt and deficit are already too high.

A Win-Win Alternative

Increasing the adoption of new technologies and raising the skills of workers to help reduce inflation by increasing the productivity of labor to offset increased wages, increasing the quality of goods to offset increased prices, and increasing the gross national product (GNP) and domestic income to further offset increased prices.

Increasing the adoption of new technologies and raising the skills of workers helps reduce unemployment by making the workers more employable, increasing the GNP and domestic spending to stimulate the creation of more jobs, and increasing the productivity and wage rates, thereby offsetting a possible reduction in hours.

The conservative alternative of having interest rates up in times of inflation and down in times of unemployment does not make sense if inflation and unemployment are simultaneously problems. That would be so if both were over 3%. Likewise, the liberal alternative of having a budget surplus in times of inflation and a budget deficit in times of unemployment does not make sense when inflation and unemployment are over approximately 3%. One can, however, stimulate new technologies and upgrade skills when inflation and unemployment are occurring simultaneously.

PROSPERITY ANALYSIS: DISPLACED WORKERS AND FIRMS

Displacement of labor means displaced because of productivity downsizing, free trade, immigration, civilian conversion, and jobs for public aid recipients, and the disabled, the elderly, minorities, and women. The issue here is how to find jobs for displaced workers. The conservative emphasis is to leave it up to the recipient to find a job and not make it a responsibility of other people. The liberal emphasis is on the welfare agency or another government agency doing most of the job-finding work. The neutral position might involve delegating the activity to a non-profit organization (Table 3).

A key conservative goal is to save tax money. That means encouraging job finding to reduce welfare payments but not incurring high fees for job finding. A key liberal goal is to find jobs for displaced workers or welfare recipients, not just to save welfare payments, but also because jobs can increase the income, quality of life, and dignity of welfare recipients. Doing so also has effects that relate to multipliers, compounding, role models, and reducing illegal activities.

A superoptimum solution alternative is to contract out to a private profit-making firm at a commission of about \$X per welfare recipient who receives long-term employment. Half of the commission is paid after four months on the job and the other half after eight months. The firm is responsible for providing training, day care, employment leads, advice, and dispute resolution. This is a good example of contracting out. The profit motive stimulates more success in finding jobs than the rate of success by a government agency or a non-profit organization. The firm also has more capability than the recipient. Tax money is saved in the long run as a result of replacing welfare with work. It may also be saved in the short run by costing less money per long-term job found than the cost with a government agency or non-profit organization.

DEMOCRACY ANALYSIS: VOTING

Conservative and Liberal Position

The United States does not look good relative to the rest of the world when it comes to voter turnout. We have roughly (these are very round numbers) 200 million adults, of which approximately half are registered to vote. But, only about half of those who are registered to vote actually do. That is about 50 million. That means if just 26 million out of the 50 million vote for a certain candidate for president, he gets in. This can be a landslide if each state is hotly contested, even though it is only 26 million out of 200 million possible voters. It is not so good when, in effect, 12% of the population can decide who will sit in the Oval office. We are not undemocratic in the sense of prohibiting people, but less democratic than we should be in the sense of facilitating voter turnout (Table 4).

The true conservative goal might be to promote the election of conservative candidates, but that is not going to be said. What is actually said is that multiple voting should be avoided. Conservatives do not want any schemes that will allow cheating at the polls. Liberals on the other hand, are concerned about people who do not vote, so they want to decrease non-voting by adults who could be eligible. The conservative position promotes a decrease in multiple voting a number of ways. One of the most extreme positions is to purge the voter rolls every 10 years and make people register over again. This would guard against individuals still being present on the voter rolls who have moved or died. It greatly decreases the number of people who register if you have to do it over and over again. Having advanced registration may make a difference with regard to decreasing multiple voting, but it will also decrease voting in general. Liberals also want advanced

Table 3 Equity versus Efficiency in Displacement of Labor

ALTERNATIVES	GOALS	
	CONS	LIB
	EFFICIENCY (merit or survival of the fittest)	EQUITY (i.e., fairness to those unemployed to no fault)
CONS		
Marketplace (leaves to the labor marketplace)	+ ^a	— ^b
LIB		
Welfare handouts with few conditions	— ^b	+ ^a
NEUTRAL		
Welfare with conditions:		
1. No able-bodied eligibles, especially males		
2. Bare minimum benefits		
3. Residence requirements		
4. Provide no due process	0 ^c	0 ^c
Superoptimum solution or win-win, i.e., job facilitators		
1. Training		
2. Wage subsidy		
3. Employment agency commissions		
4. Rising GNP		
5. Relocation		
6. Welfare conditional on training and job cooperation	++ ^d	++ ^d

^a+ = The best of the alternatives on the goal before the superoptimum alternative is considered.

^b— = The alternative is relatively adverse to the goal.

^c0 = Neither conducive nor adverse.

^d++ = The best of the alternatives on the goal after the superoptimum solution alternative is considered.

registration, but they want to make the process easier. They support ideas like postcard registration or registration at the time you get your driver's license, so-called motor-voter legislation. They also support keeping the polls open a few hours later to make it easier to vote on election day. Unfortunately, the liberal solutions in total would not make much of a difference, as more fundamental change is necessary.

A Win-Win Alternative

What really needs to be adopted is the kind of system presently being used in many countries of the world, including Canada, South Africa, Mexico, and Mozambique, and in states such as Wisconsin and Minnesota. It involves a few innovations. First, there is no requirement of advance registration. You can register in advance if you want to, but you can also register on-site the day of the election. Many people who do not vote on election day give as a reason the fact that they are not registered. If they could register on election day through on-site registration, that would no longer be a problem. Also, if elections were to be moved to a non-working day instead of a Tuesday, more people would be able to participate. In Catholic

countries like Italy, France, and Mexico, election day is on a Sunday, when people do not work. Other nations string their elections over a couple of days. Another improvement would allow people to vote in their home precinct or their work precinct, or even allow them to vote in any precinct, provided there is some way of checking to make sure they have not voted in another. Multiple-precinct voting can make a large difference in voter turnout.

With all of these facilitators, however, the conservative problem of multiple voting rises again. The way that problem is solved in South Africa, Mexico, Mozambique, and other countries, is by having voters dip their hand into a bowl of invisible ink. If you show up at a polling place anytime in the day after that, your hand is viewed under an ultraviolet lamp, and it will show that you have already voted. The invisible ink method works much better than asking people to sign their names, because names can be forged much more easily than hands. It is a good example of a win-win solution, because it would substantially decrease multiple voting and substantially decrease non-voting. It helps to achieve both kinds of goals simultaneously, like the previously mentioned outreach training, or the more cleaner profitable processes that relate to environmental protection and economic growth. All four that we

Table 4 Voter Turnout

ALTERNATIVES	GOALS	
	CONS	LIB
	Double Voting	Non-voting
CONS		
1. Advance registration	0 ^c	— ^b
LIB		
1. Postcard		
2. Motor voter	— ^b	0 ^c
NEUTRAL	0 ^c	0 ^c
Superoptimum solution or win-win		
1. On-site registration		
2. Non-working election day		
3. Multiple precinct		
4. Invisible ink	0 ^c	0 ^c

^a+ = The best of the alternatives on the goal before the superoptimum alternative is considered.

^b— = The alternative is relatively adverse to the goal.

^c0 = Neither conducive nor adverse.

^d++ = The best of the alternatives on the goal after the superoptimum solution alternative is considered.

have discussed so far are capable of achieving the conservative and the liberal goals more successfully than the conservative or liberate alternative.

On-site registration, the non-working election day, multiple precincts, and the invisible ink method all have a political feasibility problem. This problem is political in the sense that it is difficult to get such measure through Congress or a state legislature. This is because one political party is likely to have enough strength to block it, namely, the party that thinks it will suffer as a result of expanded voter turnout. As long as it has enough power to stop these measures, they will never be adopted. If, however, they ever were adopted, they are not likely to be repealed. There is a kind of ratchet effect on new facilitators once they are adopted, because the party in power will look bad if it decreases the ability of people to register and vote. These measures are likely to be adopted when the Democrats have enough influence in the Congress or in the various state legislatures. The traditional thinking is that the Democratic party benefits more from expanded voter registration and turnout than the Republican party, due to the fact that a higher percentage of non-voters consist of the people who are poorer and less educated than average. Such voters are more likely to vote Democratic. The problem is political, but it may only be a temporary problem, as hopefully all the other feasibility problems may be temporary.

BIBLIOGRAPHY

Peace

U.S. foreign policy

1. Ripley, R.B., Lindsay, J.M., Eds.; *U.S. Foreign Policy After the Cold War*; University of Pittsburgh Press, 1997; 392 pp.
2. Roberts, B., Ed.; *U.S. Foreign Policy After the Cold War*; MIT Press, 1992; 367 pp.

Crime and Criminal Justice

1. Scott, J.E., Hirschi, T., Eds.; *Controversial Issues in Crime and Justice*; Sage, 1988; 214 pp.

Prosperity

Economic policy

1. Hody, C.A. *The Politics of Trade: American Political Development and Foreign Economic Policy*; University Press of New England, 1996; 210 pp.
2. Brock, W., Hormats, R., Eds.; *The Global Economy: America's Role in the Decade Ahead*; W.W. Norton & Company, 1990; 264 pp.

Technology policy

1. Brooks, H., Cooper, C., Eds.; *Science for Public Policy*; Pergamon, 1987; 294 pp.
2. Kuehn, T., Porter, A., Eds.; *Science, Technology, and National Policy*; Cornell, 1981; 530 pp.

Democracy

Political policy

(participation/majority rule)

1. Penncock, R. *Liberal Democracy: Its Merits and Prospects*; Rinehart, 1950; 415 pp.
2. Breton, A., Galeotti, G., Salmong, P., Wintrobe, R., Eds.; *Understanding Democracy: Economic and Political Perspectives*; Cambridge, 1997; 286 pp.

Social policy

1. Darby, M., Ed.; *Reducing Poverty in America: Views and Approaches*; Sage, 1996; 390 pp.
2. McFate, K., Lawson, R., Wilson, W., Eds.; *Poverty, Inequality, and the Future of Social Policy: Western States in the New World Order*; Russell Sage, 1995; 768 pp.

Win-Win Approaches to International Peace, Prosperity, and Democracy

Stuart S. Nagel

Political Science Department, University of Illinois, Urbana, Illinois, U.S.A.

INTRODUCTION

Win-win international studies involve applying win-win analysis to problems of international peace, prosperity, and democracy. Win-win analysis refers to analyzing a policy problem in such a way as to determine what alternative might be proposed that would enable conservatives, liberals, and other major viewpoints to all come out ahead of their best initial expectations simultaneously.

In this article, win-win analysis is applied to peace issues that relate to U.S. foreign policy since the Cold War and to relations between the United States and the United Nations. The analysis is applied to prosperity issues that relate to tariffs and exchange rates. The analysis is also applied to democracy issues that relate to promoting international democracy through trade and other means.

All of these applications view peace, prosperity, and democracy in an international or multinational context rather than in a domestic context. The peace problem may be the problem of domestic violence due to drug addicts seeking money to pay for drugs or seeking territory for the sale of drugs. The prosperity problem might deal with inflation and unemployment simultaneously. The democracy problem can be facilitating voter turnout.

WIN-WIN PEACE ANALYSIS

U.S. Foreign Policy Since Cold War

Cold War U.S. foreign policy divided conservatives and liberals. Conservatives tended to favor right-wing governments and large defense expenditures, whereas liberals tended to favor left-wing governments and lower defense expenditures. Now, both sides support lower defense expenditures and are less influenced by the economic ideology of foreign governments.

The prevailing conservative orientation seems more isolationist and less interventionist than the current liberal orientation. That is shown in the greater liberal support for intervention in Bosnia, Somalia, and other places where peace making or humanitarian aid might be appropriate. A compromise position would

advocate greater selectivity in the application of isolationism or peace making.

The key conservative goal seems to be to save taxes and promote domestic business. The key liberal goal seems to be to promote peace in the workplace and domestic employment opportunities.

A superoptimum solution might be to see peace as part of a larger international package. That foreign-policy package would include the United States actively promoting peace, prosperity, and democracy throughout the world in order to help all nations become better customers for the United States, suppliers to the United States, places for investment, and better world citizens. That kind of pragmatic policy would be mutually beneficial in increasing the national incomes and quality of life in the United States and elsewhere ([Table 1](#)).

Separate analyses could be developed. An important factor in promoting peace is supporting the United Nations' peace-making activities. An important factor in promoting prosperity is free trade in the movement of products, factories, and labor. An important factor in promoting democracy is the free movement of ideas through the mass media and providing financial and technical support for democratic forces inside and outside potentially dictatorial governments.

The United States and the United Nations

The key conservative goal in this context, relating to the role of the United States in helping the United Nations to bring peace to Bosnia, is to save U.S. dollars and save lives. The conservatives were not always so oriented toward isolation. They were very interventionist in the days of the cold war, especially with regards to curtailing the expansion of the Soviet Union. But at the present time, they espouse more of an isolationist policy and less of an intervention policy than do liberals. The conservative alternative for saving U.S. dollars and lives is to keep U.S. troops at home. They feel it is more important to develop the United States first. The liberals, on the other hand, are more oriented toward being the policemen of the world, traveling with the goal of promoting world

Table 1 U.S. foreign policy since the cold war

ALTERNATIVES	GOALS	
	CONS	LIB
	Save taxes and promote business	Promote peace and employment
CONS		
Isolationism	+ ^a	— ^b
LIB		
Peace making	— ^b	+ ^a
NEUTRAL		
Selective peace making	0 ^c	0 ^c
SOS OR WIN-WIN		
Peace as part of larger international package	++ ^d	++ ^d

^a+ = The best of the alternatives on the goal before the superoptimum alternative is considered.

^b— = The alternative is relatively adverse to the goal.

^c0 = Neither conducive nor adverse.

^d++ = The best of the alternatives on the goal after the superoptimum solution alternative is considered.

peace. More idealistic or maybe more naive, they believe what is sometimes referred to as a social worker ideology. That necessitates intervention.

The win-win solution saves U.S. dollars, U.S. lives, and promotes world peace simultaneously. An approach recommended by some who specialize in international relations is the idea of a volunteer U.N. force. What this would mean is that American troops would not be sent to Bosnia, Somalia, or anywhere else. The United Nations would have its own volunteer force, a force similar to the French Foreign Legion. It would consist of people from all over the world who would join partly out of idealism, partly because it pays some money. They would not be allowed to join unless they had some military experience or some aptitude for military training. It would provide an opportunity for Vietnam War veterans, Russian soldiers who had fought in Afghanistan, and others to participate in a force that would be devoted to a righteous cause. The officers would be officers who had served in the militaries of their various countries.

As a result, there would be no American soldiers in this U.N. force. There may be Americans in the U.N. force, but they would be Americans acting on their own. They would be volunteering to serve. They would not be drafted, or as in the case of the current Americans serving in Bosnia, ordered to go because their unit was called. They are not completely volunteers. These would be pure volunteers, pure in the sense they would be joining the U.N. force with the agreement to go wherever they are sent. If they will not go, for ideological or other reasons, then an option can be provided where they can quit. In such a case, it might be fair to have them pay back something in return for their training. Such an arrangement has many possibilities.

Such a plan has been opposed in the past by the United States because it was feared that the United Nations could not be trusted to have its own volunteer force. Strictly by the numbers, the United Nations in the past has been too much under the control of developing nations or nations under the control of the Soviet Union. But now, with the breakup of the Soviet Union, the United States dominates the United Nations as the only superpower in the world. The volunteer force could not be a threat to the United States. Though a bit behind in its payments, the United States is the leading source of funds to the United Nations, which further adds to its influence. If the United States does not like what the volunteer force is doing, it could just cut off the funding, use its veto power in the Security Council, or use its influence among countries of the world. This kind of volunteer force would mean that no American lives or American dollars would be used, yet it could still be effective in promoting world peace. It is possible that the volunteer force might call upon member countries to supply additional troops at some time, but this would be handled as the need arose. This represents another kind of win-win solution (Table 2).

The feasibility problem for this issue is psychological. It would be conceived that such a force might represent the muscle behind some type of world government that could partly deprive the United States or its sovereignty. Some might fear the volunteer force is going to march in and take over. As with all the other obstacles, it is necessary for this one to be overcome if this win-win solution is going to be adopted. People have to realize that a U.N. volunteer might not be a threat to them. In no way would the people with blue helmets serving in Bosnia invade the United States.

Table 2 The United States and the United Nations

ALTERNATIVES	GOALS	
	CONS	LIB
	Save U.S. dollars and lives	World peace
CONS		
Minimum U.S. involvement (isolationism)	+ ^a	- ^b
LIB		
Substantial U.S. involvement (intervention)	- ^b	+ ^a
NEUTRAL		
Some of each Superoptimum solution or Win-Win	0 ^c	0 ^c
Volunteer U.N. force	++ ^d	++ ^d

^a+ = The best of the alternatives on the goal before the superoptimum alternative is considered.

^b- = The alternative is relatively adverse to the goal.

^c0 = Neither conducive nor adverse.

^d++ = The best of the alternatives on the goal after the superoptimum solution alternative is considered.

INTERNATIONAL PROSPERITY ANALYSIS

Tariffs

U.S. exporters and investors are helped by free trade with Mexico and other places because Mexicans can buy more U.S. products if there are no Mexican tariffs artificially raising the price of American products, Mexicans can buy more U.S. products if they have more income as a result of working in factories that have expanded as a result of American capital, and U.S. investors can make money and add to the U.S. gross national product by investing in Mexican factories that are now able to export better to the United States, because U.S. tariffs have been dropped.

U.S. consumers are helped by free trade with Mexico and other places, because they can buy products made in Mexico at lower prices, because they no longer have a U.S. tariff artificially raising the prices; they can benefit from low prices that should result from decreased labor expenses associated with some products made in Mexico, possibly stimulated with American capital; and U.S. consumers include business firms that buy producer goods less expensively from Mexico and thereby make American firms more internationally competitive.

U.S. firms and workers who are not sufficiently competitive would be hurt by the North American Free Trade Agreement (NAFTA), but this can be

minimized by retraining workers and firms so they can be more competitive in their old products or new products, making side agreements with Mexico that require upgrading of labor standards in Mexico, and helping disrupted workers and firms benefit from the increased prosperity of the U.S. as a result of more exporting, better overseas investing, and better buys for U.S. consumers.

Mexicans can benefit in the same ways as Americans by substituting for the three goal-columns: Mexican exporters and investors, non-competitive Mexican firms and workers, and Mexican consumers.

The opponents of NAFTA are referred to in [Table 3](#) as conservatives, and the advocates are referred to as liberals. That is done partly to simplify the calculation of the tools. It is also in accordance with the fact that conservatives have traditionally been in favor of high tariffs, although in recent years, that is less true than from about 1800 through the 1930s.

Exchange Rates

A high dollar value means that \$1 will buy many units of the currencies of other countries. A low dollar value means that \$1 will buy relatively few units of the currencies of other countries.

If the dollar has a relatively high value, then we have difficulty selling to other countries, because they have to give a lot of units of their currencies in order to get a dollar. If the dollar has relatively low value, then we have difficulty buying from other countries, because Americans have to give a lot of dollars in order to get the currencies of other countries.

If we concentrate on improving the quality and prices of American goods, then we can sell a lot of American goods to other countries without lowering the value of the dollar. A big effect of selling a lot more to other countries is an increase in the American national income. That enables us to have a lot more money to buy from other countries without raising the value of the dollar.

Thus, improving the quality and price of American goods through upgrading technologies and skills is a good superoptimum solution, because it can achieve both goals of increased buying and increased selling simultaneously. This is in contrast to manipulating the dollar, which increases one goal but decreases the other goal in a typical trade-off pattern ([Table 4](#)).

The superoptimum solution alternative also recognizes that trade, even without human rights conditions, can promote human rights by promoting prosperity and internal education. Democratic institutions can also promote prosperity and trade. There is thus reciprocal causation, even without explicitly linking the two.

Table 3 Tariffs

ALTERNATIVES	GOALS	
	CONS	LIB
	1. Non-competitive U.S. firms and workers	1. U.S. exporters and investors 2. U.S. consumers
CONS		
1. Opponents	+ ^a	– ^b
2. Labor unions and weak business firms		
LIB		
1. Advocates	– ^b	+ ^a
2. Intellectual liberals		
3. Business interests		
NEUTRAL		
1. Middling	0 ^c	0 ^c
Superoptimum solution or Win-Win		
1. Free trade	++ ^d	++ ^d
2. Retraining		
3. Side agreements		
4. Expanding the economy		

^a+ = The best of the alternatives on the goal before the superoptimum alternative is considered.

^b– = The alternative is relatively adverse to the goal.

^c0 = Neither conducive nor adverse.

^d++ = The best of the alternatives on the goal after the superoptimum solution alternative is considered.

Trade when used to encourage human rights needs to be presented more as a reward or bribe, rather than as a threat or punishment. Trade has worked well to encourage democratic institutions in places like South Africa. It can backfire if trade is withdrawn to the

point where extremists come to power (as in Russia), or to the point where the economy suffers long-term destruction (as could happen in Haiti).

Trade can also be used as a bargaining chip for other purposes besides human rights conditions, such

Table 4 Exchange rates

ALTERNATIVES	GOALS	
	CONS	LIB
	Buy from other countries	Sell to other countries
CONS		
High dollar value	+ ^a	– ^b
LIB		
Low dollar value	– ^b	+ ^a
NEUTRAL		
Middle dollar value	0 ^c	0 ^c
Superoptimum solution or Win-Win		
Improve quality and price of U.S. goods	++ ^d	++ ^d

^a+ = The best of the alternatives on the goal before the superoptimum alternative is considered.

^b– = The alternative is relatively adverse to the goal.

^c0 = Neither conducive nor adverse.

^d++ = The best of the alternatives on the goal after the superoptimum solution alternative is considered.

Table 5 Promoting international democracy

ALTERNATIVES	GOALS	
	CONS	LIB
	Business profits	Spread democracy
CONS		
Trade without human rights conditions	+ ^a	– ^b
LIB		
Trade with human rights conditions	– ^b	+ ^a
NEUTRAL		
Trade with slight conditions	0 ^c	0 ^c
Superoptimum solution or Win-Win		
Promoting trade and rights partly through education and communications	++ ^d	++ ^d

^a+ = The best of the alternatives on the goal before the superoptimum alternative is considered.

^b– = The alternative is relatively adverse to the goal.

^c0 = Neither conducive nor adverse.

^d++ = The best of the alternatives on the goal after the superoptimum solution alternative is considered.

as tariff reductions and the opening of investment opportunities, which conservatives would also endorse.

WIN-WIN DEMOCRACY ANALYSIS

This analysis stems largely from the controversy over how far to go in withholding trade from China until human rights are given more recognition. The problem, however, applies to many countries that would like increased trade with the United States but lack a minimum level of domestic institutions.

The conservative position is to have trade without human rights conditions or prerequisites. The liberal position is to use trade to secure better human rights conditions, including the possibility of a pluralistic political system which allows two or more political parties to meaningfully compete for votes. A neutral position would think in terms of less substantial human rights, such as visits by Red Cross workers to political prisoners, but not ending political prisoners.

The conservative position is oriented toward American business profits. The liberal position is oriented toward spreading democracy. Both sides endorse each other's goals but not with equal weight.

The object of a superoptimum alternative would be to promote trade and human rights, partly through education and communication. That means encouraging students from China and other such countries to come to the United States. Doing so means they are more likely to return with American values like democracy. That also means increasing the communication of democracy-related ideas by way of radio, television, newspapers, books, and other forms of communication to China and other such countries.

The superoptimum alternative also recognizes that trade, even without human rights conditions, can promote human rights by promoting prosperity and internal education. Democratic institutions can also promote prosperity and trade. There is, thus, reciprocal causation, even without explicitly linking the two.

Trade when used to encourage human rights needs to be presented more as a reward or bribe, rather than as a threat or punishment. Trade has worked well to encourage democratic institutions in places like South Africa. It can backfire if trade is withdrawn to the point where extremists come to power (as in Russia), or to the point where the economy suffers long-term destruction (as could happen in Haiti).

Trade can also be used as a bargaining chip for other purposes besides human rights conditions, such as tariff reductions and the opening of investment opportunities, which conservatives would endorse also (Table 5).

BIBLIOGRAPHY

PEACE BIBLIOGRAPHY

War

U.S. Foreign Policy

1. Crotty, W., Ed. *Post-Cold War Policy: The International Context*; Nelson-Hall, 1995; 344 pp.
2. Harkavy, R.E.; Kolodziej, E.A., Eds. *American Security Policy and Policy-Making*; Lexington Books, 1980; 268 pp.
3. Roberts, B., Ed. *U.S. Foreign Policy After the Cold War*; MIT Press, 1992; 367 pp.

- Ripley, R.B.; Lindsay, J.M., Eds. *U.S. Foreign Policy After the Cold War*; University of Pittsburgh Press, 1997; 392 pp.

Cross-National

- Kolodziej, E.A.; Harkavy, R.E., Eds. *Security Policies of Developing Countries*; Lexington Books, 1982; 393 pp.
- Nagel, S., Ed. *Global Policy Studies: International Interaction Toward Improving Public Policy*; MacMillan, 1991; 274 pp.
- Segal, G. *The World Affairs Companion: The Essential One-Volume Guide to Global Issues*; Simon & Schuster, 310 pp.
- Weatherby, J.N.; et al. *The Other World: Issues and Politics of the Developing World*; Longman, 1997; 303 pp.

International Interaction

- Carnegie Commission on Preventing Deadly Conflict. *Preventing Deadly Conflict*; Carnegie Corporation of New York, 1997; 257 pp.
- Nagel, S., Ed. *Global Policy Studies: International Interaction Toward Improving Public Policy*; MacMillan, 1991; 274 pp.
- Sisk, T.D. *Power Sharing and International Mediation in Ethnic Conflicts*; U.S. Institute of Peace, 1996; 143 pp.
- Zartman, W.; Rasmussen, J.L., Eds. *Peacemaking in International Conflict: Methods and Techniques*; U.S. Institute of Peace Press, 1997; 412 pp.

Crime

U.S. Domestic

- Baker, R.; Meyer, F., Eds. *Evaluating Alternative Law-Enforcement Policies*; Lexington, 1979; 226 pp.
- Nagel, S. *Improving the Legal Process: Effects of Alternatives*; Lexington, 1975; 394 pp.
- Nagel, S. *Law, Policy, and Optimizing Analysis*; Quorum, 1986; 328 pp.
- Walker, S. *Sense and Non-sense About Crime—A Policy Guide*, 2nd Ed.; Brooks/Cole Publishing Co., 1989; 276 pp.
- Wright, B.; Fox, V. *Criminal Justice and the Social Sciences*; W. B. Saunders Co., 1978; 419 pp.

Cross-National

- Dorn, N.; Jepsen, J.; Savona, E., Eds. *European Drug Policies and Enforcement*; MacMillan Press, 1996; 280 pp.
- Cole, G.F.; Frankowski, S.; Gertz, M.G. *Major Criminal Justice System*; Sage, 1981; 300 pp.
- Courts and justice abroad. In *Judicature*; American Judicature Society, May–June 1998; 81 (6).
- Terrill, R.J. *World Criminal Justice Systems—A Survey*, 4th Ed.; Anderson Publishing Co., 1999; 662 pp.

International Interaction

- Lesser, I.O. et al. *Countering the New Terrorism*; RAND—Project Air Force, 1999; 153 pp.
- Pearce, F.; Woodiwiss, M., Eds. *Global Crime Connections: Dynamics and Control*; University of Toronto Press, 1993; 269 pp.
- Raine, L.P.; Cilluffo, F.J., Eds. *Global Organized Crime: The New Empire Evil*; CSIS, 1994; 185 pp.

PROSPERITY BIBLIOGRAPHY

Economic Policy

General

- Anderson, J.E., Ed. *Economic Regulatory Policies*; D.C. Heath, 1976; 211 pp.
- Holzer, M.; Nagel, S., Eds. *Productivity and Public Policy*; Sage, 1984; 296 pp.
- Lee, Y.S., Ed. *Technology Transfer and Public Policy*; Quorum, 1997; 307 pp.
- Roessner, D. *Government Innovation Policy: Design, Implementation, and Evaluation*; MacMillan, 1988; 208 pp.

Domestic Concerns

- Brock, W.; Hormats, R., Eds. *The Global Economy: America's Role in the Decade Ahead*; W. W. Norton & Company, 1990; 264 pp.
- Hody, C.A. *The Politics of Trade: American Political Development and Foreign Economic Policy*; University Press of New England, 1996; 210 pp.
- Nagel, S. *Higher Goals for America: Doing Better than the Best*; University Press of America, 1989; 246 pp.
- Redburn, S.; Buss, T.F.; Ledebur, L.C., Eds. *Revitalizing the U.S. Economy*; Praeger, 1986; 232 pp.

Cross-National Comparisons

- Baker, D.; Epstein, G.; Pollin, R., Eds. *Globalization and Progressive Economic Policy*; Cambridge University Press, 1998; 514 pp.
- Blake, D.H.; Walters, R.S. *The Politics of Global Economic Relations*; Prentice-Hall, 1987; 248 pp.
- DeGregori, T.R.; Rodgers, H.R., Jr., Eds. *Policy Studies in Developing Nations: Poverty Policy in Developing Countries*; JAI, 1994; 306 pp.
- Groth, A.J.; Wade, L., Eds. *Public Policy Across Nations: Social Welfare in Industrial Settings*; JAI, 1985; 279 pp.
- Meyer, P.B., Ed. *Comparative Studies in Local Economic Development: Problems in Policy Implementation*; Greenwood, 1993; 202 pp.

International Interaction

1. Hogendorn, J.S. *Economic Development*; HarperCollins, 1992; 650 pp.
2. World Bank. *World Development Report 1995: Workers in an Integrating World*; Oxford University Press, 1995; 251 pp.

TECHNOLOGY POLICY

Domestic Concerns

1. Brooks, H.; Cooper, C., Eds. *Science for Public Policy*; Pergamon, 1987; 294 pp.
2. Kuehn, T.; Porter, A., Eds. *Science, Technology, and National Policy*; Cornell, 1981; 530 pp.

Cross-National Comparisons

1. Cody, J.; Hughes, H.; Wall, D., Eds. *Policies for Industrial Progress in Developing Countries*; Oxford, 1980; 325 pp.
2. Organization for Economic Co-operation and Development. *Technology and the Economy: The Key Relationships*; OECD, 1992; 328 pp.

International Interaction

1. Elliott, L. *The Global Politics of the Environment*; New York University, 1998; 312 pp.
2. Peterson, J.; Sharp, M. *Technology Policy in the European Union*; St. Martin's, 1998; 278 pp.

DEMOCRACY BIBLIOGRAPHY

Political Participation and Majority Rule

Domestic U.S. Concerns

1. Breton, A.; Galeotti, G.; Salmon, P.; Wintrobe, R., Eds. *Understanding Democracy: Economic and Political Perspectives*; Cambridge, 1997; 286 pp.
2. Crotty, W., Ed. *Political Participation and American Democracy*; Greenwood, 1991; 233 pp.
3. DeSario, J.; Langton, S., Eds. *Citizen Participation in Public Decision Making*; Greenwood, 1987; 237 pp.
4. Grofman, B.; Lijphart, A.; McKay, R.B.; Scarrow, H.A., Eds. *Representation and Redistricting Issues*; Lexington, 1982; 278 pp.
5. Nagel, S. *The Policy Process and Super-Optimum Solutions*; Nova Science, 1994; 216 pp.
6. Pennock, R. *Liberal Democracy: Its Merits and Prospects*; Rinehart, 1950; 415 pp.
7. Wasky, S.L. *Civil Liberties: Policy and Policy-Making*; Feffer & Simons, 1976; 235 pp.

Cross-National Comparisons

1. Bauzon, K., Ed. *Development and Democratization in the Third World: Myths, Hopes, and Realities*; Crane Russak, 1992; 363 pp.
2. Nagel, S.; Crotty, W.; Scarritt, J., Eds. *Policy Studies and Developing Nations: Political Reform and Developing Nations*; JAI, 1996; 331 pp.
3. Pinkney, R. *Democracy in the Third World*; Lynne Rienner, 1994; 190 pp.

International Interaction

1. Cingranelli, D.L., Ed. *Human Rights: Theory and Measurement*; MacMillan, 1988; 258 pp.
2. Hanski, R.; Suksi, M., Eds. *An Introduction to the International Protection of Human Rights: A Textbook*; Institute for Human Rights, 1997; 411 pp.
3. Hill, D., Ed. *Human Rights and Foreign Policy: Principles and Practice*; St. Martin's, 1989; 223 pp.

Merit Treatment and Minority Rights

Domestic U.S. Concerns

1. Darby, M., Ed. *Reducing Poverty in America: Views and Approaches*; Sage, 1996; 390 pp.
2. McFate, K.; Lawson, R.; Wilson, W., Eds. *Poverty, Inequality, and the Future of Social Policy: Western States in the New World Order*; Russell Sage, 1995; 768 pp.
3. Palley, M.L.; Preston, M.B., Eds. *Race, Sex, and Policy Problems*; Lexington, 1979; 271 pp.

Cross-National Comparisons

1. Davies, M.R.; Greenwood, J.; Robins, L.; Walkley, N., Eds. *Serving the State: Global Public Administration Education and Training*; Ashgate, 1998; 224 pp.
2. Gurr, T. *Minorities at Risk: A Global View of Ethnopolitical Conflicts*; U.S. Institute of Peace, 1993; 448 pp.
3. Oyen, E.; Miller, S.M.; Samad, S., Eds. *Poverty: A Global Review, Handbook on International Poverty Research*; Scandinavian, 1996; 630 pp.

International Interaction

1. Deacon, B.; Hulse, M.; Stubbs, P. *Global Social Policy: International Organizations and the Future of Welfare*; Sage, 1997; 265 pp.
2. Teitelbaum, M.; Weiner, M., Eds. *Threatened Peoples, Threatened Borders: World Migration and U.S. Policy*; Norton, 1995; 336 pp.

Win-Win Policy

Stuart S. Nagel

Political Science Department, University of Illinois, Urbana, Illinois, U.S.A.

INTRODUCTION

Win-win policy refers to public policy that is capable of achieving conservative and liberal goals simultaneously. Examples could be given from any field of public policy, such as economic, social, environmental, legal, or political policy. Win-win policies should be distinguished from compromises, where both sides retreat partially from achieving their goals in order to obtain an agreement.

AN EXAMPLE IN THE FIELD OF ENVIRONMENTAL POLICY

As an example, the field of environmental policy involves conservative and liberal approaches. Conservatives emphasize the role of consumers and the marketplace in restraining business from engaging in socially undesirable activities, like pollution. The liberals emphasize the role of the government in restraining pollution. Conservatives are especially interested in the goal of economic development, which may be interfered with by governmental restraints. Liberals are especially interested in the goal of a cleaner environment, which may not be achieved so effectively by relying on selective consumer buying.

A neutral compromise approach might involve giving business firms partial subsidies to adopt antipollution devices. Doing so would involve some requirements for receiving the subsidies but less interference than regulation and fines. Doing so would help promote a cleaner environment, but there still might be evasions by business in view of the extra expense and trouble in complying.

A win-win policy alternative instead might emphasize subsidies to universities and research firms to develop new processes (that relate to manufacturing, transportation, energy, and agriculture) that are less expensive and cleaner than the old processes. Those new processes then would be adopted by business firms, because they are more profitable, not because the firms are being forced or subsidized to do so.

The new processes thus would achieve the conservative goals of profits and economic development, even better than retaining the present marketplace. Such a win-win policy also would promote the liberal goal

of a cleaner environment, even better than a system of regulation, and without the expense of a continuing subsidy for adopting and renewing antipollution devices.

A specific example of such an environmental win-win policy has been finding a substitute for aerosol propellants and air-conditioning freon that is more profitable to manufacturers and simultaneously less harmful to the ozone layer, which protects against skin cancer. Another specific example is developing an electric car, which saves money on gasoline and maintenance, while at the same time, does not generate the exhaust pollution of internal-combustion cars. Developing hydrogen fusion or solar energy also may be examples of less expensive and cleaner fuels for manufacturing processes.

KINDS OF FEASIBILITY

Win-win policies may be capable of achieving conservative and liberal goals in theory but not in practice. To be meaningful policies, they may have to satisfy various kinds of feasibility. For example, is the policy of developing solar energy for manufacturing feasible technologically? Is there sufficient funding available to subsidize the needed research if the private sector is not so willing, due to the risks, the large amount of money needed or the long wait before payoffs occur?

Is there insurmountable political opposition from liberals who do not like government subsidies to make business more profitable? What about the opposition from conservatives who do not like government involvement in developing a research agenda? Is the program feasible administratively in terms of built-in incentives, or does it require a lot of obtrusive monitoring? Does the program violate some constitutional rights? Does the program make provision for workers and firms that might be displaced if the policy is adopted?

CREATIVITY

Win-win policies may involve creativity, but developing them is becoming easier as a result of experience

with the ideas. We now have many different approaches that can serve as a checklist in leading one to a win-win policy. For example, expanding the resources available can enable conservatives to have more money for defense and liberals to have more money for domestic programs. The government sometimes can be a third-party benefactor in providing vouchers to enable landlords and tenants, and also merchants and consumers, to come out ahead.

One also can deal with problems like abortion by getting at the causes. These causes consist of unwanted pregnancies that could be lessened through more effective abstinence programs and birth control. Thinking in terms of the goals to be achieved, rather than the alternatives to choose among, can stimulate win-win policies. Thinking in terms of increasing benefits and decreasing costs can also lead to win-win policies.

Other approaches deal with early socialization of widely accepted values, technological fixes like the non-polluting hair spray, and the contracting out of government activities to private firms (which do well as a result of the profit motive and quality specifications in the contract). Further approaches may involve combining (rather than compromising) alternatives, developing a package with something for each major viewpoint, having international economic communities, adopting a gradual or incremental win-win policy, and arranging for big benefits on one side with small costs on the other.

Win-win policy may be especially relevant as of 1994–1995, because the Democrats control the presidency and the Republicans control Congress. That division means that neither purely liberal nor purely conservative legislation is likely to pass in view of the Republican congressional majority, the availability of the filibuster to the Democrats in Congress, and the availability of the veto to the president. It is possible that some compromise legislation might pass, such as health care without an employer mandate or required forms of health insurance. Even better would be win-win policies, such as promoting real economic growth through greater productivity related to increasing labor skills, new technologies, competition, and tariff reductions.

STEPS AND CHECKLISTS

Win-win policy analysis in this context can be defined as handling policy problems by finding solutions that exceed the best initial expectations of conservatives, liberals, Republicans, Democrats, or whoever are the major groups, sides, or viewpoints in the policy dispute. Win-win is also called superoptimizing or doing better than the previous best of all major groups.

There are basically five steps to win-win policy analysis:

1. What are the major goals of conservatives, liberals, or other major groups who are disputing what policy should be adopted for a given policy problem?
2. What are the major alternatives of these groups for dealing with the policy problem?
3. What are the relations between each major alternative and each major goal? In their simplest form, these relations can be expressed in terms of a minus sign (relatively adverse relation) and plus sign (relative conducive relation), and a zero (neither adverse nor conducive relation).
4. What new alternative is there that might be capable of the following:
 - a. Achieving the conservative goals even better than the conservative alternative.
 - b. Simultaneously capable of achieving the liberal goals even more than the liberal alternative. Whatever new alternative meets these two criteria is a win-win alternative or a superoptimum solution.
5. Is the proposed win-win alternative capable of getting over various hurdles that frequently exist. These hurdles may be political, administrative, technological, legal, psychological, and economic in random order. Win-win solutions should also consider how to upgrade workers and firms that may be displaced by downsizing due to increased productivity, free trade, defense conversion, immigration, merit treatment, labor utilization, creativity, and related factors.

To facilitate developing win-win alternatives, it helps to have a positive can-do attitude that it can be done. It also helps to have a checklist of types of win-win alternatives that have worked in the past. Such a facilitating checklist of types might include the following:

1. Expand the resources.
2. Find a third-party benefactor.
3. Set higher goals.
4. Decrease the causes of the problem.
5. Redefine the problem to emphasize goals rather than alternatives.
6. Increase the benefits and decrease the costs.
7. Socialize children in widely accepted values so the problems do not occur.
8. Find a new technology.
9. Contract out via an auction to multiple firms with societal strings attached.

10. Promote international economic communities.
11. Arrange for big benefits on one side and small costs on the other.
12. Fully combine alternatives that are not mutually exclusive.
13. Develop a multifaceted package.
14. Adopt the win-win solution in steps, where the first step may be a traditional compromise.

Win-win analysis can be applied to handling dissident faculty as an example:

1. A lose-lose situation is where a department or the university seeks to destroy a dissident faculty member. The university may suffer the stigma of being enjoined for violating free speech and due process. The faculty member loses money in obtaining an injunction and subsequent enforcement.
2. A win-lose situation is where one side thinks it can be a winner and the other side a loser. Trying to reach this point often results in a lose-lose situation, as in litigation, strikes, and war.
3. A win-win situation is where the university can try to constructively encourage the potential creativity of dissident faculty. Doing so can possibly result in new ideas, inventions, and institutions to the credit of the university and the dissident faculty members. This is more likely to happen with administrators who have flexible democratic personalities rather than authoritarian personalities. All administrators, however, benefit from the funding, prestige, and quality students that go to universities that are in the forefront of new ideas, inventions, and institutional arrangements.
4. The same win-win thinking can be applied to dissident citizens in a country. The analysis is also applicable to other economic, technology, social, political, international, and legal policy problems.

WIN-WIN MATRIX

Finding solutions where all major sides achieve better than their best expectations is the goal of a win-win policy.

The win-win matrix in Table 1 indicates that basic win-win analysis is about as simple as a tic-tac-toe table. A tic-tac-toe table has three columns made by two vertical lines. It also has three rows made by two horizontal lines. That is also true of a win-win table, except it adds a fourth row for superoptimum solutions.

Table 1 Generic win-win matrix

Policies	Goals	
	Conservatives	Liberals
Conservatives	+ ^a	—
Liberals	— ^b	+
Superoptimum Solutions	++ ^c	++

^a+ = The alternative is relatively conducive to the goal.

^b— = The alternative is relatively adverse to the goal.

^c++ = The best of the alternatives on the goal after the superoptimum solution alternative is considered.

The first row indicates the types of goals to be achieved. The second column refers to conservative goals, and the third column refers to liberal goals. The first column indicates the types of policy alternatives under consideration for achieving the goals. The second row refers to conservative policies, and the third row refers to liberal policies.

A “+” sign indicates that the policy is relatively conducive to the goal in comparison with the other policies being considered. A “—” sign indicates that the policy is relatively adverse to the goal. Conservative policies usually do well on conservative goals but not so well on liberal goals, as one would expect. Liberal policies do well on liberal goals but not so well on conservative goals.

This is a classic trade-off, where adopting either of the two policies means losing some of the goals. The result is usually a compromise, whereby both sides (or all major sides) lose something and gain something. Neither side is usually strong enough to impose its solutions.

A win-win or superoptimum solution involves finding an alternative that is not purely conservative, purely liberal, or a compromise. It is, however, capable of achieving the conservative goals more than the conservative policies, and simultaneously capable of achieving the liberal goals more than the liberal policies.

The superoptimum solution thus receives a “++” on the conservative and liberal goals. In that sense, it does better than the conservative best and better than the liberal best. It should also be feasible regarding economic, technology, social, psychological, political, administrative, and legal matters.

AN EXAMPLE OF THE MATRIX

Table 2 applies the generic win-win matrix to the problem of environmental protection, especially in developing nations. The main conservative goal is economic development, and the main liberal goal is a clean environment. Conservatives also endorse a clean environment but not as much as liberals. Likewise,

Table 2 Economic development versus a clean environment

Alternatives	Criteria	
	Conservative Economic development	Liberal Clean environment
<i>Conservative</i> Marketplace	+ ^a	–
<i>Liberal</i> Antipollution regulation	– ^b	+
<i>Neutral</i> Compromise regulations	0 ^c	0
<i>Superoptimum solution</i> Improved manufacturing, agricultural, and other processes (more profitable and cleaner)	++ ^d	++

^a+ = The best of the alternatives on the goal before the superoptimum alternative is considered.

^b– = The alternative is relatively adverse to the goal.

^c0 = Neither conducive nor adverse.

^d++ = The best of the alternatives on the goal after the superoptimum solution alternative is considered.

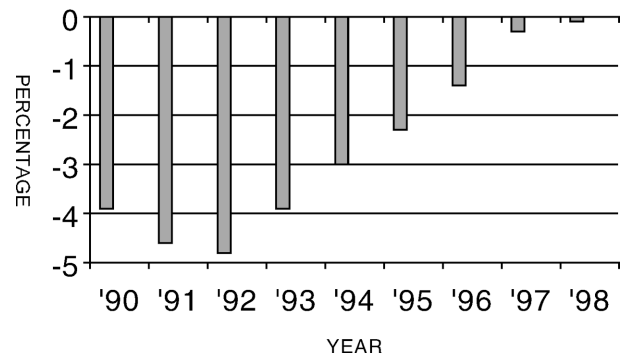
liberals also endorse economic development, but not as much as conservatives.

The main conservative policy is to rely on the marketplace on the theory that consumers would not buy products that are causing unhealthful pollution. The main liberal policy is to rely on antipollution regulation on the theory that the fines, taxes, and other penalties will cause business to reduce pollution.

The usual result consists of compromise regulations that are not as severe as liberals would like but are better than nothing. The compromise regulations are not as desirable to business as no regulation, but they are better than what the liberals are advocating.

A win-win or superoptimum solution might involve public policy working through financial and other positive incentives to generate new processes for manufacturing, agriculture, transportation, and energy that are more profitable to business but at the same time cleaner to the environment than previous processes. Such a profitable and clean policy would be an improvement on economic development and a clean environment.

An example might include the use of vehicles and machines that use less gasoline and oil but are even more productive. Such innovations are easier in terms of technology, funding, and other considerations for some kinds of pollution. Such innovations may also have a bigger impact on some kinds of pollution. The optimum allocation of incentive money is, however, a separate problem. The important consideration here is shifting conservatives and liberals toward such

**Fig. 1** Federal budget deficit as a percent of gross domestic product.

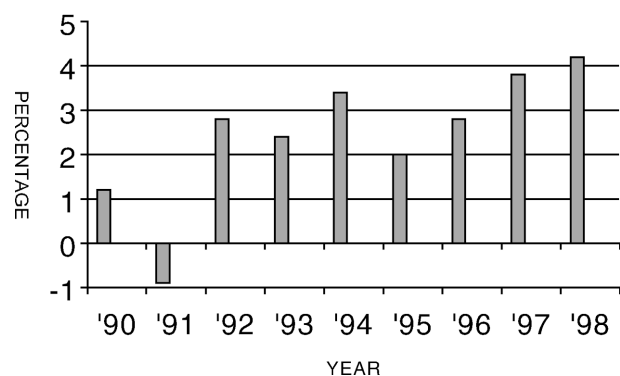
win-win solutions, rather than toward relatively ineffective marketplace and regulatory policies.

SUEROPTIMUM SOLUTIONS APPLIED TO ECONOMICS

Economists (especially Keynesian economists) have argued that an expanding economy requires (or generally requires) deficit spending in order to stimulate the economy. If the government spends more than it taxes (an undesirable occurrence), then gross national product (GNP) goes up (a desirable occurrence).

Deficits are undesirable, because they cause government borrowing which raises interest rates and hurts the economy; inflexibility in needed government spending, because so much of the budget goes to pay interest on the debt; and a class of relatively unproductive people who live off unnecessary interest paid by taxpayers. An increased real GNP is desirable, because it means more jobs, higher wages, more funds available for education and other worthwhile government programs, and improved quality of life via private purchasing.

Traditional trade-off thinking says we must accept undesirable costs to get desirable benefits. Win-win

**Fig. 2** Percent change in gross domestic product.

I. Inputs						
1. Goals	Y1	Y2				
2. Alternatives	X1	X2				
3. Relations	e.g., 1-5 scale					
II. Processing						
		GOALS		III. OUTPUTS TOTALS		
Weights =		3	1	W=1	W=2	W=3
		Y1	Y2			
ALTERNATIVES	X1	4	2	6	10	14
	X2	3	5	8	11	14

Fig. 3 Basic concepts in general policy analysis.

thinking says we can avoid the undesirable costs of big deficits and still have economic growth. Fig. 1 shows deficits turning into surpluses as of 1998. Fig. 2 indirectly shows recent GNP growth from under \$5 trillion in 1990 to over \$7 trillion in 1998, at an increasing rate of growth.

This win-win occurrence comes about through improved technology, training, competition, free trade, and public policy. Those factors cause increased GNP. An increased GNP enables the government to have increased revenue from income taxes and have decreased welfare costs. Those two factors lower the deficit, especially as a percent of the GNP.

Readers should not think that win-win analysis is biased against liberal economists like Keynes. For example, the Phillips Curve is associated with

conservatives who say we have to suffer unemployment in order to avoid inflation. The autumn 1996 issue of *Policy Evaluation* shows unemployment decreasing in the 1990s to a generation low of 4.3% and inflation simultaneously decreasing to under 1%, which is almost nothing.

COMPARED TO GENERAL POLICY ANALYSIS

The following should be considered when viewing the basic concepts in general policy analysis as shown in Fig. 3:

1. It may be highly questionable as to whether the policy analyst has used alternatives, goals,

I. INPUTS			
1. <u>Goals</u>	C, L, N		
2. <u>Alternatives</u>	C, L, N		
3. <u>Relations</u>	Just +, 0, or -		
II. PROCESSING			
		GOALS	
		C	L
ALTERNATIVES	C ^a	b	d
	L ^a	+	-
	N ^a	-	+
	0 ^c	0	0
III. OUTPUTS			
		SOS ^a	e
		++	++

^a C = conservative ; L = liberal ; N = neutral ; SOS = superoptimum solution.

^b + = The alternative is relatively conducive to the goal.

^c 0 = Neither conducive nor adverse.

^d - = The alternative is relatively adverse to the goal.

^e ++ = The best of the alternatives on the goal after the superoptimum solution alternative is considered.

Fig. 4 Basic concepts in win-win or superoptimum solution analysis.

weights for the goals, and relations between alternatives and goals, on which large-scale agreement can be obtained.

2. The results are highly sensitive to the subjective alternatives, goals, weights, and relations.
3. The totals to the right are calculated with Y2 having a weight of 1, and Y1 having a weight of 1, 2, or 3.
4. Processing steps for Figs. 1 and 2 are as follows:
 - a. Decide on winning alternative:
 - i. Which one is best?
 - ii. Which combination is best?
 - iii. What is the best allocation?
 - b. Hurdles to overcome include economic, political, administrative, legal, psychological, technology, displaced workers, and firms.
 - c. "What if" analysis.
2. It is logically assumed that conservative alternatives have a relatively positive relation with conservative goals and a relatively negative relation with liberal goals, and vice versa for liberal alternatives. It is also logically assumed that conservatives give greater weight to conservative goals than to liberal goals, and vice versa for liberals.
3. With that reasonably objective information, one can deduce alternatives that are objectively win-win. If the win-win alternative dominates both sets of goals, then it makes no difference how the goals are weighted or totaled.

Basic concepts in win-win or superoptimum solution analysis are shown in Fig. 4. The following should be noted:

1. The alternatives, goals, and weights for the goals are given by conservatives and liberals in congress, the courts, the newspapers, or elsewhere, rather than imposed by the policy analyst.

BIBLIOGRAPHY

1. Baumol, W. *Superfairness: Applications and Theory*; MIT Press, 1986.
2. Nagel, S. *Super-Optimum Solution and Win-Win Policy: Basic Concepts and Methods*; Greenwood, 1997.
3. Nagel, S.; Eckart, S. *Handbook of Win-Win Policy Analysis*; Nova Science, 2001.
4. Noyes, R. *Now the Synthesis: Capitalism, Socialism, and the New Social Contract*; Centre for Incentive Taxation, 1991.
5. Susskind, L.; Cruikshank, J. *Breaking the Impasse: Consensual Approaches to Resolving Public Disputes*; Basic Books, 1987.

Work Motivation

Hal G. Rainey

Department of Political Science, University of Georgia, Athens, Georgia, U.S.A.

INTRODUCTION

Motivation is one of the central topics in the social sciences, and work motivation plays a similarly central role in the research and practice of administration. Work motivation has received intensive theoretical development, and scholars have reported thousands of studies bearing on the topic. Scholars still struggle with definition, and reviews of the research have located over 100 distinct definitions. For present purposes, however, we can say that work motivation refers to how much a person tries to work hard and work well—to the arousal, direction, and persistence of effort in work settings.

MEASURING MOTIVATION

Motivation researchers have also struggled with different ways of measuring motivation, such as through questions on questionnaires asking people about how hard they work and how often they do some extra work. Generally, efforts to employ measurement methods such as these have not led to a consensus about a most effective procedure for measuring motivation. Researchers also employ scales of intrinsic or internal work motivation. Intrinsic work motives or rewards include those mediated within the worker, such as feelings of accomplishment, growth, and self-esteem through work well done. Extrinsic rewards are externally mediated, and are exemplified by salary, promotion, and other rewards that come from the organization or work group.

Additional Influences on Performance

Writers on work motivation also emphasize that motivation alone does not determine performance. Other factors influence performance, including ability, understanding of the task, behaviors of leaders or coworkers that can confuse or clarify, the availability of resources, and the adequacy of training. This seemingly simple point takes on great importance when superiors inaccurately attribute performance problems to subordinates' weak motivation, when in reality, the problems stem from some of these other causes.

Motivation as an Umbrella Concept

Some scholars have concluded that these complexities have moved the concept of motivation into the status of an umbrella concept that refers to a general topic rather than a precisely defined and measured research variable. They employ the term as referring to a general concept incorporating many variables and issues, such as organizational commitment, leadership practices, job involvement and intrinsic work motivation, organizational climate and culture, characteristics of work goals, and others.

THEORIES OF MOTIVATION

Theorists have developed many theories of work motivation. None explains motivation comprehensively, but each contributes an important component of a well-developed conception of motivation and its determinants and outcomes. The following examples illustrate the most prominent theories. They include the group of theories often called the content theories, that emphasize the needs and motives that serve as contents of the motivational process. This group includes Maslow's need hierarchy, Herzberg's two-factor theory, McClelland's theory of the need for achievement, and equity theory. Process theories, in the second category, do not posit particular needs or motives but rather model the psychological and behavioral processes in motivation. They include expectancy theory, operant conditioning theory, social learning theory, and goal-setting theory.

Maslow's Need Hierarchy

The psychologist Abraham Maslow proposed an influential theory—human motives or needs follow a hierarchy of prepotency, as he put it. Basic, lower-order needs dominate human motivation until they are satisfied. Then, needs at the next higher level dominate, and so on up the hierarchy to the higher-order needs, such as the needs for esteem and ultimately for self-actualization. Some details help illustrate the hierarchy. Once one has satisfied such physiological needs as hunger and the need for sleep, one becomes more concerned with the next step up the hierarchy, the

safety and security motives. Next, needs for social and love relationships dominate. With those needs reasonably satisfied, needs for esteem dominate. The highest-order need, for self-actualization, or the striving to maximize one's potential and fulfill aspirations, became Maslow's most influential concept. In his writings, Maslow argued that genuinely self-actualized persons achieve this ultimate mode of satisfaction through dedication to a duty or mission that benefits others or society in general.

Herzberg's Two-Factor Theory

Frederick Herzberg (1968) proposed one of the best-known analyses of motivational issues in his two-factor theory. From many studies of workers in organizations, he concluded that two major factors influence individual motivation in work settings—"motivators" and "hygiene factors." The distinction is critical. The absence or insufficiency of hygiene factors can contribute to dissatisfaction with the job, but the presence of hygiene factors does not stimulate high levels of satisfaction. Hygiene factors are extrinsic to the individual and include organizational, group, or supervisory conditions, or externally mediated rewards such as salary. While hygiene factors can only prevent dissatisfaction, the motivators produce a heightened level of satisfaction and increased motivation. Motivators are intrinsic to the job and include interest and enjoyment of the work, as well as a sense of growth, achievement, and fulfillment of other higher-order needs. Herzberg concluded that because motivators provide the real sources of stimulation and motivation for employees, managers must arrange job settings to provide for the growth, achievement, recognition, and other needs represented in the motivators.

McClelland: Needs for Achievement, Power, and Affiliation

David McClelland's theory about motivations for achievement, power, and affiliation has inspired thousands of studies. Need for achievement, the central concept in his theory, refers to a motivation to achieve a sense of mastery over one's environment, through success at achieving goals and outcomes through one's own cunning, abilities, and efforts. McClelland originally argued that motivation for achievement in this pattern was characteristic of persons attracted to managerial and entrepreneurial roles, but he later narrowed its application to predicting success in entrepreneurial roles. McClelland measured need for achievement through a variety of procedures, such as psychological tests and questionnaire measures. He concluded that persons high in this need choose

reasonably challenging goals and moderate risks, where outcomes are fairly clear and accomplishment reflects success through one's own abilities. McClelland claimed that need for achievement related to success in certain types of business activities and even to economic success of nations.

Equity Theory

J. Stacy Adams developed a theory about a motive for equity in organizations that makes people want to feel equitably compensated. People have a need for a sense of equity and are uncomfortable with evidence that they are inequitably compensated. They compare their own exchanges with the organization to the exchanges between the organization and other employees. If a person senses that treatment received is inequitable in comparison with others, that person will be motivated to reduce the inequity. The person may respond to inequity in various ways—by altering inputs or outcomes, mentally distorting one's inputs and outcomes, acting on the referent person (such as pressuring the person to produce more or less), or leaving by transferring or quitting.

Expectancy Theory

This theory of work motivation draws on the classic observation, attributable to various philosophers and social scientists, that humans seek to do what they think is most likely to maximize desirable results and to minimize bad results. The psychologist Victor Vroom brought this theory to prominence with an effort to formalize this time-honored observation into a mathematical formula. His formulation posited that an individual will be more strongly motivated to engage in a behavior as probabilities ("expectancies") that the behavior will lead to valued outcomes and avoid bad outcomes are perceived to be stronger. In other words, if a behavior is likely to lead to some desirable results (more pay, more interesting work) but also is likely to lead to some undesirable results (more overtime work, more stress), one's motivation to engage in the behavior will depend on just how desirable or undesirable one considers the results, and on one's sense of the likelihood of each result. Expectancy theory appeared to offer an advancement in conceiving and analyzing motivation. Organizational psychologists sought to apply it by developing questionnaires that ask employees about the likelihood of a set of outcomes (e.g., a pay raise, a promotion) from their work behaviors (such as "doing high-quality work"). The questionnaire also asked about their positive or negative valuations of the outcomes. The psychologists then developed a score for each

employee by multiplying the employee's perceived probabilities of each outcome by the employee's valuation of that outcome, and summing these products. They hoped this sum would serve as an indicator of that person's motivation.

Also fairly rapidly, problems in applying and validating the theory mounted, along with questions and criticisms. Researchers found the component constructs hard to clearly define and operationalize. All of these problems soon deflated any dramatic hopes for the theory, and theorists set about developing revised versions. They tend to state the theory as a general conceptual framework without clear postulates as to how the component constructs are supposed to be combined. Newer formulations also distinguish between *Expectancy I*, or the person's perceived probability that a given level of work effort will lead to a certain level of performance, and *Expectancy II*, the person's perceived probability that the attainment of that level of performance will lead to outcomes. These two expectancies, together with the values of the outcomes, determine motivation. The more recent versions also add other variables to the framework, such as task clarity, self-esteem, and performance assessment procedures, which can influence Expectancy I, and organizational compensation procedures, which can influence Expectancy II. In spite of the limitations of this theory, nonetheless, scholars are attempting to further develop motivation theories that typically employ expectancy theory concepts.

Operant Conditioning Theory and Behavior Modification

Theorists have also drawn on operant conditioning theory as a basis for analysis of work motivation. Operant theory derives from what psychologists have called the "behaviorist" school in psychology, which heavily emphasized the analysis of observable behaviors, and de-emphasized what goes on inside the individual. They analyzed the ways in which external influences "reinforce" or increase behaviors, and how, on the other hand, lack of reinforcement leads to "extinction" of the behavior. This interest in shaping behaviors provided the foundations for later development of behavior modification techniques in therapeutic, educational, and managerial applications. Programs of managerial applications have sometimes been referred to as organizational behavior modification programs.

The operant conditioning theorists developed concepts and principles of reinforcement that influenced later efforts to apply this type of theory in administrative settings, some of which have become

part of everyday parlance. Positive reinforcement applies a stimulus that increases the behavior on which it is made contingent. We receive payment for taking an action, and our tendency to take that action increases. Negative reinforcement increases the behavior on which it is made contingent through removal of the stimulus or condition. Operant theorists also distinguished between various patterns of reinforcement, such as how frequently and regularly a reinforcement is applied. They concluded that positive reinforcement provides the most efficient means of influencing behavior, and that punishment is less efficient and effective in shaping behavior. They usually concluded that a behavior is acquired most rapidly through frequent, regular reinforcement. A behavior reinforced with that schedule, however, will extinguish more rapidly once reinforcement is discontinued, than will a behavior that has been reinforced less frequently and regularly. This leads some behavior modification proponents to prescribe such managerial techniques as *not* praising a desired behavior constantly but on a varying basis after a number of repetitions.

This prescription serves as an example of those offered by practitioners of organizational behavior modification, which applies versions of operant conditioning theory in organizations. Organizational behavior modification often involves approaches such as measuring and recording desirable and undesirable behaviors to establish baselines, determining the antecedents and consequences of the behaviors, and determining a strategy for using reinforcements and punishments—such as praise and pay increases—to apply to the behaviors. In spite of some reports of successful programs, this approach has sharp critics and has not been widely applied.

Social Learning Theory Perspectives on Motivation

Albert Bandura has led the development of social learning theory that has influenced contemporary views of work motivation. The theory emphasizes forms of learning and behavioral change that are not tied tightly to some external reinforcement. Humans can engage in vicarious learning, anticipation, mental rehearsal and imagery, and self-rewarding behaviors (praising oneself). Applications of such processes to organizational settings have mainly involved the development of frameworks for leadership and self-improvement, using the processes mentioned above to improve performance and to enhance a person's self-efficacy. Self-efficacy refers to a person's task-specific confidence and sense of capability in accomplishing task outcomes.

Goal-Setting Theory

Edwin Locke and colleagues advanced a theory of goal setting that many reviewers have acknowledged as the most successful work motivation theory in gaining validation through well-designed research. The theory holds that difficult, specific goals to which the persons pursuing them feel a commitment, lead to higher task performance than easy or vague goals or no goals (e.g., “do your best”). The theory posits that difficult goals enhance performance by directing attention and action, mobilizing effort, increasing persistence, and motivating the search for effective performance strategies. Many studies support the basic tenets of the theory by finding that difficult, specific, accepted goals enhance task performance. One reason for the success of the theory, however, may be its compactness and relatively narrow focus on task accomplishment, often on relatively simple tasks.

RECENT THEORETICAL DIRECTIONS: INTEGRATION AND SEPARATION

Reviewers tend to agree that motivation theory is in a disorderly state. Recent proposals and initiatives for the further development of motivation theory emphasize the integration of existing theories, or they emphasize the separateness of theories, while recognizing how each theory is useful, or they emphasize some degree of integration and separateness. Recently, scholars proposed models of motivation that integrated elements of expectancy and goal-setting theory with elements of other psychological theories, such as control and attribution theory. Other recent models bring together major concepts related to motivation, including individual dispositions, needs, goals, expectancies, self-efficacy, fairness, rewards, social influences, and job design.

MOTIVATION PRACTICE AND TECHNIQUES

While theorists continue the search for theories and concepts of motivation, organizations need to motivate their members, and they take many approaches to doing so. Well-administered organizations develop well-designed performance appraisal systems. They seek well-designed compensation systems including, as appropriate, merit pay and pay for performance systems, skill-based pay, cafeteria-style benefit plans (that allow choices among benefits), work group or team-based incentives, and profit-sharing and gain-sharing programs. They utilize bonus and award systems that offer one-time awards for excellent performance or other achievements. They may use Management by Objectives and other performance targeting procedures. They recognize the value of participative management and decision making in motivating employees. They use work enhancement techniques for effective job design and enrichment. As appropriate, they initiate organizational development (OD) interventions that apply behavioral science techniques to improving such processes as communication, conflict resolution, and trust. An effectively managed organization also strives to develop an organizational culture and mission that motivates employees toward high performance.

BIBLIOGRAPHY

1. Erez, M., Kleinbeck, U., Thierry, H., Eds.; *Work Motivation in the Context of a Globalizing Economy*; L. Erlbaum Associates: Mahwah, NJ, 2001.
2. Mitchell, T.R. Matching motivational strategies with organizational contexts. *Res. Organ. Behav.* **1997**, *19*, 57–149.
3. Pinder, C.C. *Work Motivation in Organizational Behavior*; Prentice Hall: Upper Saddle River, NJ, 1998.

Working Capital

Justin Marlowe

Department of Public Administration, University of Kansas, Lawrence, Kansas, U.S.A.

INTRODUCTION

The long-term viability of any public organization is heavily dependent on that organization's effective use of working capital. Working capital provides operations managers with financial resources to meet short-term obligations, generate interest income through short term investments, take advantage of discounts and other limited-term procurement opportunities, and address many other day-to-day concerns. Moreover, sound working capital management practices can free up cash during tough fiscal times, thus mitigating the need for costly short-term borrowing. Working capital management is therefore one of the central concerns facing public financial managers. This chapter examines working capital from both an accounting and a financial managerial perspective.

THE ACCOUNTING PERSPECTIVE ON WORKING CAPITAL

The fundamental equation of accounting states that the difference between an organization's assets and liabilities is known as owners' equity.^[1] Since public sector organizations do not necessarily have owners or shareholders, the government and nonprofit variant of this basic equation labels the difference between assets and liabilities "net assets." This equation is expressed as

$$\text{Assets} - \text{Liabilities} = \text{Net Assets} \quad (1)$$

The difference between an organization's total assets and liabilities provides important insights into its ability to meet long-term obligations. Nonetheless, operations managers are primarily concerned with the organization's ability to meet current rather than long-term obligations. This emphasis on current financial condition is reflected in the prevailing governmental accounting model, which distinguishes current assets and liabilities from long-term assets and liabilities. Current assets such as cash, marketable securities, accounts receivable, and the liquid assets associated with or capital projects are accounted for in one of a number of "funds" such as the general fund, debt service fund, and capital projects fund. The difference between current assets and current

liabilities within each fund is known as "fund balance." This relationship is expressed as

$$\text{Assets} = \text{Liabilities} + \text{Fund Balance} \quad (2)$$

Because of the public sector accounting model's focus on current assets, fund balance is considered by many to be synonymous with working capital.^a Working capital is often cited as an important indicator of an organization's financial condition. Like many other indicators, such as the current ratio, it is often expressed as a percentage of current expenditures. An organization with a working capital ratio of 1:1 holds enough liquid assets to cover its current liabilities.^[3]

THE MANAGEMENT PERSPECTIVE ON WORKING CAPITAL

"Working capital management" is a term broadly construed to mean the management of current, short-term, or idle financial resources. The primary objective of a working capital management strategy is to maximize the productivity of working capital in accord with organizational objectives. There are three schools of thought on how to most effectively match resources with objectives.

The first suggests a sizable working capital reserve is necessary in order to take advantage of discounts and other short-term procurement opportunities, to prevent fees resulting from late payment of liabilities, and to protect against catastrophic losses resulting

^aThis relationship between fund balance and working capital is contingent upon an organization's choice of accounting methods. In organizations that use modified accrual accounting, fund balance represents the difference between current assets and liabilities and is thus a good indicator of available working capital. But many organizations use a cash, modified cash, budgetary, or some other comprehensive basis of accounting (OCBOA) method that focuses on resource inflows and outflows rather than current financial resources. In these organizations, working capital is synonymous with the available cash on hand, remaining appropriation, or some other measure of available financial resources. OCBOA methods typically do not receive nearly as much attention from accounting practitioners and academics as generally accepted accounting principles (GAAP), despite recent estimates that anywhere from one-third to two-thirds of all small municipal governments in the United States use OCBOA. For a further discussion of OCBOA and its related issues, see Ref.^[2].

from natural disasters and other unforeseeable events. Proponents of this view also suggest working capital provides organizations a pool of financial resources that can be used to meet debt service obligations during tough fiscal times. In turn, credit analysts look favorably upon large working capital reserves and reward effective working capital maintenance with higher bond ratings. Given these numerous advantages, organizations are encouraged to be liberal with the size and use of their working capital reserves.

An opposing perspective suggests working capital is a drag on productivity. Rather than contributing to the production of public goods and services, it instead conceals stale, unmarketable inventory, accounts receivable that may not be realized, unnecessarily quick payment of outstanding liabilities, and other slack resources should be minimized or eliminated whenever possible. Proponents of this view recognize certain business processes inevitably require some working capital, but are also quick to argue any and all idle working capital reserves are most productive when placed in short-term investments. Therefore, they claim, organizations should minimize and invest working capital reserves whenever possible.

Most public organizations, due to their missions and responsibilities, cannot afford wholesale implementation of either of these strategies. The fact that many public organizations must be prepared for worst-case scenarios such as natural disasters, terrorist attacks, fiscal downturns, and the like suggests that aggressive inventory reduction strategies may carry unacceptable levels of risk. Public organizations also rarely have the same breadth of choices among competing vendors, which greatly inhibits their ability to manage when and how goods and services will be delivered.

These and other considerations suggest a third or compromise working capital strategy characterized by the maintenance of a wide variety of working capital resources for an equally wide variety of strategic purposes. That strategy was best articulated by Shelton, Tyer, and Hembree,^[4] who suggest that public organizations ought to maintain four different types of working capital reserves:

1. Transaction balances—which allow organizations to realize economic gains by paying bills quickly and taking advantage of trade credits.
2. Compensating balances—which improve an organization's standing with banks and other financial institutions by increasing the amount of money those financial institutions have available for lending.
3. Speculative balances—which allow an organization to take advantage of near-term procurement and other opportunities.

4. Precautionary balances—which provide necessary protection against economic downturns and other fiscal shocks.

Public financial managers, according to Shelton and Tyer, are most concerned with speculative and precautionary balances. Speculative balances are often maintained in the form of reserved and designated fund balance. By designating resources for a particular use, managers can practice effective working capital management while ensuring transparent and accountable use of public funds. Precautionary balances are often equated to undesignated fund balance, which is largely a function of bond raters' stated desire for residual resources that can be used to meet debt service obligations during tough fiscal times.

WORKING CAPITAL CREATION AND MAINTENANCE

Although there are divergent perspectives on how to maximize working capital productivity, there is a general consensus that the amount of working capital available to an organization can be maximized through effective management of four key areas: cash, inventory, receivables, and payables. The central principal guiding working capital maximization for each of these areas is to “use someone else's money every chance you get and don't let anyone else use yours.” The key considerations for each of those four areas are briefly considered here.

One perspective emphasizes the importance of cash flow and has produced a number of sophisticated models and decision heuristics for cash management. Those models assign present values to resource allocation decisions which take into account the effect of individual decisions on cash flows and cash volume across the organization. Sagan^[5] highlights this fact in one of the first discussions of working capital management by claiming “From an operational point of view, the money manager's primary concern is with the current cash flows and those flows expected in the near future. He can take little comfort in a satisfactory working capital ratio when he does not have funds to meet an immediately due payment.” This is especially true in public organizations, where cash flows from tax collections, donations, government grants, intergovernmental revenue sharing, and other typical revenue sources occur infrequently.

Another potential source of working capital is tighter management of receivables and payables. Corporate organizations can significantly increase their available working capital by forcing customers, distributors, and other payees to pay as quickly as possible while simultaneously delaying payment of their own

invoices for as long as possible. But public organizations typically cannot pressure taxpayers, donors, grantors, and other revenue providers into early payment without notable consequences. Instead, many public organizations have turned to lockboxes, prepayment envelopes, discounts for early payment, and other practices that promote rather than coerce quick collection of receivables. Public organizations also have less latitude than their corporate counterparts in their relationships with vendors. Given the sheer scope of services that public agencies provide, coupled with a general lack of available providers, public agencies are often not able to renegotiate contracts or threaten to take their business elsewhere when vendors are not willing to meet inventory management schedules. These and other considerations suggest effective management of receivables and payables can free up some, but perhaps only limited, working capital.

Inventory management is also a critical working capital consideration. Inventories are a sizable investment for any organization, but are a significant drag on productivity because they earn no interest. Current corporate practices, particularly those implemented by Dell Computer Corp. and other “New Economy” firms, suggest aggressive inventory reduction strategies can generate tremendous cost savings while freeing up sizable working capital reserves. Although many of these strategies are not available to public sector organizations (as previously mentioned), the general principle that minimizing standing inventory while maximizing the frequency of inventory turnover can affect the amount of available working capital.

SUGGESTIONS FOR FUTURE RESEARCH

Public sector working capital management is clearly in its early stages. To that end, future research should consider three recent trends that are likely to affect its future development.

First and foremost, researchers should consider the state of the practice of working capital in public organizations. Studies by Wolkoff,^[6] Shelton, Tyer, and Hembree,^[5] and Marlowe^[7] suggest that many municipal governments keep large working capital reserves but lack policies and procedures governing the use of those reserves. Marlowe’s study, for example, estimates that the total general fund balance for U.S. cities and villages alone may exceed \$46 billion. A rough estimate suggests that this sum is equal to 25–50% of the municipal debt outstanding to those same governments. Future research should examine how public organizations create and maintain working capital, and whether those practices coincide with recommended practice.

A second area of inquiry surrounds recently proposed changes in financial accounting and reporting

of working capital. There is some ambiguity under the current Governmental Accounting Standards Board (GASB) model regarding the use and effects of fund balance and net assets. That model clearly states reserved fund balance, for example, ought to be legally restricted for some particular purpose where unreserved undesignated fund balance is available for appropriation. But entities are given much more latitude surrounding their use of designated fund balance. The GASB is currently considering a number of new accounting standards that will likely alter the potential uses for this key pool of working capital funds. Future research should consider the role of designated fund balance in working capital management, and how that role may be altered by changes to GASB standards.

And finally, researchers may also consider recent developments in the process of working capital management. The practice of working capital management clearly entails a wide and often contradictory set of objectives, practices, and skills. This tension was described by Sagan,^[5] who, in one of the earliest examinations of corporate working capital practice, claimed “The money manager should be a combination of a credit analyst, a commodity specialist, a money market expert, a bit of a horse trader, and a banker. As such he possesses a proper balance of art, science, and diplomacy. If he possesses these, together with an ability to read the future, his will be a successful career.” Sagan’s comments certainly place a premium on the “art” of working capital management, which is clearly that of simultaneously pursuing multiple goals under a variety of constraints.

While there is certainly no substitute for a seasoned working capital manager, recent developments in linear programming, goal optimization, and other areas of applied mathematics provide powerful tools to aid working capital decision-making. While these tools and techniques proceed from different assumptions and applications, they all provide decision makers with information about how to optimize available resources in the face of multiple goals and constraints. Linear programming approaches, such as those presented by Sartoris and Hill,^[8] Merville and Tavis,^[10] and Bierman, Chopra, and Thomas,^[9] have become increasingly popular as a means for optimizing working capital productivity and cash flow in the short term. Aman Khan^[11] provides multiple examples and potential applications of optimization models for government operations. These models are especially useful in the governmental context given the multiplicity of political actors and fragmented objectives that act upon public financial decisions. These models could certainly be applied in the working capital context, and could result in a better understanding of how to maximize working capital productivity while pursuing the previously mentioned strategic goals. Future

research should expand and develop applications of these sorts of models for working capital decisions.

CONCLUSIONS

This chapter has examined working capital, or the difference between current assets and liabilities, from both a technical and a managerial perspective. It suggests that working capital management in the public sector is a complex process, affected by a variety of often contradictory goals and objectives. It also discussed how that complexity is reflected in current public sector accounting and reporting standards, a lack of broadly accepted working capital policies and practices, and recent attempts to manage that complexity through linear programming and other mathematical models. As public sector organizations face increasing fiscal pressures, effective working capital theory, management, and technology are likely to add much value to public management.

REFERENCES

1. Granof, M.H. *Government and Not-for-Profit Accounting*; John Wiley & Sons: New York, 2001; 33–36.
2. Crawford, M. *Applying OCBOA in State and Local Governmental Financial Statements*; American Institute of Certified Public Accountants: New York, 2003.
3. Finkler, S.A. *Financial Management for Public, Health, and Not-for-Profit Organizations*; Prentice Hall: Upper Saddle River, NJ, 2001; 276–277.
4. Shelton, M.; Tyler, C.; Hembree, H. Local government reserve funds and fund balance: some applications of business concepts. *Munic. Finance J.* **2000**, *21* (1), 1–18.
5. Sagan, J. Toward a theory of working capital management. *J. Finance* **1955**, *10* (2), 121–129.
6. Wolkoff, M. An evaluation of municipal rainy day funds. *Public Budg. Finance* **1987**, *7* (2), 52–63.
7. Marlowe, J. *The Local Fund Balance: Expectations and Implications*. Unpublished Doctoral Dissertation; University of Wisconsin—Milwaukee, 2004.
8. Sartoris, W.L.; Hill, N.C. A generalized cash flow approach to short-term financial decisions. *J. Finance* **1983**, *38* (2), 349–360.
9. Bierman, H.; Chopra, K.; Thomas, J. Ruin considerations: optimal working capital and capital structure. *J. Financ. Quant. Anal.* **1975**, *10* (1), 119–128.
10. Merville, L.J.; Tavis, L.A. Optimal working capital policies: a chance constrained programming approach. *J. Financ. Quant. Anal.* **1973**, *8* (1), 47–59.
11. Khan, A. *Costs and Optimization in Government*; Quorum Books: Westport, CT, 2000.

Yield Curve for Public-Sector Debt Instruments

Theo Edwin Maloy

*Department of History and Political Science, West Texas A&M University,
Canyon, Texas, U.S.A.*

INTRODUCTION

Debt is created when a government, firm, or individual borrows money from another government, firm, or individual. The borrower gains immediate buying power in exchange for a promise of repayment, while the lender delays buying power in exchange for periodic or lump sum of interest payments in addition to the promised principal repayment. Depending on negotiated terms, the borrower repays the principal to the lender in scheduled periodic payments or in a lump sum on the due date, called the maturity.

Governments and firms primarily borrow money when annual expenditures exceed revenues, to spread the cost of long-term capital programs uniformly over the duration of the program, or when immediately due obligations exceed the short-term cash in the treasury. Individuals primarily borrow money for large expenditures that exceed cash in current liquid accounts, and when short-term cash flow does not meet current expenditures.

Governments, firms, and individuals borrow money for varying lengths of time. Lenders face varying degrees of repayment risk based on the financial conditions of the borrowers, of the financial markets, and of the debt instruments' specific terms. Given that different borrowers and debt instruments have different risks and maturities, different debt agreements can only be reached with different interest rates. At any given time, those interest rates can be charted for a particular borrower, such as the federal government or a low-risk or high-risk corporate borrower, and for different maturities, from one month to thirty years. The chart, with interest rates on one axis and maturities on the other, is called the "yield curve" because interest payments are called "yields" by investors.

MECHANICS OF BOND VALUES

The majority of government borrowing is through the sale of bonds. A bond is a contract to repay the borrowed principal upon a specific date, the maturity, in addition to making interest payments.^[1] Historically, bonds were paper certificates with coupons attached that were redeemed for interest payments. Today,

almost all bonds are issued in non-physical "book entry" form, with the issuing financial entity keeping records of all bonds and bondholders in their computer databases. Instead of receiving cash in exchange for coupons, bondholders receive their interest by direct deposit.

Federal government bonds are issued in multiples of \$1000 for maturities of up to 20 years. Governments, including the federal government, use different, but similar instruments to borrow for periods of less than one year. Some financial instruments are issued in up to \$10,000 units of face value.

Interest, the periodic fee for the use of borrowed money, is usually stated as a percentage of the face value of a bond. This yearly percentage "coupon rate" is usually divided into semiannual payments.

Although the federal government no longer issues bonds with maturities longer than 20 years, other governments and private firms issue bonds with maturities of 30 years or longer. Because of the long maturities, many bond buyers ("lenders") find it necessary to sell the bond before maturity. While bonds are initially issued in the "primary" market, there is an active "secondary" market where previously issued bonds are bought and sold by institutions and individuals.

If interest rates have risen in financial markets, the seller may receive less than the face value for the bond. If interest rates have fallen, the seller may receive more than the face value for the bond.

The market price of a bond is the present value of all cash flows at an appropriate discount rate.

A bond's present value is calculated using an elementary formula where the future value (the future cash flow) is divided by one plus the interest rate raised to the power of the number of years until maturity:

$$PV = \frac{FV}{(1 + r)^n}$$

This simplest case is called a "Zero Coupon Bond" because it has no coupon rate. The lender and the borrower agree to some principal amount and some interest rate. The lender then gives the borrower the present value of the face value when discounted at some interest rate for some number of years.

If the borrower promises to pay the lender \$1000 in twenty years with an interest rate of 5%, this is the calculation:

$$PV = \frac{\$1000}{(1.05)^{20}}$$

The divisor is 1.05 to the 20th power, or 2.653298. When \$1000 is divided by 2.653298, the present value is \$376.89.

The calculation is only slightly more complicated for a bond with yearly interest payments:

$$P = \sum_{i=1}^m \frac{FC}{(1+r)^i} + \frac{F}{(1+r)^m}$$

where P is the market value, the current cash price of a bond, M is the number of years until maturity, F is the face value of the bond, C is the coupon rate, and r is the current market interest rate on bonds of the same maturity and similar risk.

If interest is paid semiannually, then “ m ” would be doubled and “ r ” would be divided by 2. The periodic payment would also be divided by 2, because half the yearly interest is paid twice as often:

$$P = \sum_{i=1}^{2m} \frac{FC/2}{(1+r/2)^i} + \frac{F}{(1+r/2)^{2m}}$$

If a twenty-year 7% coupon-rate bond with a \$1000 face value had one year to maturity, this is the present value calculation if the current marketplace was requiring a 6% return for \$1000 one-year maturity bonds with the same risk:

$$1009.43 = \frac{70}{1.06} + \frac{1000}{1.06}$$

A buyer in the open market will pay \$1009.43 for a bond that appears to be “worth” \$1000. The reason that the price is more than \$1000 is because the \$70 coupon payment is more than the \$60 interest payment that a buyer could get on a new one-year \$1000 bond.

In the marketplace, the value of the bond is constantly adjusted to a price based on the current appropriate interest rate. If some event caused the appropriate interest discount rate to rise to 9%, the one-year bond would be valued at \$981.65:

$$981.65 = \frac{70}{1.09} + \frac{1000}{1.09}$$

If an existing bond’s market value, face value, coupon rate, and maturity are known, a similar equation

can be solved for the current market-required interest rate, which would be the discount rate.

Required Interest Rate

Lenders use several factors to determine any debt instrument’s appropriate required interest rate, which becomes the discount rate. The first component of any interest rate is the “risk-free rate of interest.” This is the interest rate a lender would require if there were absolutely no chance of default by the borrower. Historically, this rate has approximated to 2–3%. Short-term federal debt can be used to estimate the risk-free rate, and in early 2005 the interest rate on a one-month federal instrument fluctuated around 1.9–2.1%.

To calculate the interest that any particular borrower would have to pay, a “default-risk premium” is added to the risk-free rate. This is the premium for the possibility that the lender would not get all of the interest payments and the principal. It is also an insurance payment against the possibility that the interest payments and the principal will be paid but that they will be paid after the scheduled payment date. Many banks charge their “most credit-worthy” customers what they call their “prime rate.” A large corporation might pay an interest rate 1% above the risk-free rate. The loan would be about as safe as lending to the federal government, but not as safe. A state government might pay an interest rate 3% above the federal government; a local government might pay an interest rate 6% above the federal government. A state or local government might pay a lower interest rate if they pledge some specific cash flows, like highway toll payments, to repay the debt. If they pledged only their “full faith and credit,” a lender probably would require a higher interest rate. The default premium is specific to each lender and the lender’s collateral. To help lenders determine the probability of default, several financial institutions rate the danger of buying state, local, and special district debts. Lenders can also buy default insurance, and some borrowers buy the insurance for the lenders under the terms of their financial negotiations. Investors buying debt instruments in the secondary market also will base their discount rate on the debt’s risk and risk rating.

Lenders also require an “inflation premium” to make the purchasing power of interest and principal payments equal that of the original debt’s purchasing power. At the beginning of the millennium, inflation rates were quite low, but many investors remember 14% inflation rates in their lifetimes. The inflation premium will fluctuate with investors’ inflation expectations. In early 2005, the inflation premium on a 20-year federal bond appeared to be less than 3%.

Lenders may lend their money for any number of years, but they may not want to loan their money for the full term of a 20-year bond. Therefore, borrowers pay a “liquidity premium” to the interest rate to compensate for the possibility that their debt investment may not find an active market. A federal bond will have a ready market. Most state bonds will have a semiactive market. However, to sell an illiquid special district bond, an investor might have to accept a lower price than the bond’s risk would predict.

(Life insurance companies and other financial institutions buy bonds of all maturities, because they will need constant liquidity to pay death benefits. One way that a company might increase its cash flow would be to buy bonds that have large liquidity premiums. The bond will be held to maturity, so there will be no revenue loss because it is hard to find a buyer. The liquidity premium will be earned until maturity. However, a state or municipality cannot avoid paying a higher interest rate for an illiquid bond. If there is no liquidity premium, investors will buy other bonds. The liquidity premium will increase.)

Borrowers must also pay a “maturity-risk” or market premium. The maturity-risk premium relates the value of the debt instrument to the probability that the overall financial market will fluctuate. When a lender invests in a long-term debt asset, certain default and inflation assumptions exist in the market. However, those assumptions may change many times before the bond’s maturity. Because investor-specific conditions might cause a lender to sell debt investments during a time when interest rates have risen and the value of a specific bond has fallen, lenders will require a maturity-risk premium to compensate for potentially having to sell a bond when its price is low. The longer the term of a bond’s maturity, the higher the maturity-risk premium (Table 1).

The formula for the interest required from any particular debt instrument is

$$k_i = k^* + DP_i + IP_i + LP_i + MP_i$$

where “i” is the indication of each particular debt instrument.

The interest rate compensates the lender for delayed income and consumption. That is why the interest rate used for discounting in benefit–cost analysis is called

the “social rate of time preference.” The interest rate indicates how much an investor requires in interest for delayed consumption and for allowing government borrowers to consume today.

Because different lenders and borrowers have different social rates of time preference, different rates of risk tolerance, and different demands for current consumption, there is a large supply of long-term, intermediate-term, and long-term debt instruments of all degrees of risk. To determine the marketplace’s interest rate, inflation, and business condition expectations, current interest rates on bonds of all maturities can be charted into a yield curve. The United States Treasury Department publishes data to plot daily and historical yield curves.^[2]

Theoretically, a three-year interest rate is calculated by multiplying one plus the first year’s interest rate time one plus the second year’s interest rate, times one plus the third year’s interest rate, and taking the geometric cube root of the product (and then subtracting 1). Multiplying the 20 required interest rates plus 1 and then taking the geometric root (and subtracting one) would determine a 20-year maturity’s interest rate. However, as a future required interest rate is unknown, the yield curve is established by investors’ economic predictions. Conversely, the yield curve helps predict the economic future.

Yield Curve^[3]

Typically, market watchers and investors talk about interest rates going up or down as if all rates moved together. However, short-term and long-term bonds’ market prices often move independently in the marketplace, because economic conditions make some maturities less or more attractive to investors. The yield curve charts display various patterns, which can be interpreted as economic predictions and projections.

(During difficult economic times, investors might prefer high-quality, safe, but low yielding debt. During prosperous economic times, investors might prefer lower-quality bonds because they believe that the market is overestimating the debt’s default risk. Under some conditions, there might be separate yield curves for higher- and lower-quality debt.)

Ordinarily, short-term bonds of whatever risk will have lower yields that reflect long maturities’ additional risk. A “Normal Yield Curve” will slope upward to the right as maturities and yields increase. For federal government 20-year debt, the longer (right) end of the yield curve normally will be 2–3% higher than the shorter (left) end of the yield curve (Fig. 1).

When bond buyers expect the maintenance of economy at its normal pace without significant changes in inflation, the normal yield curve persists. However,

Table 1 Interest rate components

Real risk-free interest rate	k^*	2–3%
Default-risk premium	DP_i	1% or more
Inflation premium	IP_i	3% or more
Liquidity premium	LP_i	0% or more
Maturity-risk premium	MP_i	0% or more

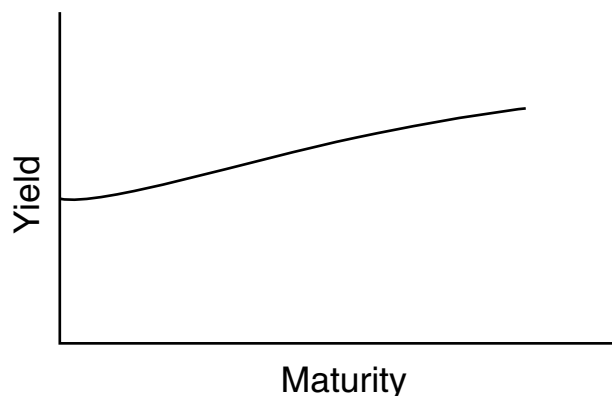


Fig. 1 Normal yield curve.

when bond investors believe that the economy will improve in the future, the yield curve will dramatically increase in slope (Fig. 2).

This shape is typical at the beginning of an economic expansion or at the end of an economic recession. Short-term interest rates are lowered by the economic recession, but long-term interest rates will rise as borrowers gather money for investment during the anticipated economic revival. Long-term lenders will demand a larger inflation premium, because economic expansions are often accompanied by increased inflation. During most economic expansions, with investor anticipation, the slope of the yield curve will increase even before the economy revives. While this is not a guarantee of economic improvement, a steepening yield curve is often accompanied by predictions of coming prosperity (Fig. 3).

Conversely, when investors believe that the economy is entering a recession, and that demand for borrowed money will decline, lowering interest rates, an inverted yield curve will be formed. Investors will anticipate that borrowers will soon reduce the supply of long-maturity debt. The investors will increase the supply of long-maturity money, lowering the price of long-maturity debt as debt sellers (borrowers) find

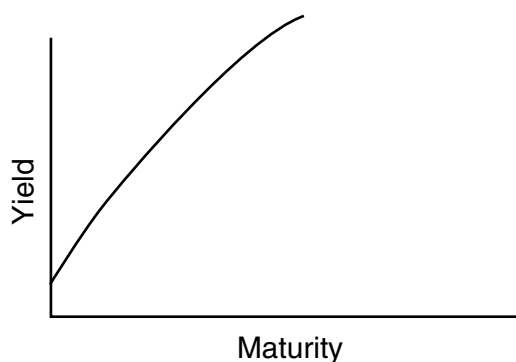


Fig. 2 Steep yield curve.

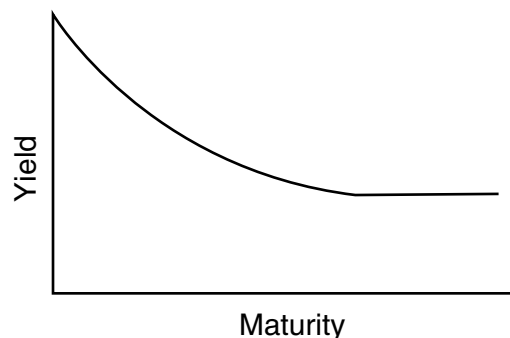


Fig. 3 Inverted yield curve.

more debt buyers (lenders) in the marketplace. Truly inverted yield curves are rare, always accompanied by predictions of an economic slowdown, and usually followed by an economic slump. Inverted yield curves are also harbinger of lower yields all along the yield curve (Figs. 4 and 5).

To become inverted, the yield curve must pass through a stage in which it is either flat or humped. Often flat or humped yield curves are interpreted as being inverted yield curves. Even flat or humped yield curves are indications of coming economic slowdowns, if not of economic slumps. For financial institutions, like banks, inverted, flat, or humped yield curves lead to lower interest income on loans.

Moving the Yield Curve

While the Federal Reserve controls only the short end of the yield curve, it is often given the credit (or the blame) for moving the curve or changing its shape. The Federal Reserve influences the short of the curve in an effort to achieve its policy goal of low and stable inflation and maximum sustainable output. A monetary policy tightening generates high nominal interest rates, but, because of its anti-inflationary effects, these rates quickly fall back. As long-term rates embed expectations of this behavior of short-term rates, they rise by only a small amount. As a result, the slope of



Fig. 4 Flat yield curve.

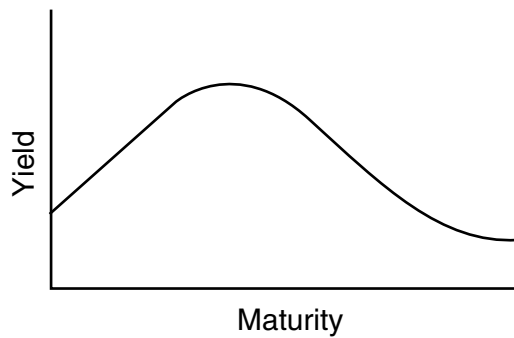


Fig. 5 Humped yield curve.

the yield curve declines when contradictory monetary policy shocks occur.^[4]

CONCLUSIONS

The sources of debt securities' supply and demand are well understood. It can be explained using elementary

mathematics as to how the marketplace prices bonds once a discount is determined. Empirical research reveals as to how lenders formulate their interest-yield discount-rate demands. The yield curve illustrates and explains this marketplace, but the yield curve is dynamic. It moves because of changing human behavior, but not mechanically. While the yield curve becomes a basis of economic prediction, the yield curve is not a determinant of the economic future. Macroeconomists face an interesting future as they attempt to explain the yield curve and its role in the macroeconomy.

REFERENCES

1. John, L.M. *Fiscal Administration: Analysis and Applications for the Public Sector*, 6th Ed.; Wadsworth/Thomson Learning, 2003.
2. <http://www.ustreas.gov/offices/domestic-finance/debt-management/interest-rate/yield.html>.
3. <http://fixedincome.fidelity.com/fi/FIHistoricalYield>.
4. <http://www.frbsf.org/publications/economics/letter/2003/el2003-15.pdf>.

Zero-Base Budgeting

J. L. Peters

*Department of Public Management, John Jay College of Criminal Justice,
City University of New York, New York, New York, U.S.A.*

INTRODUCTION

Zero-base budgeting (ZBB) rejects a previous year's budget as the basis for future resource allocation decisions. Like other performance budget approaches, ZBB seeks to maximize efficiency and eliminate waste and redundancy. But unlike other approaches, ZBB shifts attention away from a reliance on precedents established in previous budgets and focuses instead on the achievement of planned objectives. This shift is intended to emphasize the role of analysis as the basis for allocation decisions rather than subjective factors that might impede or obfuscate the process of arriving at an optimal resource distribution. Because ZBB abandons the premise that budgets are inherently incremental—that each new budget should be based upon the previous year's appropriation—it requires that every aspect of the budget be justified anew each year. As such, it has been considered by some as the most radical and controversial of all modern budget reforms.

THE ORIGINS OF ZBB

In 1940, Key Jr. asked, “On what basis should it be decided to allocate x dollars to activity A instead of activity B?”^[1] That question was his basis for arguing that analysis should play a key role in budgetary decision-making. Key asserted that while the budget reflects a set of value judgments as to how scarce resources should be allocated, traditional methods of budgeting offered no way to determine the social utility of those allocations, nor any means of evaluating the efficiency in achieving desired ends. Moreover, he argued, “If it is assumed that an agency is operating at maximum efficiency, the question remains whether the function is worth carrying out at all, or whether it should be carried out on a reduced or enlarged scale, with resulting transfers of funds to or from other activities of greater or less social utility.” (While Key did not specifically propose what eventually became known as performance budgeting, his article could definitely be seen as the catalyst for ensuing developments in that area.)

Lewis^[2] provided the fundamental premise for ZBB when he asserted that “to be justified in terms of making the most advantageous use of resources, the return from an expenditure for any activity must be more desirable and more necessary than for any alternative use of funds.” ZBB specifically requires a manager to justify an entire budget rather than just the increment being requested. In doing so, the manager is forced to redefine the “budget base to include decrements and not just increases in funding.”^[3]

ZBB first appeared in the federal government in 1962, but Burrows and Syme trace its origins all the way back to 1892, when Bastable argued against reliance on past budgets as the sole basis for future allocation decisions.^[4] The first US reference to ZBB-related concepts appeared in 1917, when Willoughby et al. published their comparative study of budgeting in Britain, France, and Germany.^[4] Early 20th century budget reformers on both sides of the Atlantic—including Higgs, Keynes, Durell, and Lewis—promoted a number of principles that were precursors to modern ZBB. Burrows and Syme credit Pyrrh with “claiming an apparently successful implementation” of what he referred to as “zero-base budgeting.” The term, however, was used by Wildavsky and Hammond several years earlier.^[5]

ZBB gained little attention until Pyrrh published his research on the use of ZBB in the private sector in the 1970s.^[6] Although the US Department of Agriculture had quickly abandoned its experiment with ZBB, it quickly gained favor with politicians such as the then Georgia Governor Jimmy Carter who saw in ZBB an opportunity to force agencies to be accountable for every dollar appropriated to them.^[7] ZBB integrated essential functions such as comprehensive planning, resource allocation, program management, and evaluation into a system designed to bring greater rationality to the budget process.

ZBB

ZBB is designed to promote both efficiency and effectiveness, enhance both long- and short-term planning, eliminate redundancy and waste, and promote greater

accountability in organizations. It encourages greater participation and coordination among all levels of management in the planning, evaluation, and budget processes by initiating the decision-making process at the bottom tiers of the organization and aggregating them as they rise through the hierarchy. Zero-base budgeting makes no distinction between existing and new programs. The basis for budgetary decisions is a comparison that weighs the relative effectiveness of one alternative against another. Both new and existing programs are subjected to the same level of scrutiny and evaluation. Managers use the results of their evaluation to establish priorities among the existing and alternative programs and, as necessary, reallocate resources from lower to higher priority activities.^[8] Managers also periodically reevaluate policy objectives. On such occasions, new objectives may be set or existing ones may either be maintained or adapted to reflect current need. Ongoing evaluation of existing programs assures that planned objectives are being met.

The techniques used in the budget-making process are standardized so that all budget submissions are consistent in content and formulation. These objectives can be achieved by requiring managers to consider a range of alternative strategies and trade-offs as well as consider the potential consequences of implementing each one as they attempt to arrive at the appropriate allocation decision.^[6] ZBB, therefore, theoretically facilitates zero-sum reallocation. An increase in expenditures in one program area can be offset by reductions in spending elsewhere in the budget. In practice, however, forcing managers to make such trade-offs can be difficult. Sunk costs, longstanding commitments to clients, and strong public or political support are potential impediments to such trade-offs.

STEPS IN THE ZBB PROCESS

Sarant identified four basic steps in the ZBB process: identification of decision units, development of decision packages, review and ranking of decision packages, and preparation of the budget.^[9] (Sarant identified four basic steps in the zero base budgeting process—the model used here. Cozetto, Kweit and Kweit identified five steps, while others, including Pyhrr, defined as few as two). But before an organization can undertake the ZBB process, it must conduct a thorough analysis to determine its suitability for the system. Once that determination has been made, the organizational structure and managerial hierarchy should be reviewed to identify the organization's "decision units." A decision unit is a discrete program or service area within an organization under the supervision of a manager. A decision unit should encompass

all aspects of the activity, yet be the smallest unit for which basic allocation decisions must be made. The budget provides a good starting point for identifying decision units. Large decision units may be subdivided into several smaller ones so that a meaningful review of budget, objectives, and performance may be undertaken. Decision units should be carefully accounted for to assure that no duplication occurs.

The decision units become the focus of planning activities to set long-term goals and objectives, establish programs, and create policies. In the public sector, these steps may initially be the responsibility of the legislature rather than the organization. However, under ZBB, it is essential that the organization defines its own missions, goals, and objectives, as well as measures against which the program's performance will be compared.^[10] The performance measurement system should consider both the effectiveness of program activities and the efficiency with which that activity is meeting its objectives.

ZBB is known as a decentralizing budget analysis technique because it essentially works from the bottom up, although the process begins at the top. Upper management establishes goals and objectives for lower management to address, such as service levels, expected output, and funding targets. The remaining steps are undertaken by lower management, which carefully reviews its programs and proposes a range of alternatives (the number will vary) for achieving operational objectives.^[11] At every level of the organization, managers use their expert knowledge and judgment to select the preferred alternative on the basis of relative value, and by that selection, determine the resources necessary to implement them.^[12] The manager submits this information to upper management in the form of "decision packages."

A decision package is a brief report that specifies goals and objectives and provides relevant information that facilitates decision-making as the process moves forward, including information on program costs and benefits, performance measures, and alternative proposals for achieving them. The alternatives represent different levels of funding activity or performance, generally at the minimum level, current level, incremental level, and decremental level.^[13] During this phase of the process, careful consideration will also be given to the potential consequences of not providing the service at all. This information gives upper management a full range of funding options to use in their deliberations.

The third step in the ZBB process is the ranking of decision packages based on organizational goals and objectives. Decision packages are ranked in descending order according to priority, enabling managers to consider each alternative in the context of the others as well as the consequences of each option. In a large

organization with a multitiered hierarchy, the preparation and ranking of decision packages will occur several times. As decision packages ascend through the hierarchy they are consolidated into logical groupings. It is essential during this step to make decisions at the appropriate management level. Certain decisions are better made in the upper tiers of management, while others may be made at lower levels of the organization.

The fourth and final step in the ZBB process is the preparation of the budget. Done properly, the previous three steps provide the organization with not only a detailed and accurate assessment of resources needs but also the means with which to fund programs at levels appropriate in relation to competing priorities. Budget preparation proceeds according to guidelines set forth by the executive and the central budget office. Zero-base budgeting provides program managers with an essential tool for defending their budget requests.

Once a budget is approved and implemented, ZBB relies on evaluation and monitoring procedures similar to those employed in any other type of budget. Program performance, in particular, is closely monitored to assure that objectives are being achieved. Feedback is acquired through various means to assess problems or opportunities that arise during the fiscal year. All of these data are critical inputs in the ZBB system.

SHORTCOMINGS OF ZBB

Many of the shortcomings discussed in this section are not necessarily unique to ZBB, but they are particularly problematic for a results-oriented budget. An effective budget—one that successfully links an organization's output with what it intends to accomplish—relies upon timely, consistent, and reliable information and that is a fundamental challenge for ZBB. Schick^[14] argued that “the true test of performance measurement is in its use as the means by which organizations are managed and resources are spent.” Too often, however, performance measures fall far short of the mark for a variety of reasons. One problem is the unavailability of relevant measures at critical points in the budget cycle. Performance measures are generally collected at the end of a program year, while the budget calendar might demand that information at a much earlier date.

Forsythe^[15] outlined a number of others problems such as a lack of consistency between the information used in budget accounts and the information available to organizational units. Data on crosscutting programs—programs that cut across a number of agencies—is especially difficult to produce. Some budget systems require cost accounting, an expensive and difficult undertaking. Another obstacle to the

development of good performance measures may be the existence of divergent agendas among the legislature, agency officials, and other stakeholders participating in the development of performance measures. This can result in the collection of data not suitable for budgetary decision-making.

Even when relevant information is available, it cannot tell the whole story. Budgeting inevitably involves projections into the future and it is impossible to make accurate estimates of future needs or resources.

Conflicting objectives are also an impediment to an effective budget. Fossett, Gais, and Thompson^[16] coined the term “dissonant spillover” to describe the phenomenon of having agencies working in the same general policy sphere pursue conflicting objectives.

Legislation at the state and federal levels, which mandates minimum levels of service for certain programs presents another impediment to successful implementation of ZBB. Such requirements compel a government to maintain spending levels regardless of a preference for more efficient and effective programs.

Management and measurement of performance data are critical aspects of ZBB, which was first introduced during an era before desktop computers became standard office equipment. At that time, only the largest organizations had the capacity to do the detailed analysis necessary to develop meaningful measures. Although managers eventually overcame this problem, ZBB continues to have one major shortcoming: many managers find it simply too complicated and overwhelming, even with currently available technology. Zero-base budgeting requires a huge amount of data to be properly implemented, and managers and politicians alike often balk at the challenge of dealing with the volume of work. In addition, the entire process requires a sizable commitment of time.

In 2001, Metzgar and Miranda^[17] reviewed the new technology available to budgeteers in an effort to determine whether or not it would have an impact on budget reform. They cautiously concluded that while advances in technology can facilitate the adoption of budget reforms such as ZBB, “the behavior of budget actors and the structure of budget processes matter.” In his study of ZBB, Wholey^[18] found that lacking the means and commitment to make ZBB work, managers fell back into their old habits of devoting valuable resources to programs that were not a high priority, focusing more attention on the size of appropriations than on the efficacy of the programs being funded. In addition, some found that, for all their effort, they got a relatively minimal return. Savings in tax dollars and improvements to government programs were not enough to justify the high implementation costs of the ZBB system. In some cases, the alternatives ultimately adopted by the agencies using ZBB only reflected incremental changes from

the previous budget.^[19] Williamson concurred, citing that at least one experiment with ZBB found that it served a “negligible role” in shifting financial resources.^[20]

TODAY'S ZBB

Although ZBB was politically unpopular when President Jimmy Carter tried to bring it to Washington in the 1970s, it continues to have its proponents. It has been the foundation of Georgia's budget process as Carter was the state's governor.^[21] Colorado, Nebraska, Idaho, Iowa, and Oregon also use ZBB to some extent. Each uses a modified form of the system, pairing it with program budgeting or other forms of performance budgeting. Other states such as North Dakota have implemented the system on a trial basis, eventually opting not to use it but to adopt some of its elements.^[22] The system, however, remains largely unpopular with many practitioners who denounce it as a waste of time and effort. Perhaps the most interesting evidence of ambivalence about ZBB can be seen in Nebraska where the executive branch has embraced a results-oriented, performance-based approach while the legislature persists in utilizing the incremental approach.^[23]

The National Conference of State Legislatures (NCSL) as well as other professional organizations such as the National Association of State Budget Officers and the Government Finance Officers Association recommends the use of a modified form of ZBB. They base their recommendation on the fact that an annual review of existing and proposed programs is simply not feasible. All of the states that currently use ZBB, according to NCSL, employ a modified version of the system. Many opt instead to use target-based budgeting—a modified form of ZBB that does not require a full justification from base zero every year.^[21]

Interestingly, the 1994 Republican Contract with America^[24] proposed that ZBB be implemented at the federal level. Realistically, however, full review of some programs on an annual basis may be inappropriate, particularly when legislation mandates minimum service levels.

CONCLUSIONS

Zero-base budgeting is based on the nearly irresistible premise that it would help government to make optimal resource allocation decisions. It has ultimately failed to fulfill its promise because, in its purest form, ZBB runs counterintuitive to the behavior of practitioners involved in the budget process. Wildavsky and Hammond echoed the sentiments of many critics

of ZBB when they concluded “comprehensive ZBB vastly overestimates man's ability to calculate and grossly underestimates the importance of political and technological constraints.”^[5] Key's challenge has yet to be met.

REFERENCES

1. Key, V.O., Jr. The lack of a budget theory. In *Classics of Public Administration*, 4th Ed.; Shafritz, J.M., Hyde, A.C., Eds.; Harcourt Brace: Forth Worth, 1997; 109–113.
2. Lewis, V.B. Toward a theory of budgeting. In *Classics of Public Administration*, 4th Ed.; Shafritz, J.M., Hyde, A.C., Eds.; Harcourt Brace: Forth Worth, 1997; 178–191.
3. Tyer, C.; Willand, J. Public budgeting in America: a twentieth century retrospective; http://www.iopa.sc.edu/publication/Budgeting_in_America.htm (accessed 25 April 2005).
4. Burrows, G.; Syme, B. Zero-base budgeting: origins and pioneers. *Abacus* **2000**, *36* (2), 226–241.
5. Wildavsky, A.; Hammond, A. Comprehensive versus incremental budgeting in the Department of Agriculture. *Admin. Sci. Quart.* **1965**, *10*.
6. Pyhrr, P.A. *Zero-Base Budgeting: A Practical Management Tool for Evaluating Expenses*; John Wiley & Sons: New York, 1973; 33–35.
7. Patillo, J.W. *Zero-Base Budgeting: A Planning, Resource Allocation and Control Tool*; National Association of Accountants: New York, 1977; 1 pp.
8. Wholey, J.S. *Zero-Base Budgeting and Program Evaluation*; Lexington Books: Lexington, MA, 1978; 6 pp.
9. Sarant, P.C. *Zero-Base Budgeting in the Public Sector: A Pragmatic Approach*; Addison-Wesley: Reading, MA, 1978.
10. Sarant, P.C. *Zero-Base Budgeting in the Public Sector: A Pragmatic Approach*; Addison-Wesley: Reading, MA, 1978; 34 pp.
11. Patillo, J.W. *Zero-Base Budgeting: A Planning, Resource Allocation and Control Tool*; National Association of Accountants: New York, 1977; 25 pp.
12. Sarant, P.C. *Zero-Base Budgeting in the Public Sector: A Pragmatic Approach*; Addison-Wesley: Reading, MA, 1978; 73–75; Cozetto, Kweit, M.G.; Kweit, R.W. *Public Budgeting: Politics, Institutions, and Processes*; Longman Publishers: White Plains, NY, U.S.A.; 204 pp.
13. Patillo, J.W. *Zero-Base Budgeting: A Planning, Resource Allocation and Control Tool*; National Association of Accountants: New York, 1977; 3 pp; Sarant, P.C. *Zero-Base Budgeting in the Public Sector: A Pragmatic Approach*; Cozetto, Kweit, M.G.; Kweit, R.W. *Public Budgeting: Politics, Institutions, and Processes*; Longman Publishers: White Plains, NY, U.S.A.; 203 pp.
14. Schick, A. Getting performance measures to measure up. In *Quicker, Better, Cheaper: Managing Performance in American Government*; Dall, W.F., Nelson, A., Eds.; Rockefeller Institute of Government: Albany, NY, 2001; 39–60.

15. Forsythe, D.W. Pitfalls in designing and implementing performance management systems. In *Quicker, Better, Cheaper: Managing Performance in American Government*; Dall, W.F., Nelson, A., Eds.; Rockefeller Institute of Government: Albany, NY, 2001; 519–551.
16. Fossett, J.; Gais, T.; Thompson, F.J. Federalism and public management: health insurance food stamps and the take-up challenge. In *Quicker, Better, Cheaper: Managing Performance in American Government*; Dall, W.F., Ed.; Nelson, A. Rockefeller Institute of Government: Albany, NY, 2001; 207–244.
17. Metzgar, J.; Rowan, M. Bringing out the dead: can information technology resurrect budget reform? *Government Finance Rev.* **2001**, *17* (2), 11.
18. Wholey, J.S. *Zero-Base Budgeting and Program Evaluation*; Lexington Books: Lexington, MA, 1978; 7 pp.
19. Cozetto, D.A.; Kweit, M.G.; Kweit, R.W. *Public Budgeting: Politics, Institutions, and Processes*; Longman: White Plains, NY, 1995; 208 pp. See also Wholey, J.S. *Zero-Base Budgeting and Program Evaluation*; Lexington Books: Lexington, MA, 1978; 8 pp.
20. Williamson, D. *Zero Base Budgeting*; <http://business.fortunecity.com/discount/29/zbb.html> (accessed 22 July 2005).
21. National Conference of State Legislatures. *Fundamentals of Sound State Budgeting Practices*; <http://www.ncsl.org/programs/fiscal/fpfssbp.htm> (accessed 22 July 2005).
22. North Dakota Legislative Council. *Zero-Base budgeting*; <http://www.state.nd.us/lr//assembly/55-1997/interim-info/memos/99268.html> (accessed 19 July 2003).
23. National Association of State Budget Officers. In *Budget Processes in the States*; National Association of State Budget Officers: Washington, DC, 2002.
24. Republican National Committee. *Republican Contract with America*; <http://www.house.gov/house/Contract/CONTRACT.html> (accessed 31 July 2003).